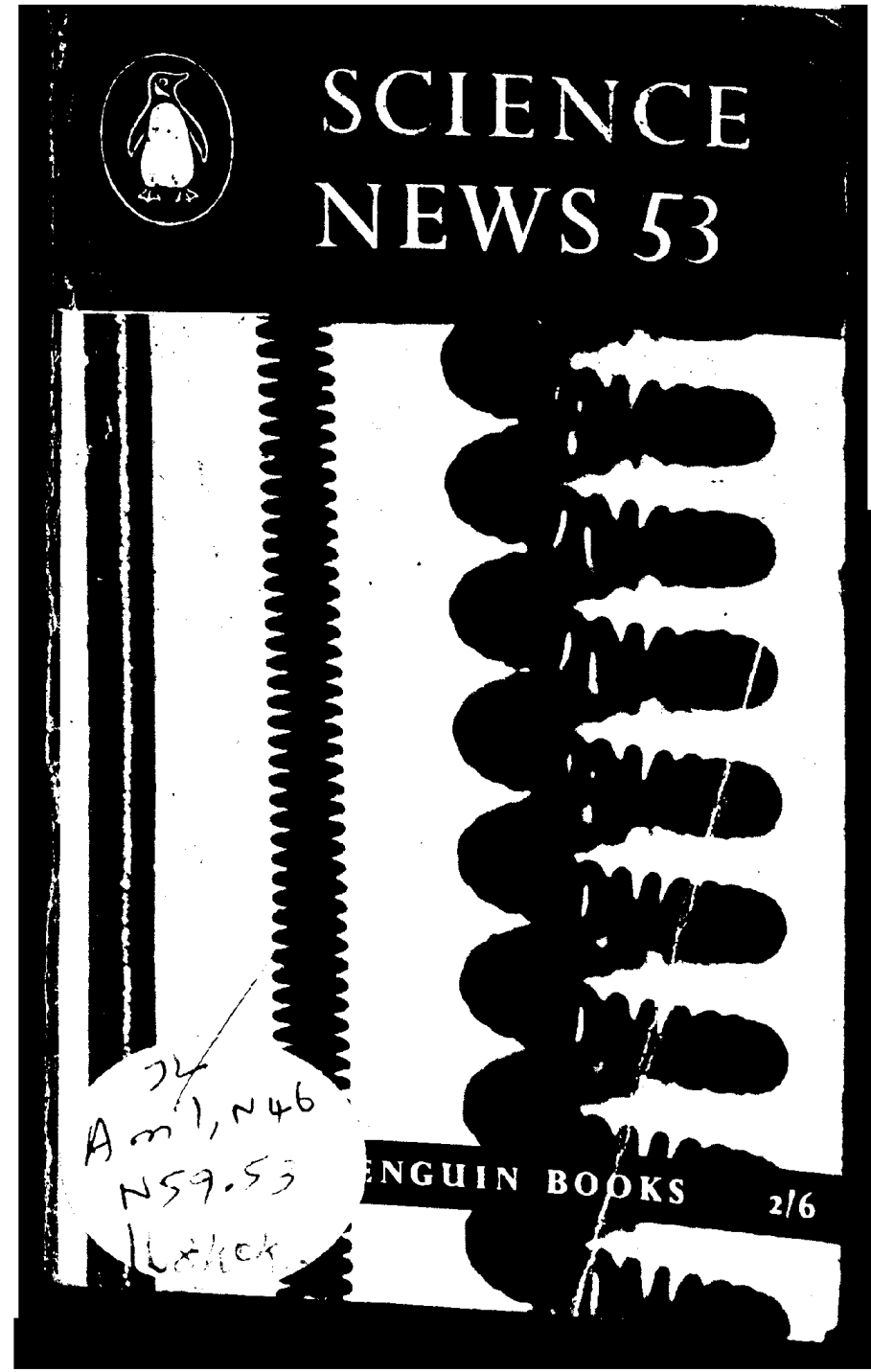


SUBSCRIPTION RATES

Owing to the printing dispute there will be only three issues of *Science News* in 1959, appearing in February, May, and October.

The subscription in the United Kingdom is 11/– for four issues, or in the United States 3 dollars, post free, payable through your bookseller or direct to Penguin Books, whose address can be found on the back of the title page.

Particulars of Australian subscriptions can be obtained from booksellers or from Penguin Books.

Numbers 5, 6 and 21–52 inclusive are still available. If there is any difficulty in obtaining any of these from a bookseller they may be had by writing direct to Penguin Books.

A microfilm version of the issues of *Science News* for 1959 is in preparation, and will be obtainable by subscribers to *Science News* on application to University Microfilms Ltd, 44 Great Queen Street, London, WC1.

SCIENCE NEWS

EDITED BY ARCHIE AND NAN CLOW

53

PENGUIN BOOKS

DOMESTIC SYNTHETIC DETERGENTS

J. R. P. MONKMAN

THE dramatic rate of increase in quantity and variety of synthetic detergents since 1946, when manufacture for domestic use was merely a few thousand tons of pure product annually, to an estimated 40,000 tons by 1955, has been unequalled in any other similarly based industry. This has made many people believe that these products are post-war inventions, and soon may be the only form of detergent available. Neither of these beliefs is correct.

The history of synthetic detergents goes back nearly a hundred and thirty years to 1831, when 17-year-old Edmond Frémy, research chemist, born at Versailles, at one time a pupil of Gay-Lussac, sulphonated olive and castor oils, which on neutralization were probably the first synthetic detergents ever made. Apart from a limited adoption of these products as auxiliaries by the textile industry of that time, further development was delayed until 1916 when, due to an extreme shortage of fats, Germany felt compelled to re-examine the potentialities in this field.

It rapidly became apparent that the initial degree of success in achieving a simple substitute for soap was to have far-reaching consequences, for many disadvantages of soap were being sidestepped by the synthetics. They appeared to function equally well in moderately hard and soft water, and nearly as well in supplies of very high hardness. They would tolerate without breakdown or precipitation some degree of acidity as well as the presence of calcium and magnesium and soluble heavy metal ions. Such increases in versatility appealed strongly to the textile trade, which encouraged these researches more than any other contemporary industry.

Before the different types of synthetics that developed from these original researches are examined in detail, it would perhaps be helpful to define the three essential properties that a detergent solution must display, and the manner in which it works.

(*a*) It must show strong surface activity and adsorption properties at liquid/solid, and liquid/ liquid interfaces.

(*b*) Very dilute solutions must decrease the surface tension of water and the interfacial tensions between oils and water to very low values.

(*c*) It must promote rapid and stable emulsification.

In more general terms, it must readily 'wet' both the dirt and the supporting material, separating one from the other, meanwhile suspending the liberated soil in such fine particles that redeposition is prevented.

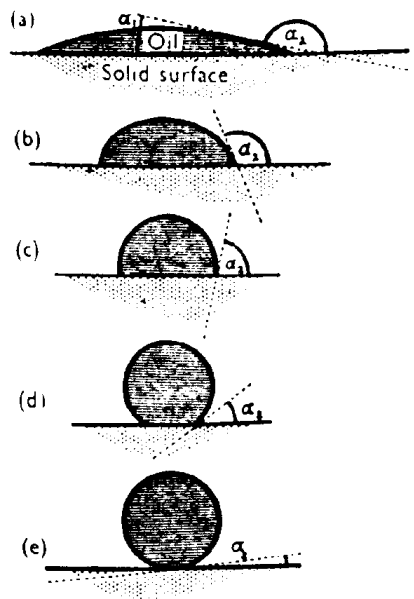


Fig. 1. Diagrammatic representation of removal of oily soil from a plane surface.

THEORY OF DETERGENT ACTION

The theory and mechanism whereby such wetting and desoiling is accomplished was first described in 1937 by N. K. Adam, who, through the medium of photomicrographs, contributed substantially to our knowledge in this field.

Imagine a solid surface contaminated by an oily soil. It will be seen from the diagram (Figure 1*a*) that the contact angle α_1 measured between the oily surface and the solid, at the edges of the affected area inside the oil, approaches zero, and the complementary angle α_2 measured between the surface of the oil and the clean area of solid approaches 180°. It is the behaviour of the contact angle α_2 we must observe while following the progress of detergency.

By definition, detergent solutions must decrease the interfacial tensions between oils and water to very low values, and as the relation between such forces obtaining in the system shown in Figure 1*a* may be expressed as

$$T_{so} = T_{sw} + T_{wo}.\cos\alpha_2 wo$$

where T is the appropriate interfacial tension, and '$_{so}$', '$_{sw}$' and '$_{wo}$' refer to solid/oil, solid/water, and water/oil respectively. Then by rearrangement,

$$\cos\alpha_2 wo = \frac{T_{so} - T_{sw}}{T_{wo}}$$

Since the adsorption of detergent at the water/solid and water/oil interfaces would result in a decrease in either or both of these interfacial tensions (T_{sw} or T_{wo}) and either trend will be seen by inspection to have the effect of increasing the value of $\cos\alpha_2$, it follows that the natural angle α_2 will be decreased.

The effect of a progressive reduction in the value of α_2 on the behaviour of the oil spot on the solid surface is shown in Figure 1 (*b* to *e*). As the contact angle approaches zero, the oil spot, which may have been merely a few molecules thick and by comparison immensely large in area, gradually shrinks in

diameter, thickening in depth, until finally it assumes nearly the shape of a sphere, attached insecurely to the solid base only by a small area of contact. In practice, the oil spot would probably break up into a series of 'islands' which would behave individually in this way.

While theoretically the separation of the oil from the solid could be achieved by the simple reduction of the contact angle to zero, in practice this ultimate reduction seldom if ever occurs, but the high adsorptive properties of the detergent contribute to remedy this deficiency. The detergent molecules have a greater attraction for the solid surface than does the oil, and the practical effect is to drive an infinite number of 'molecular wedges' between the oil and the solid until the residual area of contact, oil/solid, is finally almost eliminated. The combined effect is such that at this stage, a slight degree of agitation is enough to detach the soil from the base, and the object of the detergent is achieved.

Redeposition of soil

There remains the important question of preventing the dirt being redeposited elsewhere on the solid base. In the case of fabric washing, which is the chief function of domestic detergents, the swirling garments act as fine sieves through which the dirty wash liquor constantly passes. This filtering action may easily be responsible for redeposition of dirt.

The principle whereby this can be avoided is twofold. Firstly, the dirt particles are encouraged to exist in as small a dimension as possible, which enables them to pass through the interstices of of the fabric without being jammed like tennis balls into the netting surrounding the court; secondly, they are conditioned by emulsifiers so that once broken down they have no tendency to aggregate and form larger groups.

Certain chemicals, carboxymethylcellulose (CMC) being a particularly good example, behave excellently as emulsifiers to this end. Some differences of opinion exist as to the exact manner in which this effect is accomplished. Either the CMC encapsulates the dirt particles and neutralizes their inherent attraction for each other, or the fibres themselves adsorb the chemical, thus providing an 'insulating barrier' which repels redeposition. Or it may be a combination of both mechanisms.

In 1955, Stüpel set out to investigate the matter. He found that acridene orange would stain CMC and its presence could be detected by a red colour when viewed under the ultra-violet microscope. When fibres of cotton, wool, rayon, and ramie were treated in dilute solutions of CMC, such as might be found in typical laundry liquors, no CMC was detectable on the fibres, but soil, such as soot, natural oils and fats, mineral oils and fatty acids, deliberately added, or present accidentally, rapidly became thickly coated with CMC.

This work is still in progress, so no final conclusions can be drawn, but the indications are, so far, in favour of the encapsulation effect.

'Builders'

Carboxymethylcellulose is an example of a 'builder'; a chemical that can beneficially be added to pure synthetic detergents to enhance their properties. Many cheap inorganic alkaline chemicals possess some detergent properties of their own, and when combined with a synthetic may permit of a reduction in its quantity, due to their own contribution, without any total loss of detergency. This results in a less costly product.

It was recognized in the early part of 1940 that heavily soiled fabrics, cottons in particular, were not being effectively cleansed by the simpler domestic synthetics then available. This problem was partially solved by the introduction into the detergent formulation of massive doses of tripolyphosphates, and from that date onwards the separate development of light-duty 'unbuilt' detergents from heavy-duty 'built' products is discernible.

These ancillary chemicals fall into two main functional groups, those that assist detergency or emulsification directly, as already described, and others that perform a different specific duty, of which more will be said later. From this it will be realized that the packaged detergents bought over the counter are not 100 per cent pure active synthetic.

A hypothetical but typical analysis of such a product might be as follows:

	Per cent by weight
Pure active detergent	21
Sodium silicates	15
Sodium phosphates	36
Sodium chlorides and sulphates	22
Sodium perborate	3
Carboxymethylcellulose (CMC)	0·5
Lauryl alkanolamine	2
Optical white	0·5
Direct blue/green dyestuff	trace
Perfume	trace

It will be seen that nearly 80 per cent of the compounded detergent consists of ancillary material, although it should be noted that the chlorides and sulphates, being unremoved by-products of manufacture, act simply as diluents.

The main types of chemicals available for the building of home detergents are the sodium tripolyphosphates just mentioned and other condensed phosphates, as well as ortho, meta, and sesquisilicates, perborates and percarbonates, carboxymethylcelluloses (CMC), sodium-starch glycollates, methyl-ethyl or hydroxyethyl celluloses, water soluble cellulose ethers, alkanolamines, optical whites or fluorescents, plus occasionally traces of 'direct' coloured dyestuffs, clays such as bentonite, and more rarely sodium carbonate or sesquicarbonate, the last two being quite strongly alkaline and generally considered less suitable for domestic purposes.

THE ACTIVE DETERGENT

The pure detergents themselves can be divided into three main classes according to which portion of the molecules behaves as the active component.

Pure detergent

Anionics | Non-ionics | Cationics

The anionics dissociate in water yielding a large amphipathic anion (i.e., receptive to both oil and water) and a small metallic cation. The anion is the 'working unit' hence the class-name 'anionic'. The cationics behave in a reverse manner, as it were, the large active amphipathic ion being the cation, its anion partner being usually a small passive halide. Non-ionics, on the other hand, do not dissociate in water at all.

The cationics generally are inefficient in the alkaline conditions that are normal for domestic purposes, and as a consequence the field of retail synthetic detergents is filled either by the anionics or the non-ionics.

The compatibility of one class with another varies. Non-ionics may be mixed in all proportions with either of the other two, and in the case of the anionics, are often so used to boost efficiency, but generally anionics and cationics are mutually destructive.

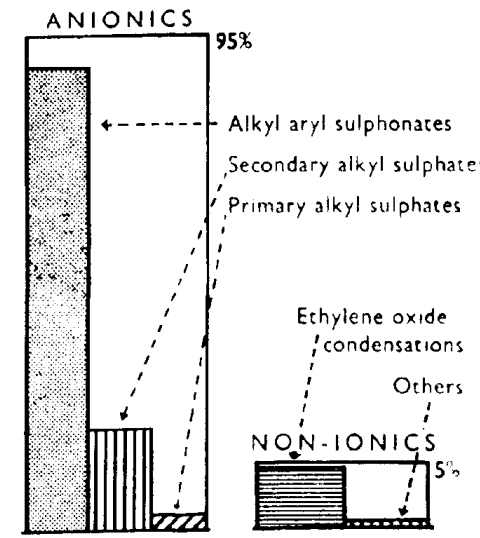


Fig. 2. Proportions of various types of active detergent in use at present.

As will be seen from Figure 2 anionics form 95 per cent of all domestic synthetic detergents; within this group are the spray dried powder products such as SURF and OMO of Lever Bros., TIDE and DAZ of Thomas Hedley, or FAB of Colgate Ltd. The remaining 5 per cent are non-ionic and consist of the concentrated liquid detergents of which STERGENE of Domestos Ltd is probably the best known.

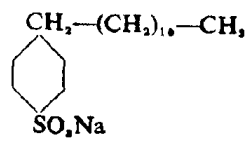
The anionic class can now be further divided into three subgroups (see Figure 2).

(*a*) Primary alkyl sulphates
(*b*) Secondary alkyl sulphates
(*c*) Alkyl aryl sulphonates

By far the largest group is the last; almost 95 per cent of the total anionics, and almost certainly the alkyl aryl sulphonates, constitute the active detergent in the domestic package, because as dry non-hygroscopic powders they are eminently suitable in this capacity.

They are manufactured in stages by alkylation of an aromatic hydrocarbon, probably benzene, with either a petroleum fraction or similar synthetic hydrocarbon such as polypropylene, of carbon chain length 9–15 atoms, in the presence of the catalyst aluminium trichloride, by the Friedel-Crafts condensation. The resulting product is then neutralized by caustic soda, purified by distillation, and sulphonated either by concentrated sulphuric acid or oleum (i.e., H_2SO_4 + added SO_3), the sulphonation stage being kept at a moderate temperature to avoid discoloration.

The resulting alkyl aryl sulphonic acid ($R.C_6H_4.SO_3H$) is again purified, and finally neutralized by soda alkali, some residual sodium sulphate remaining on the final drying of the product. A typical compound would be sodium dodecyl benzene sulphonate.

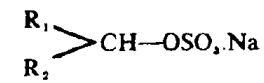


It is interesting to consider briefly the effect on the ultimate

product if the number of carbon atoms in the alkyl radicle is varied. With a very long hydrocarbon chain, the solubility of the final compound is much reduced, if indeed the product can be prepared at all, because ease of sulphonation is dependent on chain length. On the other hand if the aromatic group is made larger, naphthalene being substituted for example, this also reduces the solubility. Alternatively if the carbon chain is shortened, solubility increases hand-in-hand with wetting power, but detergency is reduced. By a combination of these variations, such as increasing the size of the aromatic group at the same time as the alkyl chain is shortened, detergency is saved, although the products are still primarily wetting agents.

During the years 1938–49 secondary alkyl sulphates were used extensively, and were the basis of many domestic detergents during that period, but following the development of the alkyl aryl sulphonates, the slightly sticky, hygroscopic nature of the secondary alkyl sulphates caused them to fall into disfavour and now they represent less than 5 per cent of all the anionics.

They are produced by sulphonation of the secondary fatty alcohols, or olefins from petroleum fractionation, and have the typical formula:



where R_1 is an aliphatic group of carbon chain length between 12 and 18 atoms, and R_2 is probably methyl (CH_3).

A typical product of this type, sodium alkyl sulphate, is marketed by the Shell Chemical Co. under the name of TEEPOL. It is offered either as a concentrated liquid, or it can be spray dried to powder form, in which state it used to be incorporated in the earlier domestic detergents before the common adoption of the alkyl aryl sulphonates.

The tiny remaining group of primary alkyl sulphates, a mere half per cent of the total anionics, owe their insignificant position to costs of production. They are excellent detergents, though admittedly slightly inferior in wetting power to the secondary alkyl sulphates. Derived from natural fats and oils,

they cannot compete in price with the almost illimitable flood of cheap petroleum bases from which the alkyl aryl sulphonates are compounded. Sodium dodecyl sulphate (CH_3—$(CH_2)_{11}$—OSO_3.Na) is typical of this primary alkyl sulphate group; it was used in the formulation of the earliest soapless shampoos of the 1930s, DRENE being an example.

Of the other main class of domestic detergents, the liquid non-ionics, the most important are now the condensations of polyethylene oxide with either dioctyl phenol or cresol. These have the general formula

$$R—O(CH_2—CH_2)_{20\text{-}23}—OH$$

where R can be either an alkyl or aryl group.

Lissapol N, manufactured by Imperial Chemical Industries Ltd, is typical of this class. After some modification it is supplied to the domestic market by Domestos Ltd under the name of STERGENE.

The manufacture, briefly, is as follows; phenol and polypropylene are caused to interact in the presence of boron trifluoride catalyst to yield the intermediate, the alkylated phenol

$$\underset{\text{phenol}}{C_6H_5OH} + \underset{\substack{\text{polypropylene}\\ n=9 \text{ or } 12}}{(CH_3.CH.CH_2)_n} \xrightarrow[\text{catalyst}]{BF_3} \text{alkylated phenol}$$

This intermediate is then reacted with ethylene oxide under slight pressure at about 150° C, in the presence of caustic soda, to yield the end product, Lissapol N.

The non-ionics, as a class, one suspects have as yet merely crossed the threshold of development. They display traits of behaviour which, even for the non-technical domestic user, distinguish them as being in some way subtly different from the general run of synthetic detergents.

One factor, for example, which is typical of their singularity, concerns solubility. In the early stages of development, it was common to find the clear homogenous solution of the detergent in cold water becoming cloudy with precipitated surfactant as the temperature was raised. Opinions varied as to the degree to

which detergency was lost by this partial solubility at elevated temperatures, but theory apart, it was agreed it was expedient to investigate the possibilities of the molecule remaining truly soluble at all temperatures.

One method investigated was by sulphating the terminal hydroxyl group (OH); but then the product ionized in solution and so ceased to be a true non-ionic. Some increase in solubility in hot solutions was however achieved by increasing the amount of ethylene oxide in the final manufactured product. Yet another difference is the much lower general level of foaming displayed by solutions of proper concentration, when compared with anionics.

Many domestic users, conditioned by years of experience with soap, and later the dry powder anionics, regard the low foam level of the non-ionics as indicative of an insufficiency in quantity of detergent, so excess is added. This might prove to be a super-excess, as the fact that they are liquids tends to cause most of us to overestimate the initial requirement anyway. In the early days, STERGENE suffered some loss of sales in consequence, but this product has recently been modified by builders to permit of its competition on more level terms with the general purpose built anionics. Whilst a sufficiency of foam may be necessary to satisfy the instincts of the user of the agitator type of domestic washing machine, commercial laundry operators using the rotary washers find the reduced foam level of the non-ionics extremely beneficial in avoiding the excessive foam overflow common to some anionics.

A new product, 'HYFAC 431', a mixture of fatty acids derived from hydrogenated fish oils, has been recently patented by Thomas Hedley Ltd. When added in concentrations of about 3 per cent of the total weight of detergent, it has been found to act as an effective suds depressant on the alkyl aryl sulphonate of heavy-duty anionic detergents, without reducing their efficiency.

While the development of HYFAC 431 indicates the active interest taken by this firm in attempting to overcome some disadvantages of anionics met in practice, it is understood that

HYFAC 431 is neither currently in use as a builder nor can supplies of it be made available.

Solid Form Synthetics

Many attempts have been made to produce non-ionics in solid form. Since they possess an inherently high detergent efficiency, only about 10–15 per cent of the final product need be pure detergent, and the added builders are able to absorb this low ratio of liquid surfactant satisfactorily, to give free flowing powders, but their lack of popularity, to date, has been largely due to their very high density, which compares unfavourably in the mind of the domestic purchaser who sees offered larger packets, for similar cost, containing the lighter, fluffier anionics.

THE QUESTION OF FOAM

So far, either abundance or apparent deficiency of foam resulting from the different types of detergents has been treated as a subsidiary effect, for in spite of the instinct to associate it with detergent potential, it is more correct to say that foaming and detergent power are independent consequences of surface activity of detergents. This is not always fully appreciated. Frothing and foaming result from surface activity being manifest at water/air surfaces; but detergency depends on adsorption effects at water/oil and water/solid surfaces, mainly below water level.

The strong surface activity of most detergents is due to the large amount of hydrocarbon in the molecule. This always tends to come out of solution in water owing to the greater attraction that water molecules have for each other rather than for the hydrocarbon molecule, or even the attraction that the hydrocarbon molecules themselves have for each other. The consequence is that the water molecules band together in a concerted attempt to oust the intruder hydrocarbon.

Sisely has shown that while most detergents foam readily, the presence of foam is by no means essential for detergent action, for when soap foam is deliberately destroyed or repressed by antifoaming agents (e.g., tributyl phosphates) little

reduction in detergent power is discernible. Further, some substances, saponin for example, foam extravagantly, but possess hardly any detergent power at all.

Nevertheless, it would be incorrect to believe that foam is of no value whatever in practical detergency. It can assist in soil removal by the mechanical scrubbing action exerted by the slight rigidity of the bubbles, which also contribute their mite in assisting the breakdown of dirt globules into smaller particles, the better to be emulsified.

Foam and public health

At first thought there seems to be little direct relation, but during recent years their interdependence has become a matter of serious concern for sewage authorities. They have found that normal and unavoidable agitation during treatment of domestic effluent containing waste synthetic detergents produces an enveloping blanket of foam which can smother and destroy the bacteria of the purification beds on whose action depends the successful conversion of domestic sewage into innocuous liquors. Also, certain types of detergents show extreme resistance to breakdown and destruction by bacteria, and these two factors have seriously threatened efficient working of this vital service.

As early as 1953, the problem of excessive foaming under such conditions was recognized and methods to control it explored. The commonest was that of spraying large volumes of water on to the foam. Even up to 10 per cent of the total volume of effluent flow was needed in some cases. Initially, when many of the detergents were based on the secondary alkyl sulphates and the early alkyl aryls, sufficient success followed this method to dispel immediate anxiety, but progressively, as public demand required greater stability and volume of foam, specific builders were added to satisfy this requirement and the problem again began to assume serious proportions. At one sewage works the capital cost of plant and machinery for spraying purposes was calculated to be £20,000, with an additional cost of £15,000 a year for maintenance and operation. The use of chemical defoamants is not generally acceptable to sewage authorities,

partly because of the high cost of such chemicals, but more because of the fear that such treatments might themselves be equally deadly to the bacteria.

Recent researches by the Royal Dutch Shell Chemical Co. of Amsterdam have traced the cause chiefly responsible for this defiance to bacteria to the molecular configuration of the alkyl aryl sulphonate group of anionics, which, it will be remembered, form the basis of the vast majority of the dry powder products in popular use.

Their finding has been that the hydrocarbon side chain attached to the sulphonated benzene ring is commonly very complex in its branching and arrangement, but when modified to a simpler, straight chain, the molecule of sodium dodecyl benzene sulphonate is rendered 'biologically soft', in their own words, and much more vulnerable to subsequent bacterial destruction, without any loss in detergent efficiency.

This newly created product, Dobane JN, shows a satisfactory head of foam when fabrics are boiled, which the housewife regards as essential, and yet promises to avert the unhappy and dangerous consequences at the sewage works already described. Modifications to the Newhaven plant of Shell Chemicals are in progress, and it is anticipated that within 12–18 months, an output of 30,000 tons per year will be available for use in retail products.

Pursuing the question of foam a step further affords the opportunity to examine the uses of some specific builders. It will be appreciated now that different basic surfactants have different inherent degrees of foaming ability. Often, there is natural to the product an adequate amount for aesthetic or practical reasons, but if for some particular end use more is required, small additions of lauryl alkanolamine remedy the deficiency to a remarkable degree.

WHITER THAN WHITE

The problem of maintaining whiteness of fabrics and garments throughout their lifetime is a perpetual source of concern to the average housewife. To this end manufacturers commonly

incorporate in domestic synthetics either percarbonates or perborates, which act as a source of the bleaching agent, hydrogen peroxide. The mildness of this oxygen-rich bleach compared with the extreme vigour of the hypochlorite liquors retailed separately, coupled with the effect of an automatic rationing of the quantity used, due to an instinctive aversion to wasting detergent, ensures that the bleaching action is confined within bounds that result in minimum harm.

Progressive damage to fabrics is an inevitable consequence of the attack of chemical bleaches on fibre and soil alike, and premature weakening of garments can often be attributed to improper and excessive use of these agents. This unfortunate circumstance, unhappily all too common, has encouraged the use of another specific builder, the optical or fluorescent white.

Yellowing of white fabric is due mainly to absorption of the bluish element of the visible spectrum. These optical whites, chemically similar to the direct cotton dyestuffs, absorb the invisible near ultra-violet radiation of natural daylight, and to a lesser extent artificial light, between the wavelengths 300–400 mμ. They re-emit this energy within the wavelengths 400–500 mμ, which is in the blue region of the visible spectrum. This fluorescence is not only effective in correcting the shade of white, but additionally it donates an extra brilliance. This has become the focal point of considerable commercial advertisement.

These optical whites, mainly derivatives of diaminostilbene or benzimidazole, were discovered originally by Krais in the mid-1930s, but their use in domestic detergents has only become common during the past four or five years. Between 50 and 100 compounds are known already, but many are unsuitable for inclusion in domestic products, for reasons of solubility or specific substantivity, but recent products, the benzoyl derivatives of 4 : 4 diaminostilbene 2 : 2 disulphonic acid, promise to combine the requirements of adequate water solubility, an extreme fastness to light, and a high intensity of fluorescence to weight ratio.

As an example of the difficulties of compromise, the effects of

choosing optical whiteners of different fastnesses to light and washing can be quoted. Textile bleachers and finishers are currently using these whiteners to a wide extent to give bleached goods (e.g., poplins for shirtings, etc.) a quality of whiteness and brilliance previously only barely achieved by deep bleaching. The reduced treatments in the destructive hypochlorites that are consequently needed, when in combination with the optical whiteners, permits the fabrics to leave the bleach-croft measurably stronger.

It is obviously of advantage that the light and wash fastness of such original treatments should be of a high order, so that no loss in whiteness should occur in storage or use. On the other hand, it is not uncommon to receive, in textile laboratories, garments that have been vat (fast) dyed, in pastel shades in particular, with the complaint that following successive domestic launderings the pastel colour has washed out. On examination, it is frequently found that successive launderings in domestic detergents containing substantive optical whiteners has built up a level of fluorescent dye to such a degree that the original pastel shade is being progressively blocked out. Appreciable changes in colour can occur in this way.

A very severe acid oxidation treatment is often required to remove this heavy over-dyeing of optical whitener, and in a large majority of cases it can be shown that the vat shade remains underneath, unchanged, which vindicates the textile dyers' original choice of a wash fast colour. Such severe restorative treatments as this appreciably weaken the fabric, often to a degree making it unsuitable for further wear, and in no sense can they be regarded as a cure, but the point emerges that had the detergent manufacturer chosen an optical whitener of poorer washing fastness, each successive home laundering would have tended to have removed the previously applied layer, as the new deposit was imposed, and the colour changes described would have been greatly delayed, if in fact they became noticeable at all during the garment's life. There is, of course, also the possibility that the housewife has used more detergent than she should have.

Effect on the Skin

Fears are sometimes expressed of the dangers of dermatitis resulting from the use of synthetic detergents. A summary of conclusions based on inquiries made by the British Launderers' Association, the Ministry of Health, and detergent manufacturers themselves, in addition to experimental work, both in this country and the U.S.A., on the irritant effect of synthetics on humans and animals, shows that although dermatitis of one form or another can and does occur, the incidence is probably no greater than that resulting from the use of other washing products, and in many cases may have been due to the use of unnecessarily strong solutions.

CONCLUSION

What of future trends in the domestic field of synthetic detergents? Work on the production of the solid non-ionics continues, as do attempts to produce a synthetic bar soap for toilet purposes, which so far have also been only partially successful, this market being still almost completely satisfied by soap-based products.

The low solubility of the alkyl aryl sulphonates has prevented their competitive use as concentrated liquid detergents, this demand being filled by the two main manufacturers of the non-ionics, but recent work using organic alkalis, ammonia or triethanolamine, in place of caustic soda for neutralization after sulphonation, perhaps in combination with solubilizing agents such as alcohol, suggests that in this way the field of liquid detergents may be extended.

In regard to formulation, the established pattern of creating an increasingly complex compound by grafting on to the pure root synthetic a number of builders, designed to accentuate and correct the characteristics of the pure detergent, plus others which serve specific duties, will probably continue in an attempt to produce a balanced domestic product.

Regarding production, while the annual tonnage of pure synthetic base for domestic purposes continues to rise, the spec-

tacular rate of increase between 1951–3 of 14,000–29,000 tons has never been repeated, and indications are that further increases in consumption will be more modest, probably dependent chiefly on increases in population.

The question 'Is soap going to disappear from the market altogether?' can best be answered by stating that a balance now appears to have been struck between demands for synthetic detergents and soap based products, and no further significant swing in favour of synthetics is anticipated in the foreseeable future.

WHAT IS THE ORIGIN OF SPECIES?

C. H. WADDINGTON

ALTHOUGH the book that sold the theory of evolution to the scientific world was called 'The Origin of Species', it is precisely the problem of how species originate that remains one of the most mysterious questions in the whole complicated edifice of theory and fact that has been built up in the hundred years since Darwin wrote. We have made great advances in understanding how the character of a group of animals or plants may gradually change as the generations pass. This is the progressive aspect of evolution. It does not necessarily involve the splitting of one original population into two or more distinct groups which it is worth while to name as different species. Certainly, when one follows such a sequence of changes throughout a long period, for instance, in geological formations that contain a continuous fossil record of some particular type of creature, one may find that the earliest representatives and the latest ones differ from each other sufficiently to be ranked as different species. But these are species that exist at different times, one at a geologically later period than the other. Such species differences are not quite comparable to those that we find existing contemporaneously in the world of today. In the origin of a later species from an earlier ancestral one, there need never be a stage at which the population of animals is split into two separate groups; the later species may be formed merely by a gradual progressive alteration of the whole population. On the other hand, the many different, but related, species that exist simultaneously at a single time, must, if they are produced by evolution, have been formed by a process that involved splitting up an originally single population into a number of distinct sub-groups.

Since Darwin's day we have discovered a great deal about the nature of the differences that exist between such contemporaneous and related, but yet definitely distinct, species. We know that in many cases the species difference may involve changes in the arrangement of the chromosomes, which may have been broken and rejoined in various ways. Such chromosomal alterations can be detected by cytological study, in particular of hybrids between the two species, if those can be produced. The evidence suggests, however, that such comparatively crude changes as a reshuffling of the chromosomes are not an essential feature of species divergence. They seem more usually to be secondary effects which have occurred after the initial stages in the splitting of an originally single species into two sub-units.

A much more important fact is that each species has a characteristic integrated system of heredity, and that these systems differ from one species to the next. A species consists of a group of populations which share, to a large extent, a common hereditary endowment. One can regard this endowment as consisting of all the genes which commonly occur in the species. As Th. Dobzhansky, in particular, has emphasized, this common 'gene pool' is not just a haphazard assemblage of individual genes; on the contrary it is integrated or organized, in the sense that genes belonging to one gene pool fit in reasonably well with the other members of the pool. Thus any individual, whose genes one can think of as being drawn at random from the gene pool, will in general be a reasonably well-adapted being. On the other hand, genes belonging to one species may not fit in at all well with the gene pool of some nearly related species; the hybrids will be unhealthy, or infertile, or ill-balanced in some other way.

This organization of the hereditary endowment of a species is spoken of as the 'co-adaptation' of the gene pool. Co-adaptation occurs not only within the species as a whole, but to a minor extent at least within separate populations belonging to the same species. Differences in co-adaptation tend to occur whenever two populations, for one reason or another, fail to interbreed. The barrier between them may be merely distance, so that the members of one population, although they would be

perfectly ready to breed with the members of a second population, and would be perfectly fertile with them, never in fact come across them. If two populations, belonging to the same species, remain for a considerable time with little interbreeding, it is frequently found that their co-adapted gene pools have diverged to some extent. It seems certain that the most typical cause which produces the differences in co-adaptation between different species is of essentially the same kind, namely, their failure to interbreed. But the lack of cross-breeding must continue for a long time if two populations are to diverge sufficiently to be ranked as different species. The co-adaptation of the gene pool of a particular group must arise only very gradually, as natural selection weeds out those genes which, although not specially harmful in other combinations, happen not to fit in well with the constellation of genes that is becoming characteristic in this population. Quite a small amount of cross-breeding would very greatly slow up this gradual divergence, or even abolish the difference entirely. For this reason it seems unlikely that mere distance would be a sufficiently effective barrier to allow two groups to evolve into different species; if there is no absolute barrier to migration from one end of the range to the other, there will always be a slow drift of genes throughout the whole area that the species occupies, and the local groups are not likely to diverge very much from one another.

That brings us back to the other difference which, we know, very frequently distinguishes two nearly related species; that is, they frequently exhibit some other barrier to cross-breeding that is not dependent on distance. This very commonly takes the form of a disinclination of the individuals of one species to mate with those of another; and, further, such matings may be relatively infertile, or produce unhealthy or inferior offspring. The existence of a reproductive barrier of this kind between species is not universal, but it is so common that it has been considered by some people as a criterion for deciding whether two groups should be ranked as different species or not. Clearly when it is in existence it will make it much easier for the gene pools of the two groups to diverge, since it will prevent the

mixing that would tend to bring them back to uniformity.

One of the great puzzles in understanding the origin of species is to discover how these reproductive barriers arise. If one has a uniform population of animals, all of which are perfectly ready to mate with any member of the opposite sex they may come across, any aberrant individual who starts being unduly selective, and exhibits a reproductive barrier that cuts him off from some of his potential partners, will be at a disadvantage and is likely to leave fewer offspring. That is to say, natural selection would tend to eliminate him. We therefore have to think of some special method by which the first steps in sexual isolation could occur. From a theoretical point of view two suggestions have been put forward. One, which is often associated with the name of H. J. Muller, attributes the origin of sexual isolation to pure chance. This amounts to saying that we just have to accept it as an empirical fact that, when the gene pools of two groups start to become differentiated from each other, one of the first signs will be that the populations begin to acquire different sexual preferences, which will cause individuals to tend to mate within their own population rather than crossing with members of the other. The alternative suggestion is to point out that as soon as the gene pools of two groups have diverged somewhat the hybrids between the populations will tend to be inferior, and that there will therefore be a natural selection against the readiness for cross-breeding. This point has been particularly emphasized by Dobzhansky, who has suggested that this type of selection against readiness to cross-breed may be involved even in the very first stages in the evolution of a reproductive barrier.

So far little experimental work has been done on this problem. It is obviously very difficult to investigate in nature the situation of species as they begin to become distinct from one another. One can, of course, discover populations which seem to be in the first stages of species divergence and one can show that they may differ in such matters as mating preference, but it is almost impossible in such cases to discover how these differences in reproductive behaviour have originated. We are in fact

thrown back on laboratory investigations. In our group in Edinburgh we have recently made two studies; one along the lines suggested by Muller and the other on those put forward by Dobzhansky. In both cases the organism used was the well-known fruit-fly *Drosophila melanogaster.*

Laboratory Studies

In the first case we wanted to discover whether, if stocks of this species are kept isolated from one another for many generations, tendencies towards reproductive isolation will occur by chance. We have in the laboratory many strains that have been isolated from one another for a long time because they have been involved in other experiments, for instance programmes of selection or inbreeding. In the investigations of reproductive isolation we used six inbred stocks that had a relatively normal hereditary constitution and appeared indistinguishable from normal flies; three other inbred stocks, each of which contained one mutant gene (yellow body colour, vermilion and white eyes); four lines that had been selected for many generations for a high number of bristles on one of the segments of the abdomen, and another four lines that had been similarly selected for a low number of bristles. The stocks of each of the three groups (of wild type inbreds, of marked inbred, and of selected lines) were tested against each other by allowing males to choose between females of their own and of another strain, and also by allowing females a similar choice between males from different strains. We found that the marked inbreds stocks differed strongly from each other in general sexual vigour, and there were some definite tendencies towards reproductive barriers. For instance, all normal females show an aversion to yellow males, while yellow females, perhaps because they are rather feeble, are very receptive towards males of any kind. However, these peculiarities of reproductive behaviour in obviously unusual races are not perhaps a very good guide to what might happen in nature. The normal-looking inbred strains, and also the selected strains that still remain fairly normal in appearance, are more interesting in this connexion. We found that among

the inbreds there were in fact quite well marked tendencies for individuals to mate preferentially with their own kind. In four out of the six inbred lines tested, the tendency for the male to choose his own type of female rather than any other was quite definite, and in two of these four cases there was the same tendency amongst the females. In the lines selected for high or low bristles number, however, there seemed to be almost no tendency for intra-group mating: either sex, given a choice between its own strain and some other, seemed to choose more or less at random.

The fact that in the normal-looking inbred strains there was a tendency for an individual to prefer to mate with its own group is rather good evidence that at least the first steps in reproductive isolation may arise by chance. It is interesting, too, that it arose amongst the inbreds and not amongst the selected strains. According to genetic theory, one would expect strains that had been inbred in reproductive isolation from one another for many generations to come to differ in a large number of very varied hereditary qualities, while strains that have been selected for particular qualities will tend to differ only in respect of the genes that have a strong influence on those particular characters. It is therefore more likely that some sexual isolation will occur by chance between the inbred strains than that it will arise between the selected races.

Although here we have some evidence in favour of Muller's suggestion of a chance origin of tendencies towards sexual isolation, this trend had not gone very far in our inbred strains, even though they had been isolated from one another for about sixty generations. The Muller mechanism is in fact hardly likely to be the whole story. What it may do is to provide a first step towards a reproductive isolation. In another experiment we learned that the mechanism suggested by Dobzhansky – namely a selection against cross mating – can also be effective, and could therefore come in as a second stage to strengthen an incipient reproductive barrier.

For these experiments we took two strains that were recognizable because each contained a mutant gene. In one case this was

ebony, which produces a dark body colour, while in the other the strain contained vestigial, which reduces the wings to small stumps. Flies of the two strains were put together in a large container and allowed to mate. If two ebony flies mate their offspring will be ebony, and similarly if two vestigial flies mate they will produce vestigial progeny; but if an ebony mates with a vestigial the next generation from this pairing will be of the wild type, showing neither the ebony nor the vestigial character. From each generation we therefore picked out some ebony and some vestigial offspring to carry on the population, rejecting all the wild type individuals which had been produced. Thus we were exercising strong selection against cross-breeding, and the population was perpetuated only by those matings of like with like. We found that in point of fact this selection did lead to a decrease in the frequency of cross matings; there was a considerable increase in the tendency of vestigial flies to mate with vestigial, and a slightly less well marked increase in the tendency of ebonies to choose to mate with ebonies .

This gives good evidence that the mechanism suggested by Dobzhansky can actually work. If hybrids between two co-adapted populations are relatively at a disadvantage, then natural selection should be able to build up a reproductive barrier.

In nature, as opposed to the laboratory, there is another factor which may play a considerable part in the evolution of reproductive isolating mechanisms. Modern evolutionary theory tends to pay little attention to the fact that the mode of life of a species is partly under its own control. An animal can, to some extent, choose to live in one place rather than another, and to conduct its life in one of the possible ways open to it. Consider a species which has been living in one area, but which is now spreading into some new region – a region with a rather different character from that to which the immigrants have become adapted. We are prepared to admit that natural selection will tend to change the hereditary constitution of the population that moves into the new area so as to adapt it to the special conditions; and we can see that it will be a gain to the immigrants

if they bred with one another and do not cross back with the parent populations and thus lose the hereditary adaptations that natural selection is building into them. What is often forgotten is that the fact of moving from the old area into the new one, the process of taking up a slightly different form of existence, will probably involve differences in behavioural preferences. And if one set of behavioural preferences become changed, it seems probable that others, for instance those involved in mating choice, will be changed also. Thus alterations in mating behaviour seem rather likely to occur purely by chance, according to Muller's suggestion, simply because the whole behavioural system of the animals must be in a fluid state to allow the colonization of the new area to occur at all.

REFERENCES

Dobzhansky, Th.: *Genetics and the Origin of Species* (Columbia University Press, 1941).

Knight, G. R., Robertson, A., and Waddington, C. H.: 'Selection for sexual isolation within a species', *Evolution*, 1950 **10** 14–22.

Muller, H. J.: 'The Darwinian and modern concepts of natural selection', *Amer. Phil. Soc.*, 1949 **93** 459.

Santibañez, S. K., and Waddington, C. H.: 'The Origin of sexual isolation between different lines within a species', *Evolution*, 1959 **12** 485.

THE END OF CHEMISTRY IS ITS THEORY

P. COOK

THIS passing year, 1958–9, is the centenary of a beautifully aesthetic scientific theory – and I do not allude here to the much publicized doctrine of Natural Selection which, like antisepsis, in some quarters may be considered aesthetic – but to the theory of the chemical nature of carbon and the molecular structure of its compounds. The theory was put forward independently, and almost simultaneously, by August Kekulé and Archibald Scott Couper, at the time both young men in their twenties. It was not simply a satisfying theory; it was richly rewarding, a sort of open-sesame. Before 1858 progress in organic chemistry had been mainly analytical, with a piling up of a mass of largely uncoordinated facts; after 1858 there came a period of synthesis, with a rapid and guided production of an immense number of carbon compounds – almost half a million have been isolated and named. A theory was a necessity, that is a successful theory, for in the first half of the nineteenth century there was no lack of theories, but, unfortunately, they just did not fit the facts, or at least all of them.

The big three in chemistry, Berzelius, Dumas, and Liebig, each had a theory of his own. In this, however, they agreed, that the radicals of organic substances were of a complex nature (even if one believed or not in the existence of a vital force in their formation) and they held that a study of these radicals was necessary to achieve 'a natural classification of organic compounds'. Berzelius only demanded that the radicals should not transgress his earlier electrochemical theories: Liebig hunted down radicals as one might run after the eland in uncharted Africa, to see what the beast was made of and how many more

of its kind there were; Dumas for a time was with Liebig, but later suggested a law of substitution, that an atom of hydrogen in an organic substance could be replaced by an atom of chlorine or oxygen without any change of type. To Berzelius this was utter heresy. The organic radical was electropositive and chlorine was eminently electronegative. As well might Dumas stick the horns of the eland on a leopard and claim to have still a species of antelope. And so they corresponded and argued with each other, but always in a firm and gentlemanly manner. When their theories approached there were congratulations, when they diverged there were sighs. But woe betide any speculative chemist who strayed into their field. Their comment, all in the name of science, was very bitter, very personal. Poor Laurent was severely savaged, so much so that his friend Gerhardt could say, 'I was an impostor, the associate of a brigand – and all this for an atom of chlorine put in place of an atom of hydrogen!'

Confusion was made more confused by the lack of unanimity with regard to the exact meanings of the terms atom, molecule, and equivalent, which, as used, were quite interchangeable, thus revealing a part conscious disbelief in the existence of an ultimate particle. There existed more than one list of atomic weights; one followed one's fancy. The very symbols varied. For acetic acid alone nineteen different formulae had been proposed. The chemists of that period paid a price of sore bewilderment for neglecting the hypothesis of the Turinese physicist, Amedeo Avogadro. Well might Wöhler complain, 'Organic chemistry is enough to drive one crazy. It is like a primeval tropical forest, full of the most remarkable things, a monstrous and boundless thicket with no way of escape, and into which one may well dread to enter.'

By the eighteen-fifties Berzelius had departed for the Elysian Fields, where no doubt he was vigilant in combating any revolutionaries who doubted his electrochemical theories; Liebig had withdrawn to a reflective life in Munich; and the ambitious Dumas had become a politician of the Second Empire, his life in laboratories almost behind him. The views of Berzelius and

Liebig were inherited by Hermann Kolbe, who developed the theory of the compound radicals with remarkable insight and who also challenged his rivals in the same polemical spirit as his illustrious predecessors. He had much to challenge, for unquestionably by then the leading theory belonged to the Paris school as put forward by Charles Gerhardt in his Theory of Types which sought to show that organic compounds conformed to certain inorganic types. There was a water type, an ammonia type, a hydrogen type, and a hydrochloric acid type. Into the moulds of these four types were squeezed all known organic compounds. Its four-square nature is rather reminiscent of the belief of the ancients in the four elements of fire, air, earth, and water. It was not enough that compounds were; as with the flowers and fossils of the nineteenth century they had to be like. The formulae were easy to write as the following simple examples may show.

Water Type

Ethyl alcohol $O\begin{cases}C_2H_5\\H\end{cases}$ Acetic acid $O\begin{cases}C_2H_3O\\H\end{cases}$

Ammonia Type

Benzamide $N\begin{cases}C_7H_5O\\H\\H\end{cases}$ Triethylamine $N\begin{cases}C_2H_5\\C_2H_5\\C_2H_5\end{cases}$

Hydrogen Type

Ethane (Ethyl hydride) $\begin{cases}C_2H_5\\H\end{cases}$ Butane (Di-ethyl) $\begin{cases}C_2H_5\\C_2H_5\end{cases}$

This theory of types, which now lies lightly in the dustbin beside so many of its forerunners, and which would be largely incomprehensible to a young chemist of today, had a considerable hold in its time. It was accepted with a measure of relief by many chemists who found in it a handy teaching method, and who no doubt detected in the fitting of organic compounds to inorganic models a final fading away of the unsatisfactory belief

in a mysterious but inexplicable vital force. There was also the commendation that the theory rested on the work of the newly arrived chemists, on Williamson, Hofmann, and Odling in London, and on Wurtz and his co-workers in Paris. In Germany Kolbe would have none of it; the 'typists' he said were simply playing with formulae. One of its supporters, however, was Kekulé, who had indeed been studying in Paris under Gerhardt while the theory was being formulated, and who had earlier worked with Williamson in London. He suggested mixed types, models which could, say, be part water and part hydrogen types. Then he rather spoiled the four-square and inorganic aspect by proposing a further type – a marsh gas type patterned on the formula CH_4. Deeper consideration of this type led Kekulé in 1858 to publish his famous paper on the chemical nature of carbon. Briefly put, he stated that carbon had a combining capacity of four, and that carbon atoms had the power of going together in such a way as to satisfy part of their combining capacity. Here indeed was a compass to guide the seeker through the 'boundless thicket' of organic chemistry. Many years later Kekulé explained how the theory originated in his mind. 'During my stay in London', he said, 'I resided for a considerable time in Clapham Road near the Common. However, I often spent my evenings with Hugo Muller at Islington, at the opposite end of the great town. We talked of many things, but most often of our beloved chemistry. One fine summer evening I was returning by the last omnibus, "outside", as usual, through the deserted streets which, at other times, are so full of life. I fell into a reverie, and lo, the atoms were gambolling before my eyes. Whenever, hitherto, these diminutive beings had appeared to me, they had always been in motion; but up to that time I had never been able to discern the nature of their motion. This time, however, I saw how, frequently, two smaller atoms united to form a pair; how a larger one embraced two smaller ones; how still larger ones kept hold of three or even four of the smaller; whilst the whole kept whirling in a giddy dance. I saw how the larger ones formed a chain dragging the smaller ones after them. The cry of the conductor, "Clapham Road", awakened me from

my dreaming; but I spent a part of the night in putting on paper at least sketches of these dream forms.'

And that, claimed Kekulé, was the origin of his Structure Theory. Although this delightfully imaginative tale has been quoted often, one might be pardoned for wondering if Kekulé was speaking with his tongue in his cheek, for his statement was made at a considerable celebration in his honour. Certain it is that in his paper of 1858 he hesitated to put down formulae showing chains of carbon atoms. Kekulé was concerned to develop the idea of the combining capacities of the elements (variously termed basicity, affinity, and atomicity, now valency) with reference to the compounds of carbon. By demonstrating clearly that carbon was 'four-atomic' and by assuming a 'carbon-skeleton' he was able to work out the atomicity of many of the radicals. Having done this he continued to write what few formulae he gave according to the Theory of Types. This is not surprising as the supporters of that theory held that formulae merely illustrated the changes undergone by a compound, how formed or how decomposed. The very title of Kekulé's paper is noteworthy – 'Uber die Konstitution und die Metamorphosen der chemischen Verbindungen und über die chemische Natur des Kohlenstoffs'. True, Williamson had said that formulae may be used 'as an actual image of what we rationally suppose the arrangement of the constituent atoms in a compound, as an orrery is an image of what we conclude to be the arrangement of our planetary system'. But this was quite contrary to the views of the other adherents. Wurtz stated that 'by our rational formulae we do not have the pretension of representing the intimate constitution of compounds. These formulae represent changes only, that is, facts which can be demonstrated by experiment.'

Gerhardt was quite explicit on this point; for him formulae expressed possible double decompositions, and for this reason the same compound might be given more than one formula. He was bold enough to declare, 'Chemical formulae are not destined to represent the arrangement of the atoms.'

It was left to another to show not just that this could be done

but that it was wholly necessary. In fact at the moment of publication of Kekulé's paper the theory of molecular structure had been formulated by Archibald Scott Couper working in Paris; his thesis, 'On a New Chemical Theory' was actually in the hands of his professor. It had indeed been too long in the hands of the professor, Adolphe Wurtz, who had been dilatory in having it presented to the French Academy. When Couper read Kekulé's paper he was much disturbed. He felt chagrin at being forestalled, and anger against Wurtz for being the cause of it. At once he took Wurtz to task for the delay. There were words, and Wurtz, considering Couper to be insolent, expelled him from his laboratory. As soon as possible Couper had his theory presented to the Academy, and shortly afterwards had it published in full in both French and English. It is a classic paper, a brilliant feat of the human intellect. Couper suggested structural formulae for carbon compounds and he wrote them almost as we write them today, but forty years before the discovery of the electron and nearly seventy years before the electronic nature of the valency bond was understood.

For a scientist Couper's background was quite remarkable. At the time of his theory he was 27 years of age and had been studying chemistry for little more than two years. Educated at home by his mother, he had attended the Humanity classes of the University of Glasgow. Every summer during vacation he toured or studied in Germany, which country seemed to hold a compulsive fascination for him. In 1852 Couper switched to the University of Edinburgh where he read logic and metaphysics under Sir William Hamilton, famous for his advocacy of the Common Sense school of philosophy. This contact with Hamilton was all important for Couper, for the almost traditional Scottish doctrine of Common Sense with its stress on the inductive method of reasoning was peculiarly in line with scientific procedure. Further, of all the philosophers in Britain, Hamilton was sharply aware of the importance of the German philosophers, in particular of Kant and his critique of knowledge. It is significant that at the end of three years Couper transferred to Berlin to continue his studies, and it was there that, for some

unknown reason, he turned to chemistry. It is tempting to think that Couper, fresh from his contact with Hamilton and his study of Kant, may have been influenced by the current positivist philosophy of Comte, who taught that since the Absolute was unknowable, it was best for philosophers to direct their energies to appraising and systematizing the methods and findings of science. After a year spent in gathering analytical skills Couper left Berlin for Paris, where he put himself under the direction of Adolphe Wurtz.

In contrast with Kekulé, who arrived at his conclusions as a direct development of the Theory of Types, Couper put forward his theory to replace the type theory, which, he regarded as scientifically unsound, even pernicious. Indeed Couper's paper must be approached as an essay in the philosophy of science. In it there are echoes of Descartes, of Reid and Hamilton. While historical comment frequently and properly accompanies the introduction to a new theory, Couper's opening sentences must be unique in a paper on chemistry. He begins, 'The end of chemistry is its theory. The guide in chemical research is a theory. It is therefore of the greatest importance to ascertain whether the theories at present adopted by chemists are adequate to the explanation of chemical phenomena, or are, at least, based upon the true principles which ought to regulate scientific research... There are two conditions which every sound theory must fulfil: (1) It must be proved to be empirically true; (2) It must no less be philosophically true.' And on purely metaphysical grounds, Couper proceeds to combat Gerhardt's Theory of Types. He makes a striking analogy. 'Suppose that someone', he says, 'were to systematize the formation of letters into words that formed the contents of a book. Were he to begin by saying that he had discovered a certain word which would serve as a type, and from which by substitution and double decomposition all the others are to be derived – that he by this means not only could form new words but new books, and book *ad infinitum* – that this word also formed an admirable point of comparison with all the others – that in all this there were only a few difficulties, but these might be ingeniously overcome – he would

state certainly an empirical truth. At the same time, however, his method would, judged by the light of common sense, be an absurdity. But a principle which common sense brands with absurdity is philosophically false and a scientific blunder.' Of course Couper is dealing with Hamilton's common sense – not, as that erudite man would have said, a sound understanding applied to vulgar objects, but a universal principle of human knowledge understood through analytical reflection on experience itself.

Commenting on Gerhardt's views on formulae and constitution, Couper is particularly scathing. 'Look where Gerhardt's generalization of Williamson's generalization leads him, and legitimately too – a fact which his logical spirit clearly discerned. He is led not to explain bodies according to their composition and inherent properties, but to think it necessary to restrict chemical science to the arrangement of bodies according to their decomposition, and to deny the possibility of our comprehending their molecular constitution. Can such a view tend to the advancement of science? Would it not be only rational, in accepting this veto, to renounce chemical research altogether?' But Couper is not finished when he has dealt with the Theory of Types; he goes on to demolish at a stroke any lingering remnants of the radical theories by stating emphatically that science demands the strict adherence to the philsophical principle, which it dare not ignore, and without which research cannot advance a step, namely that a whole is simply a derivative of its parts.

Having thus strewn the field with theories run through with his philosophical sword, Couper raises his flag. As might have been expected of one trained in metaphysical rather than physical studies, he commends Descartes, with no conscious irony, to the chemists of Paris. 'The sure and invincible method of arriving at every truth which the mind is capable of discovering is always one and the same. It is that, namely of throwing away all generalization, of going back to first principles, and of letting the mind be guided by these alone. It is the same in common matters. It is the same in science.'

Acting on his own advice Couper advanced what he termed 'a more rational theory of chemical combinates'. His theory is firmly based on a consideration of the elements and of their chemical affinities, or, as we should call them today, their valencies. In this respect he makes a remarkable forecast. 'It is such a property as affinity that is required to form the base of a system. Nor would its suitableness for this purpose be affected by the discovery that the elements are themselves composite bodies, which view the chemist is perhaps not unwarranted to adopt. For in such a case the necessity would doubtless still be found to exist of adopting the principle of affinity, or something at least equivalent to it, as the basis of the explanation of chemical combinates.'

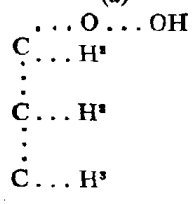
Although Couper does not mention the name of Frankland, who is generally recognized to have been the first to suggest the valencies of the elements, he makes no claim on his own behalf to be the originator of the idea, thus differing from Kekulé who, firmly but mistakenly, believed that it began with himself. What Couper stresses is that 'carbon enters into chemical union with carbon, and that in the most stable manner'. This property, with the combining power of carbon taken as four, and sometimes as two, suffices to explain for him all that is characteristic of organic chemistry. The theory is stated with arresting clarity and unusual boldness. Couper was manifestly miles ahead of the 'typists', and paradoxically this was his misfortune, for while they had obliged him by exhausting many sources of error, their minds were still held captive by those very errors. It would require not one paper but a campaign to free them. For instance, Kekulé in seeking to explain the Theory of Types by the potent idea of valency, was actually demolishing the theory, and yet he continued to write the formula of a compound after the type fashion, as representing its possible chemical reactions and no more. But Couper was *seeing* the molecule. He could talk of one carbon atom being situated between two others; he could refer to 'one side of the molecule'; and he could suggest the possibility of a molecule having two poles – with special reference to the molecule of an organic acid in which he

supposed the oxygen atoms to be electrically different, the one attached to the hydrogen atom being electropositive. This, it is made clear, is different from previous electrochemical theories (i.e. of Berzelius) which could never be applied throughout to organic chemistry. It is of no value to read into this a sign of the modern Theory of Resonance; it is of value to recognize how Couper was placing atoms around the molecule. Thus for the first time a chemist was able to write structural (or constitutional) formulae.

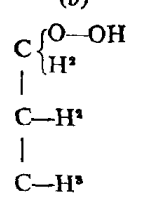
In his formulae Couper erroneously took the atomic weight of oxygen = 8 and as a consequence had to consider oxygen as having the same combining power as hydrogen or chlorine. In each formula, therefore, there are twice as many oxygen atoms as there should be. In the paper submitted to the *Philosophical Magazine*, Couper joined the atoms with dotted lines, whereas in *Annales de Chimie* he used straight lines and brackets, the latter to connect atoms of carbon with atoms of different elements. How advanced his formulae were over those of the type theory may be shown by the following examples where (*a*) is the English version, (*b*) the French version and (*c*) these versions modified by using straight lines throughout and applying oxygen = 16.

Propyl Alcohol

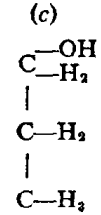
(*a*)



(*b*)

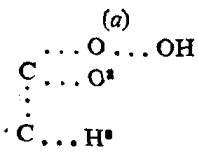


(*c*)

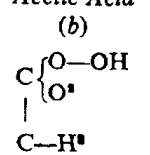


Acetic Acid

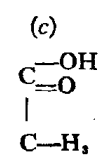
(*a*)



(*b*)

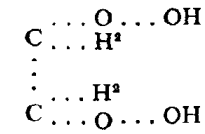


(*c*)

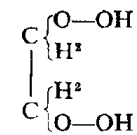


Glycol

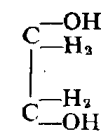
(*a*)



(*b*)

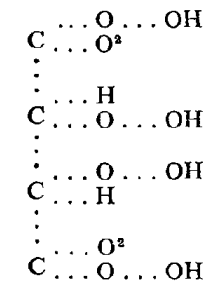


(*c*)

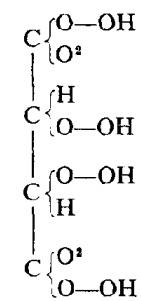


Tartaric Acid

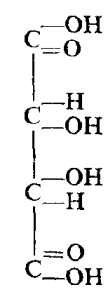
(*a*)



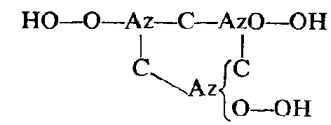
(*b*)



(*c*)

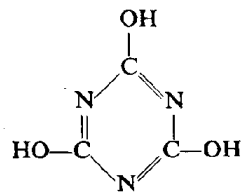


It will be seen that these formulae are really a halfway stage between constitutional and graphical formulae. By taking the oxygen atom to have the combining capacity shown, Couper avoided the necessity of inventing the double bond, but at the same time this forced him to propose an electric attraction between the two oxygen atoms of the C . . . O . . . OH group. He did not boggle at attempting to write the formula of glucose, nor even later of salicylic acid, but the most striking of all his suggestions occurs at the end of the French version where he says that the formula of cyanuric acid will be



In this formula he states that 'the atoms of carbon and nitrogen are linked by 2 units of affinity'. Thus Couper is brought to

make use of double links in this the first suggestion of a ring formula. The modern formula of cyanuric acid may be taken to be



Couper's error lies in taking the combining power of nitrogen to be 5. His readiness to write such formulae with little experimental work to back them up rendered them distasteful to such as Wurtz. Nevertheless, here was offered not a compass only but a chart to negotiate the boundless thicket of organic chemistry, and what a deal of exploration has been done since then. But it would be an error to imagine that the theory was welcomed. It was severely criticized. Wurtz, who may be remembered as the arch-example of how not to commence a book on the history of science – 'Chemistry is a French science. Its founder was Lavoisier of immortal memory' – reviewed the theories of Kekulé and Couper in a publication of his own formation. Of Couper's formulae he claimed that they were too far removed from experience, too arbitrary, too hypothetical; they presumed too much, and their author was at fault in presenting them *comme la loi and les prophètes*. Even a year later a commissioned paper could be read at the British Association of 1859 purporting to give the recent progress and present state of organic chemistry, and carry but one line mentioning Couper, to correct him on a point regarding oxygen in which he was in error. This point was taken from Couper's paper, but his theory received not the slightest mention.

Following his expulsion by Wurtz from the laboratory in Paris, Couper departed the scene. He made no attempt to answer his critics; he made no attempt to defend nor to extend his theory; he was heard of no more. One wonders if Darwin, having presented his thesis, had retired to his couch and had

not written the Origin of the Species, nor got himself a champion in T. H. Huxley, how much his name would be heard today. Couper had withdrawn; the field went to Kekulé, than whom there was no more industrious chemist on the Continent. Within a year he began publication of his important *Text-book of Organic Chemistry* which served as a platform for his theoretical views. His attitude to formulae had changed so far as to allow him to attempt a representation of the structure of simple compounds. Thus, although the corpse was carried around for a little while longer, the Theory of Types was dead. In the volume published in the early part of 1859 the first of these formulae are encountered. Though Kekulé must have been influenced by Couper, whose work he had read, this is not apparent from the first sight of them, of which the following are typical examples:

Acetic Acid

Ethyl Alcohol

Much nonsense has been written about Kekulé's early training as an architect (a profession which he pursued for a short period before turning to chemistry) having influenced him in his approach to the molecule and its structure. The representations he suggested have more relation to the German table than to any architectural structure, and it was inevitable that Kolbe should damn them as 'roll and sausage' formulae. Regrettable as it may be to those who seek such influences it was the student of philosophy who showed the sound grasp of form and structure – it may also be regrettable to others that he should be a student of common-sense philosophy. In fact, two or three years were to pass before chemists generally could bring themselves to apply Couper's structural formulae, and then there were several claimants to have the honour of being responsible for them. Of all people Wurtz, who would seem to have had a faulty memory, could claim that these formulae were used for the first time in his college lectures in 1863. Kekulé, by his inspired

suggestion of a ring formulae for benzene, gave the thebry a far-reaching extension. The era of synthesis had begun. Few scientists have been honoured in their time as was Kekulé. He became professor at the University of Bonn, was appointed tutor to the Crown Prince of Germany, the future Kaiser Wilhelm, and was raised to the rank of baron; he was even fêted by his scientific colleagues. At a celebration in his honour in 1890 he could declare with modesty, 'Some ideas at some time lie in the air; if one does not give expression to them another will soon do so.'

But Couper's name had almost disappeared. Text-books are inclined to beget text-books of their own kind where errors continue to be perpetuated and where omissions continue to be omitted. It is true that reference, usually brief, to Couper's paper could be found in the books of the historians of chemistry, but few active scientists ever read such books, unless circumstances force them. One who was forced to turn back was Richard Anschutz, successor to Kekulé in the chair of Chemistry in the University of Bonn. As a labour of love Anschutz had chosen to write a biography of his illustrious predecessor, and by an ironic twist of fate, in compiling his material, he browsed through the literature at the time of Kekulé's discoveries some fifty years before. He looked up Couper's paper, 'On a New Chemical Theory'. Anschutz was bowled over by the lucidity and the far-sightedness of the author. He could not conceive how anyone with such penetration of thought should have been heard of no more, nor could he rest content till he had found out what had happened to Couper. Over some years he sent out inquiries to Paris, to London, and finally to Edinburgh, where he enlisted the support of Professor Crum Brown. The registers of the University of Edinburgh were combed with no success as to the whereabouts of Couper. It was a chance remark by Sir James Dewar which furnished a successful clue.

'Couper was long before my time,' said Dewar. 'It's like a dream to me, as if I had been told that he had to be put in an asylum.' Crum Brown applied to the Board of Lunacy for Scotland and found their records were better than those of the university. Archibald Scott Couper had been a patient in a mental

institution during the early summer of 1859. He was discharged, readmitted, and finally discharged at the end of 1862. Thereafter he lived for many years in the care of his mother in his native place of Kirkintilloch, a small town near Glasgow, where he was a familiar figure in the streets walking about accompanied by an attendant. It is a pathetic picture. Priority may be of little consequence to the progress of science, but it is of consuming interest to the scientists involved, deny it as they may; research is yet a human activity. Nor is a consideration of priority lacking in value for those who come after, for, in studying first one side, then the other, one is forced to see the problem at issue in terms of the workers involved, and not in terms of those who know the whole answer. And then, as W. P. D.Wightman has pointed out in a recent article on the origins of the Periodic Law, 'in the meticulous search and analysis which such an objective entails the whole question often appears in a different light, revealing new perspectives of the greatest historical and philosophical interest'.

The theory of the chemical nature of carbon and of the molecular structure of its compounds abounds in such perspectives. The concept of the atom was too simple; the problem was to recognize the molecule. Its recognition marks not just a revolution in the science of chemistry; it marks a revolution in man's thought – that the infinitesimal outside visual experience had form and structure that could be represented by formula almost as if it were number. Who will say, for instance, that this had no influence on the thought processes of Pasteur – a chemist and in Paris – as he evolved from about the year 1859 onwards his ideas of micro-organisms bringing about fermentation and disease? Or that Clerk Maxwell, in the same year, as he applied his statistical and dynamical methods to the justification of the molecular conception of gases, was not aware of the import of the theory. Indeed the theory, firmly based on valency, was a challenge to physics, a challenge that has not yet run out.

These two young men, Kekulé and Couper, were fated to be the agents of an illuminating theory, which was not merely to be of transitory interest, but which, by its extension, became

Organic Chemistry itself, and more besides. Well might Couper, with his reliance on philosophical principles, have written, 'The end of chemistry is its theory.'

REFERENCES

Kendall, J.: *Young Chemists and Great Discoveries* (London, Bell, 1939).

Partington, J. R.: *A Short History of Chemistry* (London, Macmillan, 1948).

Findlay, A.: *A Hundred Years of Chemistry* (London, Duckworth, 1937).

Meyer, Ernest von: *A History of Chemistry* (London, Macmillan, 1898).

Anschutz, R.: 'Life and Chemical Work of Archibald Scott Couper', *Proc. Royal Soc. Edinburgh*, 1909 **29** 193.

Kekulé, F. A.: 'Über die Konstitution und die Metamorphosen der chemischen Verbindungen und über die chemische Natur des Kohlenstoffs', *Annalen der Chemie*, 1858 **106** 129.
Also in Ostwald's Klassiker Nr. 145.

Couper, A. S.: 'On a New Chemical Theory . . .', *Phil. Mag.*, 1858, 4th series, **16** 104.
'Sur Une Nouvelle Theorie Chimique . . .', *Annales de chimie*, 1858 **53** (3) 469.
Also in Alembic Club Reprints No. 21.

Japp, F. R.: Kekulé Memorial Lecture, *Journal Chem. Soc.*, 1898, 73 97.

THE MICROSOMAL PARTICLE: THE TYPE-SETTER OF THE CELL

J. M. BARRY

IN the last century biologists discovered that living cells contain a nucleus surrounded by a viscous fluid which was called the cytoplasm; and they surmised that the cytoplasm possessed an intricate structure which lay beyond the power of their microscopes to resolve. Recently this intricate structure has been revealed by the beautiful electron micrographs of G. E. Palade of the Rockefeller Institute for Medical Research in New York (Inset 13). In these pictures are usually seen large sausage-shaped particles called mitochondria, and also what appear to be layers of membranes called the endoplasmic reticulum; scattered over this reticulum are tiny dark particles.

Some years before the electron microscope was invented scientists could see that the cytoplasm contains particles, and they tried to isolate these particles by breaking up cells in salt solution and centrifuging the suspensions at various speeds; but no clear-cut separation could be obtained. Then, about ten years ago, it was discovered that the reason for this failure was that many of the particles burst when they entered the salt solution, and that this bursting could be prevented by breaking up the cells in a solution of cane sugar of high osmotic pressure. If animal cells 'homogenized' in sugar solution are slowly centrifuged, the cell nuclei and any unbroken cells first separate. If the clear fluid above the sediment is then centrifuged at a higher speed a sediment separates which the electron microscope shows to be a collection of mitochondria; and if the clear supernatant is now centrifuged at a very high speed, to give some 100,000 times the force of gravity, a precipitate separates which consists of small particles that have been called microsomes.

Do the two types of particle we isolate by centrifuging exist as such in the cytoplasm? The mitochondria which are isolated look very similar to those seen within the cell in Inset 13 and are certainly little changed by isolation. They contain the enzymes which are used to oxidize acetic acid to carbon dioxide and water, which is the principal way in which the cell gets energy. The mitochondria are therefore tiny power units and, as would be expected, they are found in large numbers in cells that need large amounts of energy, such as the muscle cells. The microsomes however do not exist as such in the cell, but are fragments of the endoplasmic reticulum – when the cell is ground up the reticulum breaks into thousands of small microsomes. We see in Inset 13 that the endoplasmic reticulum is covered with tiny particles; each microsome which is isolated also has these tiny particles attached. These particles, which I shall call microsomal particles, appear to be the more fundamental part of the microsome, and it is these that will be discussed in this article. Recent evidence suggests that cytoplasmic proteins (and hence cytoplasmic enzymes) are built up from amino acids on the surface of these particles; and that they are the messengers which carry genetic information from the nucleus to the cytoplasm.

Before describing the evidence that proteins are built up on the surface of these particles a brief word on the structure of proteins will be given. Proteins are organic compounds of very high molecular weight; but in spite of their high molecular weight and their vital function in living cells, there is nothing magical about them, and the great discoveries about the structure of proteins have been made by a few men, like F. Sanger of Cambridge, who were not afraid to consider them as organic compounds. Like other organic compounds, proteins are crystalline when pure, and a pure protein like any other pure compound is a collection of molecules of identical structure. When a protein is broken down with acid the only products are a mixture of about twenty simple amino acids. Protein molecules have been proved to be long chains of these amino acids linked together in an apparently jumbled order; but in a pure protein the jumble in one molecule is identical with that in the next. Protein molecules can be compared to very long words made from an alphabet of twenty letters. (Some protein molecules have two or more amino acid chains cross-linked to one another and these molecules can be compared to hyphenated words.) A gram of a pure protein corresponds to a single word repeated many millions of times. Almost all the chemical reactions of a cell are catalysed by enzymes and every enzyme that has been isolated is a protein. Different enzymes have different properties because their amino acids are arranged in different sequences – they are like words with different spelling.

Why now do we believe that most proteins are built up from amino acids on microsomal particles? The first discovery that led towards this conclusion was made by H. Borsook and his colleagues at the California Institute of Technology in 1950. They knew that if an amino acid containing radioactive carbon atoms is injected into an animal, and the animal is killed after a few days, all the proteins of its body will be found to contain greater or smaller amounts of the injected amino acid, which can, of course, be detected by its radioactivity. Borsook realized that the way to find at which point in the body cells amino acids are first incorporated into proteins was to inject a radioactive amino acid into an animal and kill it, not a few days, but a few minutes after the injection; after this short time the only proteins to contain large amounts of the radioactive amino acid would probably be those at the site of protein formation, because the new molecules would not have had time to become distributed throughout the cell. He and his colleagues therefore did the following experiment. They dissolved in a little water a few crystals of three amino acids which had been made in the laboratory to contain radioactive carbon-14, and injected the solution into a guinea-pig. Thirty minutes later they killed the guinea-pig, rapidly dissected out its liver, and homogenized it in sugar solution in a Waring blendor (the machine that has now been adopted by coffee-bars for making milk shakes). They then centrifuged the homogenate at three different speeds and obtained the liver cell nuclei, the mitochondria, the microsomes, and the supernatant liquid. In each of these fractions they

measured the protein content per gram (which is simply done by a nitrogen analysis), and also the intensity of radioactivity per gram (which is also simply done by placing a known weight of material in a standard position under the window of a Geiger counter). From these measurements they calculated the proportion of the protein in each fraction which had been replaced by newly made radioactive protein. The proportion was twice as high in the microsomes as in any other fraction. Similar experiments were repeated by Borsook and other people and the result was always the same: although a long time after injecting a radioactive amino acid into an animal the protein of the microsomes was not the most radioactive protein of the cell, a few minutes after the injection it always was. The microsomes therefore seemed to be the point in the cell at which amino acids are converted to proteins.

At this time it was not known that a single microsome does not exist as a particle in the cell but is a fragment of the endoplasmic reticulum; nor was it known that a single microsome has scattered over it a number of smaller microsomal particles. Since 1950 many biochemists have followed the lead given by Borsook, and the most important discoveries have been made by P.C. Zamecnik and M.B. Hoagland and their colleagues at Boston. They discovered, soon after Borsook did his experiment, that if microsomes from the liver of a rat are washed with the detergent sodium deoxycholate, and the suspension is centrifuged at 100,000 times gravity, instead of the original microsomes separating, a sediment forms which the electron microscope showed to be tiny particles of uniform size (Inset 14). It was clear that these particles had formed part of the large microsomes that had disintegrated. At about the same time Palade demonstrated that the intact cell contains an endoplasmic reticulum on which are scattered tiny particles (Inset 13) and it was soon shown that these particles were the same as those which the detergent releases from the microsomes.

Zamecnik's group found that the microsomal particles contained about 50 per cent protein and the question then arose: are proteins built up from amino acids on these microsomal particles or on the remainder of the microsome (that is, on the remainder of the endoplasmic reticulum)? They answered this question by the following series of experiments. In each experiment they anaesthetized a rat with ether, opened its abdomen and exposed its liver. They then injected a solution of an amino acid, which contained radioactive carbon, into the rat's tail vein. Then, between 2 and 20 minutes after the injection, they rapidly removed part of the liver and plunged it into ice-water. They then ground it up in sugar solution and separated the microsomes by centrifuging; these they then washed with deoxycholate solution and got the microsomal particles. They found that when the liver was removed 2 minutes after the injection the microsomal particles contained up to seven times more radioactivity in every milligramme of protein than the remainder of the microsome; this difference in radioactivity persisted for several minutes after the injection. They concluded that the radioactive amino acid was being built up into protein on the microsomal particle and not in the remainder of the microsome. Later experiments have fully supported this conclusion.

Zamecnik and his colleagues have also made another important discovery: that each amino acid undergoes a number of reactions in the solution of the cell before it is built up into protein on the surface of a microsomal particle. They have discovered these reactions by the classical approach of biochemistry: by breaking up animal cells and attempting to extract those components which aid the incorporation of an amino acid into a microsomal particle. The most important reaction they have demonstrated is that each amino acid, apparently catalysed by its own particular enzyme, reacts with adenosine triphosphate (ATP) to give an activated form of the amino acid. In this way ATP provides energy for protein synthesis (as it also provides energy for a muscle to contract, for a firefly to give light and for an electric eel to give a shock). However, there is not space here to describe precisely what is meant by the activation of an amino acid, or to describe these experiments on extracts of cells in detail.

What will rather be attempted is to show why the conclusion

that amino acids are converted to proteins on microsomal particles, if it is proved correct, is one of the most fundamental advances in biochemistry in recent years. That the conclusion is correct seems increasingly probable. Particles of a similar size and composition to the microsomal particles of liver have been isolated from a number of cells and have been shown to be able to incorporate amino acids into protein. Some cells, such as bacteria, contain no endoplasmic reticulum, but they still contain the tiny particles. However, the final proof will be the demonstration that particles isolated from a particular tissue will manufacture a protein that is specific to that tissue; at the time of writing this has not been done.

Let us assume that each microsomal particle is responsible for the synthesis of only one type of protein – and the particles are in fact so small that this assumption is probably correct. We believe that twenty kinds of activated amino acid molecules are presented to the particles at random, while off each one comes a protein chain with the same amino acids always in the same place in the chain. That is, following on the previous analogy, each microsomal particle is presented at random with twenty letters and turns them into words of the correct spelling. It is for this reason that themicrosomal particles were called the type-setters of the cell. There are two possible mechanisms by which each microsomal particle could have this property of a type-setter. The first is that each particle is animated by a mind which causes the selection of amino acids and their arrangement in the correct order; but biochemists have never yet needed to resort to this non-mechanical type of explanation. The only alternative mechanism is that the structure of each particle is such that it directs the different amino acids to link together in the correct sequence; and if the structure of one particle does in fact differ from that of the next, then either one particle must contain molecules of different structure to the next, or it must contain molecules of the same structure arranged in different patterns. Because the particles are so small the difference between them almost certainly lies in the structure of the large molcules they contain. Because a microsomal particle can control the sequence of amino acids in a protein it can conveniently be said to contain 'information' about the structure of the protein. But it should be remembered that this is only distantly related to the information carried in a human mind; it more resembles the information carried in the death-mask of Oliver Cromwell about the structure of his face.

Now we know that the proteins of a child have exactly the same amino acid sequence as those of its father or its mother. The information about their structure must therefore be transmitted in the egg and sperm. From what has been said about the microsomal particles the reader might expect to find that one of each kind of these is carried in each egg and sperm, but this is not so. For one thing, sperms contain no ribonucleic acid (RNA) and microsomal particles consist half of RNA and half of protein. In fact, the information about protein structure is almost certainly transmitted from parent to offspring in the structure of the desoxyribonucleic acid (DNA) of the chromosomes. The chain of evidence for this belief is roughly as follows. The information is certainly carried in the nucleus of the germ cell: if the nucleus is removed from the egg of one species of sea-urchin, and the egg is then fertilized by the sperm of a second species, the fertilized egg develops into a sea-urchin of the second species. The information is certainly carried in the chromosomes of the nucleus: the movements of the chromosomes at cell division are exactly those which would be expected of the bearers of the hereditary information. The information is almost certainly carried in the DNA of the chromosomes and not in their other component which is protein: biochemists have transmitted information about protein structure from one strain of bacteria to another by extracting pure DNA from one strain and adding it to the other. It seems probable that the only precise information that is passed from parent to offspring is information about the amino acid sequences of thousands of proteins of the adult body, and that this information is carried in the structure of the DNA of each gene.

Therefore, if cytoplasmic proteins are built up on the surface of microsomal particles, it follows that when a fertilized egg

develops into an adult, the information carried in the DNA of the genes must be transferred to microsomal particles. Two great expanses of research seem to open out from this fact. Firstly, how is the information transferred from the DNA to the microsomal particles and, from them, to proteins? This transfer corresponds to the translation of a message in a code of four letters to one in a code of twenty. This is because each molecule of DNA is, like a protein, a chain made by combining simpler units (called nucleotides) in a jumbled order; but whereas the protein chain is made of twenty different units or 'letters', the DNA chain is made of only four. Since the information in DNA is carried in the shape of the DNA molecule, it is presumably carried in the spelling of its four-letter alphabet. Somehow each gene must control the formation of a microsomal particle in such a way that the spelling of the DNA of the gene results in a particular spelling in either the protein or RNA of the particle. There is some evidence that suggests that the information is carried in the RNA of the particle (RNA molecules like those of DNA correspond to words in a four-letter alphabet). But what the chemical reactions are by which the information in the gene is transferred to the particle and from it to the protein is quite unknown.

The second field of research that opens out seems even more fundamental; it is to discover how the ability of the nuclear DNA to transfer information to a microsomal particle, which in turn causes the formation of a cytoplasmic protein, is controlled. For if every gene of the nucleus sent off microsomal particles at top speed as messengers to cause the formation of cytoplasmic proteins, then, when an egg develops into an adult, no differences could arise between tissues, since every cell of the body would contain every possible protein. With a knowledge of how microsomal particles are regulated may come an understanding of the mechanism of development; also may come an explanation of the wild, uncontrolled formation of proteins in cancer.

STEREOPHONIC SOUND REPRODUCTION

J. MOIR

UNTIL fairly recently the majority of sound reproducer systems, whether radio, tape, disk or sound film, have employed a single transmission channel between the microphone in the studio and the listener in the home or theatre. This is a monaural or monophonic system. In the early stages of the art the 'naturalness' of the reproduction provided by such a system was limited by defects in the individual components, but normal development has gradually brought the best monaural sound channels up to the stage where little further improvement in performance is to be expected.

There is no evidence to suggest that any extension of the frequency range beyond the present limits of roughly 20–15,000 cycles, a reduction of the distortion components below 0·1 per cent, an increase of the signal to noise ratio about 60 dB, or any modification of the phase characteristics would produce a significant increase in the naturalness of reproduction. In spite of this semblance of finality there is also no doubt that a discerning listener could never be deceived into believing that even the best monaural reproduction was an original performance. In what respects do our systems fail?

Many of the answers to this question are still only seen rather dimly, but recent work has answered that of the reproduction when the system transmits an impression of the size of the sound source, and the position of the instruments on the stage. A monaural system provides no indication of the width of the orchestral stage though it does, in fact, give an impression of the depth* of the stage, but the image is distorted to the extent that

* It is commonly believed that the hearing system estimates depth, i.e., distance, by noting the ratio of the amount of reverberant sound to the amount of direct sound.

sounds originating at any point on a given radius from the microphone all appear to come from the same point, i.e., the virtual stage has depth but no width. In this respect the performance is unaltered by an increase in the number of microphones in the studio or loudspeakers in the home.

Sound reproducer systems that do give an impression of the width and depth, and hence the size of the studio stage, have been known since somewhere about 1880 and are termed 'stereophonic', a word derived from the Greek *stereos* meaning solid. In practice a stereophonic reproduction is found to have several probably unsuspected advantages over the best monaural performances. They may be listed as –

1. An improvement in the clarity of reproduction. Any monaural reproduction of the music of a large orchestra tends to be a jumble of sounds in which it is difficult to recognize the individual instruments or to separate the soloist from the orchestra.
2. The acoustic atmosphere of the concert hall is reproduced in the listening room. This is equivalent to adding the background to a portrait.
3. The majority of the acoustic difficulties that occur in music studios or concert halls are due to low frequency phenomena and are exaggerated by monaural reproduction. These difficulties are minimized when the reproducer system gives a faithful impression of the acoustic conditions in the studio or concert hall.
4. Speech intelligibility is improved, particularly when the speech is being delivered in the presence of other sounds.
5. A full symphony orchestra produces maximum loudness levels in the region of 105–110 phons and, appropriately used, these loudness levels are very effective in stimulating the senses. T. Somerville and others have shown that a monaural reproduction of the same music in a small room is disliked if the loudness exceeds about 85–90 phons, a loudness reduction of about five times. Stereophonic reproduction of the original sound leads to the acceptance and appreciation of loudness levels much nearer the original.
6. A full orchestra will have significant components over the frequency range up to at least 12–14 kc/s though no investigator has yet found an audience (uneducated in the pseudo hi-fi sense) that wishes to hear the full frequency range reproduced over a monaural system. There is evidence, though not yet complete proof, to suggest that stereophonic reproduction leads to an appreciation of a frequency range much nearer that of the original music.

A completely satisfying explanation for these improvements is not easy to find, for we are woefully ignorant of the relative importance of all the factors that contribute to a 'high quality' sound picture. The exposition that follows may therefore be incomplete.

The hearing system has the capacity to separate from a jumble of simultaneously presented sounds just those sounds to which it wishes to listen. This is a reflex action that is employed continuously and unwittingly in all our everyday activities, for without it, conversation in a noisy street or across the table in a busy restaurant would be impossible. Its effectiveness is indicated by the ease with which it is posssible to listen to the noise from one tappet in an engine having twelve such noisy tappets. A major contribution to this faculty is the hearing system's 'steerable directivity', which allows it to concentrate on sounds approaching the head from one particular direction to the partial exclusion of similar sounds approaching from other directions. The eyes have an analogous facility in that objects are seen in sharp focus only over a relatively narrow angle of a few degrees though they appreciate the surroundings over an azimuth angle of about 140 degrees. A significant movement at any point within this wide field immediately results in the eyes bringing that point into sharp focus.

The hearing system has similar characteristics though they are less well understood. Kock[1] has shown that our two-eared listening system has a discrimination of 12–16 dB in favour of wanted sounds and that this discrimination is dependent upon the difference in the direction of approach of the wanted and unwanted sounds. In practice a symphony orchestra will occupy a concert stage, 70–80 feet wide, with the orchestra subtending

an angle of about 60–70 degrees at a seating position about half way down the hall. When listening to a 12-inch loudspeaker at a distance of 10 feet in the average living room the same orchestral sounds will subtend an angle of less than 3–4 degrees, thus all the discrimination due to the spatial separation of the instruments in the studio is lost. This is thought to be the major part of the explanation of the increased clarity of a stereo reproduction.

The directional facility is also the explanation of the reduced trouble with the acoustics of the studio. Two-eared directional listening reduces the amount of generally reverberant sound by discriminating against those reflections that approach the head from the side and rear. Directional microphones have some of the same advantages but they do not duplicate the human hearing system in performance.

The improvement in the intelligibility of speech is also attributable to the directional properties of the ear. Noise, i.e., any unwanted sound, masks conversation by competing with the wanted signals for the attention of the hearing system and thus any directivity is of value in reducing the amplitude of the noise components.

The preference shown by listeners for restricted loudness and frequency ranges is also due to the loss of directionality introduced by the monaural reproducer. When listening to a large orchestra, one listens primarily to a small section of the ensemble only, the remainder of the instruments forming an accompaniment to the focus of attention. This focus wanders continuously over the stage as the conductor and the score varies the emphasis on the different groups of instruments. The subjectively judged loudness (though loudness is hardly the right term) is not that measured by an omnidirectional loudness level meter of the usual type but is a function of the acceptance angle of the hearing system. Any particular group of instruments can be reduced in effective intensity and modified in spectral energy distribution, by the simple and automatic process of listening away from them. Thus the hearing system changes both the effective loudness and the effective spectral

energy distribution of the sound source to that giving the most pleasing impression. Use of a single channel reproducer system destroys this facility for all the sounds approach the listener over a narrow angle from one direction, that of the loudspeaker.

It will be seen that the transmission of an impression of the size of the sound source is likely to offer a real opportunity of increasing the naturalness of reproduction, and it remains to be seen how it may be accomplished and what instrumental complication is involved. At this point it is worth emphasizing that the size of the reproduced sound image cannot be increased by increasing the size or number of loudspeakers. Though widely believed, and indeed a logical deduction from superficial thinking, this solution is fallacious.

The problem is best approached by considering just how the human hearing system is able to judge the size of a sound source and to solve the allied problem of determining the position of the source relative to the listener's head. As with so many other subjective judgements, the actual details of the process are not crystal clear and there is room for differences of opinion about the relative importance of the various clues to source position.

The high accuracy of location in azimuth achieved by the hearing system is not generally appreciated. Using speech as a signal, J. Moir and J. A. Leslie[2] found that the average angular error between the true and estimated position of a sound source is in the region of 1·2 degrees, a result that received a substantial measure of confirmation from the work of F. C. Klump[3] and Sandel who, using pure tones, found errors that were only slightly larger.

The possession of two ears spaced apart by the head is clearly necessary, for this enables the hearing system to sample the sound field at two points and by studying the differences in the sounds reaching the two ears, to come to a decision about the position of the source. This is an easy conclusion, but it is difficult to decide just what differences are of importance to the hearing system. Some of the suggestions will be examined though it is fair to say that experimental evidence of their relative importance is still lacking.

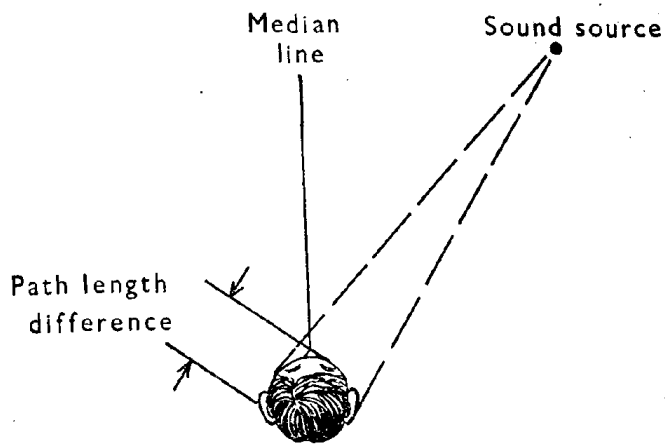


Fig. 1. Origin of the path length difference due to a sound source off the median line.

Sounds originating off the azimuth median line (Figure 1) will result in signals at the two ears which differ in many respects. The first obvious difference is in time of arrival, for the sound reaching the farther ear has farther to travel. The maximum time difference will occur when the sound source is out on either side of the head and along the line joining the two ears. Under these conditions any particular wave front will reach the far ear about 0·65 millisecond after it reaches the near ear, the difference corresponding to a path length round the head of roughly 22 cms. When the sound originates on the median line between the two ears the path to both ears is identical in length and there will be no time of arrival differences. Thus as a sound source moves round the head the time of arrival will differ by 0·65 millisecond with the source on one side, fall to zero with the source in the centre and rise to 0·65 millisecond again when the sound source is in line with the ears on the other side of the head, the signal at the nearer ear always leading in time. If the hearing system can detect small time differences then it is clear that information about the position of the source is available

in this form. Klump's work has revealed that inter-aural time differences of 6–12 μsecs can be detected, corresponding to an angular discrimination of about 1·25 degrees.

Sounds originating off the azimuthal axis will also result in loudness and intensity differences at the two ears, for the mass of the head lies between the ears, throwing a diffraction shadow across the far ear and producing pressure doubling at the near ear. The actual intensity difference will be a function of the signal frequency, for the pressure increase at the near ear and the depth of the shadow cast across the far ear, are functions of the diameter of the head measured in wavelengths. Calculations of the pressure difference due to diffraction can only be very approximate in the present stage of the art, but F. M. Weiner has produced some experimental data which is reproduced in Figure 2. From these it will be seen that the loudness difference

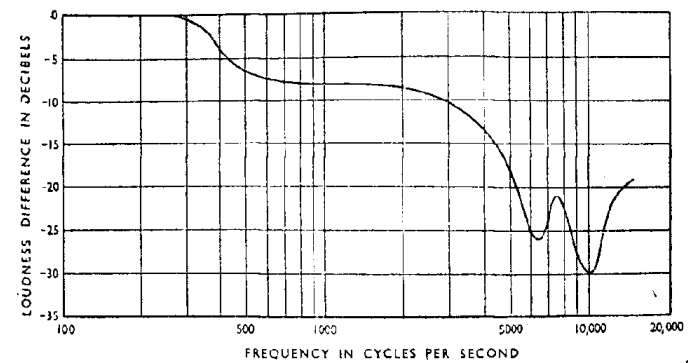


Fig. 2. Loudness difference produced in the right ear when a pure tone source is moved from the right to the left of the observer.

rises from almost zero for signals of low frequency, to 25–30 dB when the signals have a frequency above 5 kc/s. For reasons that will emerge at a later stage, it is unlikely that loudness difference itself is the important factor, the initial amplitude difference of the acoustic wavefronts being much more closely related to the angular position of the source as it is less diluted

by the general reflections from the boundary surfaces. Moir has shown that in a typical room the difference in the amplitudes of the initial wavefronts may be as much as three times the difference in loudness.

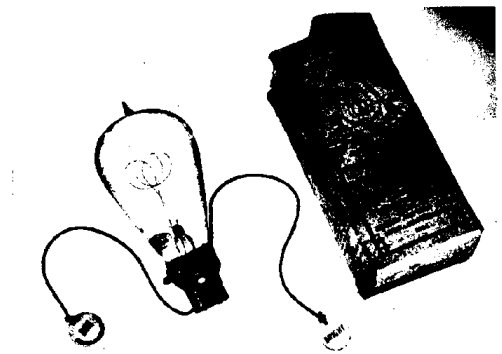
Both the time differences and the initial amplitude difference at the two ears are seen to be capable of providing satisfactory clues to source position and both are independent of the acoustic environment. It has been found that the accuracy of source location is not critically dependent upon the acoustics of the room, a fact that suggests that time or initial amplitude differences provide the primary clues to source position.

Complex sound waves such as speech and music will have their power spectra appreciably modified by the frequency dependent diffraction effects and this difference in the power spectra at the two ears may in itself provide an indication of the position of the sound source quite independent of the loudness difference produced. When trying to locate the position of a source emitting a steady rough note composed of a large number of harmonics, there is a strong subjective impression that spectrum balancing is in fact being employed, for the head is turned until the note has the same tonal character in both ears.

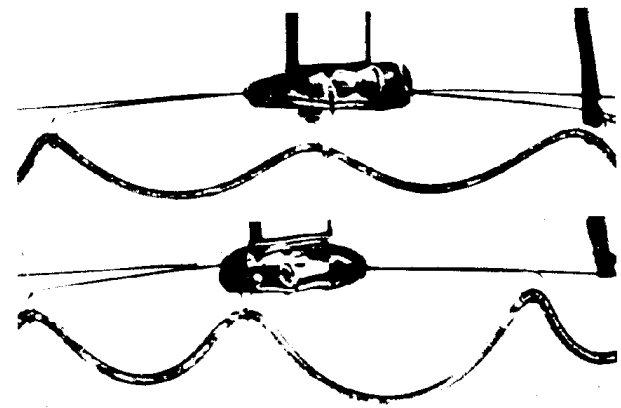
While all these differences, time, initial amplitude, loudness or power spectrum differences might be employed by the hearing system to locate the source position, it is difficult to ascertain just which of the differences is employed in practice. It seems likely that all are employed, the first approach being made on time difference, confirmed by initial amplitude difference, and finalized by equalizing the power spectrum in the two ears. Loudness difference, the commonly accepted explanation, would appear only to introduce confusion into the location process.

Whatever the means employed by the human hearing system there is little doubt that at least two separate signals must be transmitted from the studio to the listener. The simplest method of achieving this is to install two microphones in a dummy head, conveying the two signals by separate channels to two headphones worn by the listener. Apart from the natural objection to wearing headphones there is an unsuspected difficulty with

THE TUNGSTEN FILAMENT LAMP

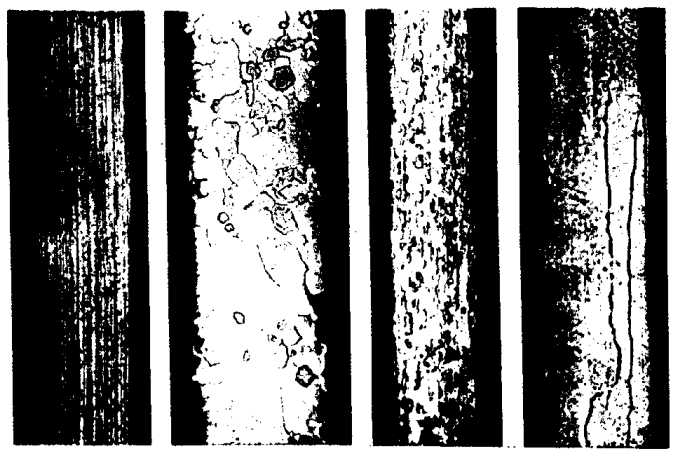


1. This novel design amongst the many forms of carbon lamp existing at the turn of the century was patented in the U.S.A. in 1903. Like a gas mantle it can be turned up and down in brightness by pulling the appropriate string!



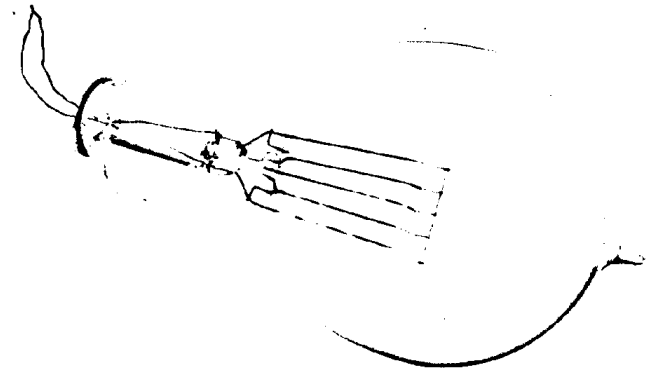
2. Two single coil filaments after 1,000 hours burning. The greater sag of the lower coil increases its length, resulting in increased heat loss to the gas filling. This lowers the luminous efficiency and light output of the lamp. The much improved stability of the top filament gives better lamp quality and makes possible lamp design modifications, giving further quality gains.

3. An 'off-set' in a tungsten wire after burning (magnification × c. 50), the result of crystal slip along a boundary extending across the wire diameter. This gives a brittle characteristic to the wire and was a common fault in early tungsten wires. Research into factors affecting crystal size and growth led to wires which were less brittle.

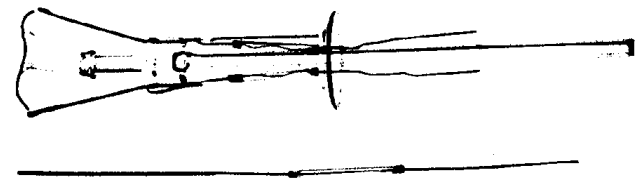


A B C D

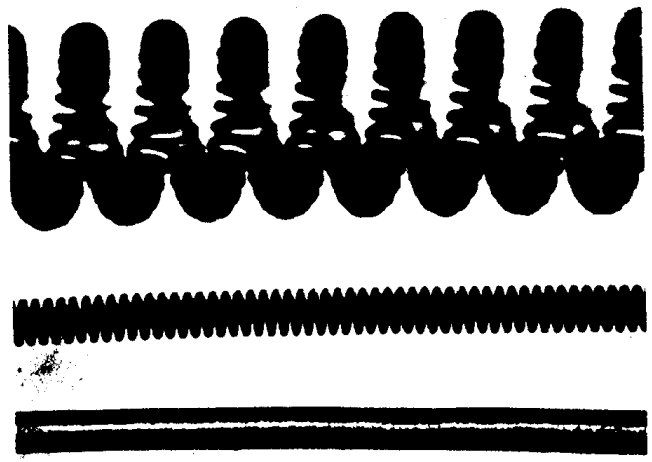
4. Microstructure of tungsten wires (magnification × c. 100).
(A) As drawn, before recrystallization. (B) Recrystallized commercially pure tungsten (0·05 per cent impurities). Large and small crystals with no orientation of growth. Possibility of crystal boundaries eventually extending across the wire diameter, leading to 'off-setting'. (C) Recrystallized thoriated tungsten (0·70 per cent thoria). Fine crystal structure giving a strong non-brittle wire, but with increased tendency to excessive sag. (D) Recrystallized alumina doped tungsten. Large overlapping crystal structure, showing long crystal boundaries parallel to wire axis. This structure gives strong wire with excellent 'non-sag' properties.



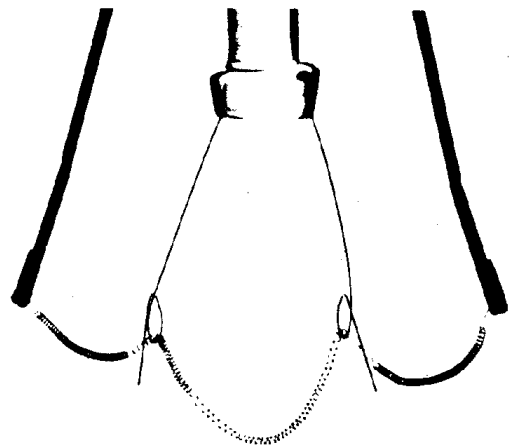
5. One of the first coiled-coil filament lamps made at the General Electric Company Research Laboratories in 1924.



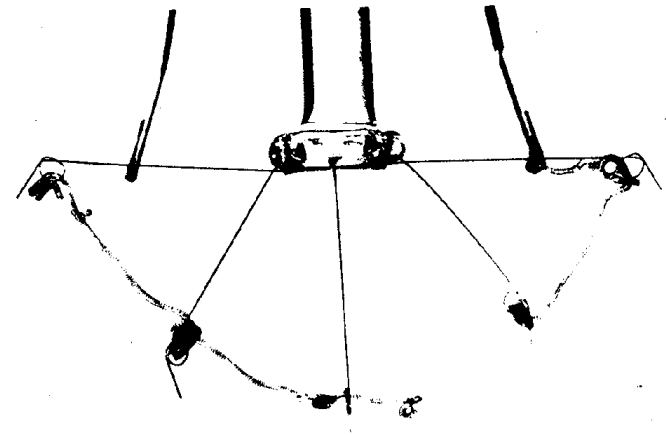
6. Top: Modern 'Osram' G.L.S. lamp filament mount, showing a glass-sheathed fuse in each lead.
Bottom: Lead wire, showing glass-sheathed constantan fuse and 4-part construction.
These fuses within the 'Osram' lamp prevent blown house fuses, or bursting lamps possibly resulting from lamp failure, and afford complete protection to the user.



7. 240v. 40W. coiled-coil and 240v. 15W. single-coil filaments compared to a human hair (magnification × *c.* 75), illustrating the extreme fineness of the wire used in modern general service lamps.

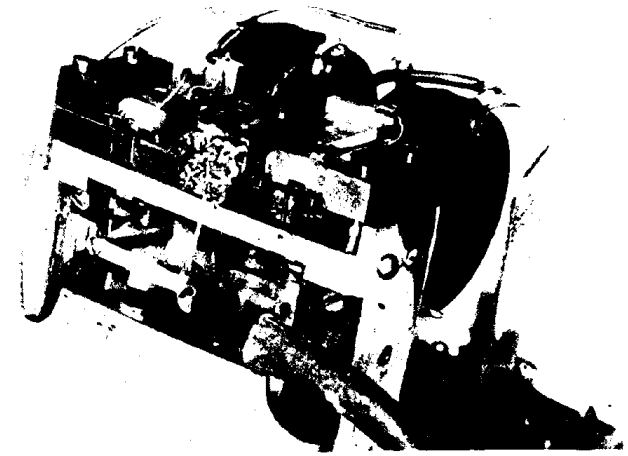


8. Sag and severe pitch distortion produced by deliberate contamination of the filament surface with a minute quantity of a metal salt.

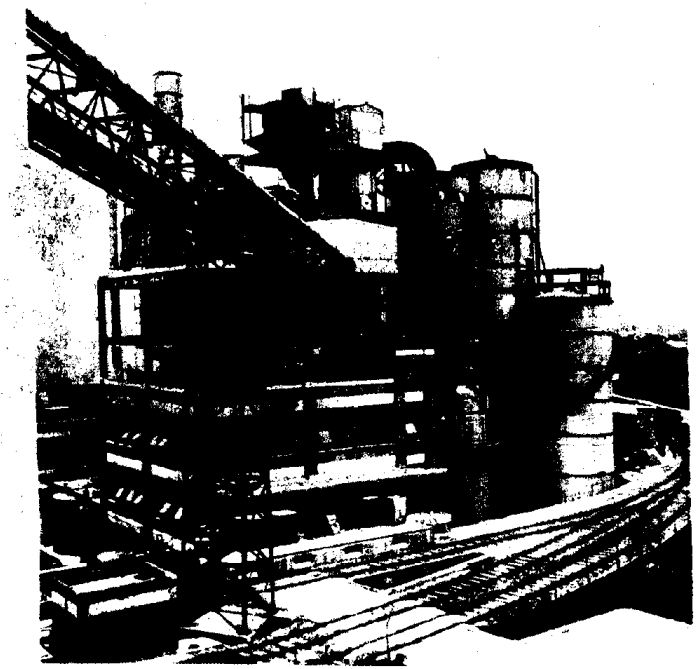


9. Artificially produced filament 'squirming' due to a very low partial pressure of oil vapour in the filling gas.

STEREOPHONIC REPRODUCTION



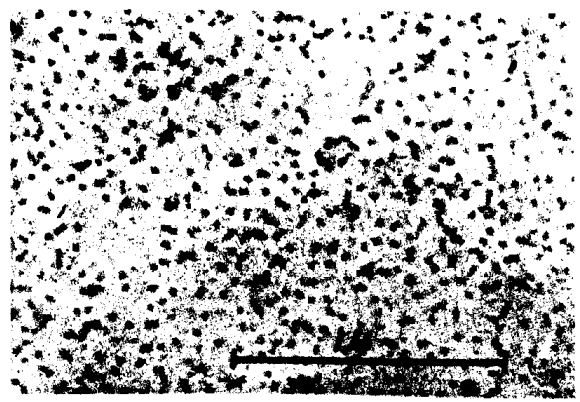
10. The original Blumlein stereophonic disc cutter. Two magnetic assemblies drive the chisel-shaped stylus which is free to move in two directions at right angles. *(Courtesy of E.M.I. Ltd.)*



11. External view of detergent-making plant at the Manchester factory of Thomas Hedley and Co. Ltd.

12. Mixing dry materials in different proportions during the manufacture of detergents. (*Courtesy of Thomas Hedley and Co. Ltd.*)

13. Electron micrograph of a section of a pancreatic exocrine cell of a guinea-pig. On the left is part of the nucleus and on the right parts of two mitochondria. Between them lies the endoplasmic reticulum on which can just be seen the dark microsomal particles. (× c. 30,000.) (*Courtesy of Dr G. E. Palade, Rockefeller Institute of Medical Research, New York.*)



14. Electron micrograph of deoxycholate-insoluble particles from microsome fraction of rat's liver. These particles are thought to be similar to those on the endoplasmic reticulum in Plate 13. (× c. 40,000.) (*Courtesy of Dr P. C. Zamecnik, Massachusetts General Hospital, Boston, Mass.*)

this elementary technique. It seems highly likely that localization in the vertical plane is largely due to the change in signal differences at the two ears as the head is turned towards the source. This information is lost if head phones are worn, for both sound source and room then appear to move with the listener's head, a rather eerie effect.

If the listener cannot be transported to the room by wearing headphones it may prove possible to transport the room to the listener. Technically the simplest, but economically the most impossible scheme, is to erect a screen of several hundred microphones across the studio at a distance of 30–50 feet from the orchestra, coupling each microphone through an amplifier to a speaker similarly placed in a screen of several hundred loudspeakers across the reproducing room.

Though there is no doubt that this is technically satisfactory, it is economically impossible, so the next step is to reduce the number of separate microphones and loudspeakers in the curtain. Remembering that localization in the vertical plane is necessarily poor (our ears are spaced apart in a *horizontal* plane) the first step is to remove all except the bottom row of microphones and speakers. When this is done it is found that there is a surprisingly small loss in the stereo effect suggesting that the number of speakers in the remaining rows might well be reduced. In fact it is found that excellent stereo effects can be obtained with one row of five microphones and loudspeakers, very good results are possible with three speakers, and acceptable results can even be secured in a room of domestic dimensions from two speakers.

Five separate channels between studio and listener appear to be economically justifiable in large halls and are now being used in the 70 mm sound film installation such as Todd-A.O. Three channels are certainly satisfactory in almost all cinemas and may eventually be justifiable in domestic reproducer systems, while two channels can be made to produce a very significant improvement over a single channel in domestic surroundings.

That two channels can give such an acceptable performance remains somewhat of a mystery for the sounds from both loud-

speakers reach both ears and thus might be expected to produce utter confusion rather than stereophonic effects. However, H. A. M. Clark, G. F. Dutton, and P. B. Vanderlyn have shown mathematically that using two microphones and two speakers the amplitude difference between the resultant-signal-at-the-right-ear-from-left-and-right speakers and the resultant-signal-at-the-left-ear-from-the-left-and-right-speakers, is proportional to the phase difference resulting from a sound source off the microphone axis. Thus they claim that by appropriate choice of microphones and circuit, the E.M.I. two-channel 'stereosonic' system produces a true stereophonic sound image.

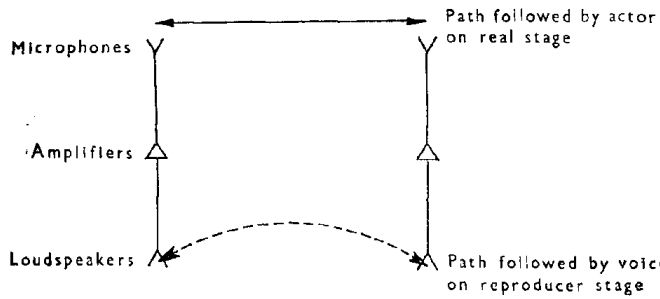


Fig. 3. Concave path followed by the virtual position of a voice in a two-channel stereophonic reproducer system.

In practice the main penalty for using only two channels is that the centre line of the virtual stage appears to be concave as in Figure 3 and the fraction of the total floor area over which stereo effects are produced is decreased to that suggested by Figure 4. Both effects can be minimized but not eliminated by appropriate microphone technique in the studio.

Microphone placement is the main practical problem with any two-channel stereo system and in consequence all the recording organizations are extremely reticent about their studio practice. Cardiod directional microphones are usually used but the spacing may vary from zero as in the E.M.I. Stereosonic system to ten or fifteen feet as used by some of the Ameri-

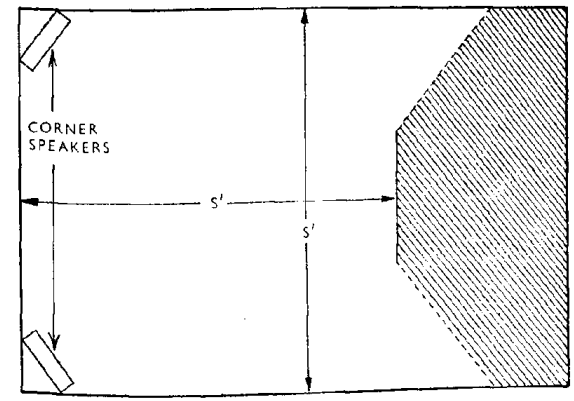


Fig. 4. Usable floor area associated with a two-channel stereophonic system. The front row of listeners should be no nearer to the loudspeakers than the distance between the speakers.

can studios for recording large orchestral works. None of the arrangements described are as satisfactory as a three-channel system but this should not be taken to mean that they are commercially inadequate for domestic use.

However many channels are used in a stereophonic reproducer system, it is essential that the signals in each channel be kept out of the other channels. Using 70 mm sound film there is no objection, other than the economic one, to providing six separate sound tracks, while four tracks can easily be accommodated on standard 35 mm sound film. Four tracks are now feasible on 0·25 inch magnetic tape, though two tracks are the present standard.

Stereo transmissions by radio are more difficult, though two separate channels may be obtained quite simply by using two transmitters, but this is probably an unacceptable use of frequency space. In the VHF and higher regions where frequency modulation is common practice, a second channel may be obtained by modulating a sub-carrier spaced 50–60 kc/s off the main carrier. A single carrier, on the other hand, may be

simultaneously amplitude and frequency modulated to give two separate channels though the cross talk margin is only barely satisfactory.

Gramophone records provide the major source of music in the home, but with these the addition of a second channel is a more difficult problem. A second channel may be accommodated either by using half of each surface for each channel and providing a double-headed pick-up for replay, or by recording each signal on one side of the record and playing both surfaces simultaneously. Neither system is really satisfactory.

As far back as about 1928 A. D. Blumlein, of what is now the E.M.I. organization, suggested that two separate signals might be recorded in a single groove, with one signal modulating the groove as a hill and dale recording, while the other signal produced lateral modulation and he successfully demonstrated the technique. As this is similar to the system that is being commercially exploited in the stereophonic records now being released to the public, the technique is worthy of a more detailed explanation.

Two separate moving coil or moving iron assemblies drive the chisel-shaped recording cutter which is free to move along two axes at right angles, one coil producing motion of the stylus in one direction, while the second coil produces motion along an axis at right angles. This is diagrammatically illustrated by Figure 5, while Inset 10 shows the original Blumlein cutter. In the ideal design there would be no interaction between the two drive systems.

It is interesting to study the resultant movement of the cutter when driven in two directions mutually at right angles. For convenience that channel carrying the signal that produces vertical movement of the cutter will be referred to as the vertical channel, and the channel carrying the signal that produces lateral motion will be called the lateral channel. A sinusoidal signal of maximum amplitude in the vertical channel with no signal in the lateral channel will produce sinusoidal movement of the cutter in a vertical direction, i.e., pure hill and dale recording. A signal of maximum amplitude in the horizontal channel with

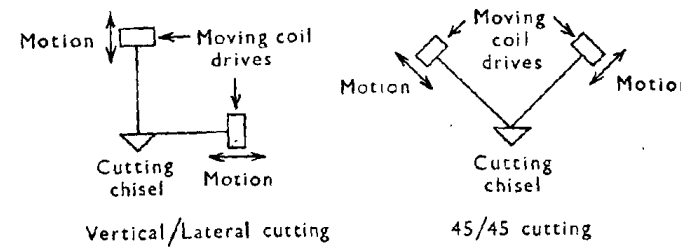


Fig. 5. Schematic arrangement of record cutters for stereophonic recording.

no signal in the vertical channel will result in the maximum lateral amplitude of the cutter tip.

Equal amplitude signals in both vertical and lateral channels will produce a resultant motion that is the vector sum of the motion due to each signal, and thus the cutter will have a sinusoidal motion along an axis at 45 degrees to the surface of the disk. If the two signals have the same frequency and are of equal amplitude but are not in phase, the cutter will perform an elliptical movement that becomes a circular motion when the signals are of equal amplitude but are 90 degrees out of phase. It is perhaps an elementary precaution to repeat that the cutter tip motion is at all times the vector sum of the motions that would result from the signal currents in each channel acting individually.

The pick-up used for replay must include two voltage generating elements each responsive to motion of the stylus along its own particular axis and ideally having no response to motion of the stylus along the orthogonal axis. In practice this is a difficult target that requires further work by pick-up designers to ensure its adequate achievement.

Satisfactory stereophonic reproduction demands that the signals in the two channels be equally treated (or ill treated) in their passage through the reproducer system, though in fact this cannot be ensured if hill and dale recording is used for one channel and lateral recording for the other. In a classic paper

published as far back as 1932 F. V. Hunt showed that harmonic distortion produced by a lateral recording was only about one-third of that produced by a hill and dale recording of the same peak to peak amplitude. In the commercial recordings now being released this basic difficulty has been side-stepped in a very neat manner, the axis of the system being rotated through 45 degrees to bring the direction of stylus movement of each system to an angle of 45 degrees to the surface of the record. Thus the performance of both channels is equalized by degrading the lateral channel and upgrading the vertical channel. This is the stereo recording technique known as the 45/45 system that has been adopted by all the major recording organizations in Europe and America, and it is a tribute to the genius of the late A. D. Blumlein that he described the technique in detail in his patent 394,325 (dated 1931) indicating that the actual idea had occurred somewhat earlier.

In most other respects the recording standards adopted for the new stereo disks are the same as those used for 33 r.p.m. LP records, but the stylus radius has been reduced to 12·5–15 microns and the maximum radius of the bottom of the groove to 5 microns. Any stereo pick-up should give a better performance on existing LP records than its monaural counterpart. A good monaural pick-up having adequate vertical compliance will produce a satisfactory monaural reproduction from a stereo record and produce no undue increase in wear on the records provided a stylus having a tip radius of 0·7 thou. or less is used. While stereo pick-ups remains expensive and stereo records form a minor part of a user's collection, the economic approach suggests that separate pick-ups be used for stereo and monaural records.

Gramophone records providing stereo reproduction are now being released in increasing numbers by all the leading recording companies in England and America. Most of the recordings provide a satisfactory standard of performance, but many of the American recordings give the impression that the orchestra has been divided into two well separated sections with a microphone for each. On replay each loudspeaker reproduces the

music played by half the orchestra, but there is little or no virtual signal from the space between the speakers. This is not stereophony. A good stereo reproduction is characterized by the small amount of sound that appears to come from the visible speakers and the even distribution of sound in the space between the two speakers.

These are no doubt only teething troubles that will disappear with growing experience, for the best of the stereo recordings are very satisfactory.

REFERENCES

1 Kock, W. E.: 'Binoral Localization and Masking', *Journal of the Acoustical Society*, 1950 **22** 801.

2 Moir, J., and Leslie, J. A.: 'The Stereophonic Reproduction of Speech and Music', *Journal of the British Institution of Radio Engineers*, 1952 **12** 360.

3 Klump, F. C., and Eady, H.R.: 'Measurements of Interaural Time Differences', *Journal of the Acoustical Society*, 1956 **28** 859.

THE TUNGSTEN FILAMENT LAMP

J. A. MOORE

AFTER World War II it was a popular belief that the tungsten filament lamp would be superseded by the hot cathode fluorescent tube, having four or five times the efficiency and at least five times the life. In fact, this has not occurred, and considerations of compactness, ease of handling, and simplicity of operation of the tungsten lamp evidently have outweighed to some extent the merits of high efficiency, low surface brightness, choice of colour, and long life, obtainable with the fluorescent tube.

Indeed, the number of tungsten filament lamps in use has increased considerably. This increase is partly due to a more imaginative approach to lighting design by lighting engineers and architects, and partly to a growing appreciation of the value of good lighting on the part of users. In addition, there is a rapidly mounting demand for tungsten projector lamps used in colour transparency projectors, and for very high wattage lamps in film and television studios, where the broad spectral band of incandescent radiation is essential for colour work.

These facts underline the continued importance of the tungsten filament lamp in the lamp industry, and the tungsten lamp 'family tree' has grown rapidly until there are now some 2,000 types, ranging from the surgeon's tiny 'grain of wheat' lamp to the 20 kW. studio lamp.

The incandescent filament lamp may be said to have been the forerunner of the scientific revolution that has taken place in the last fifty years. The discovery of electron emission from heated tungsten and carbon filaments, and the work of Edison, Fleming, J. J. Thomson, and Richardson, arose from work associated with the development of incandescent vacuum electric lamps at the turn of the century. This led to the radio valve

and the birth of the radio and electronics industry. Since the lamp industry was established by the time the radio valve was developed, there was much valuable experience available in component manufacture, pinch making, assembly, and pumping which enabled the valve industry to develop very rapidly. In fact, the radio valves used by the British forces during the First World War were produced in the factories of the large lamp manufacturers. The electrical supply industry also owes much of its early growth to the great demand for the electric carbon and tungsten lamps at the beginning of the century.

THE TUNGSTEN LAMP UP TO 1930

To summarize the early history of the electric filament lamp it is necessary to go as far back as 1840. At that time W. R. Grove, the British inventor of the cell named after him, was experimenting in his laboratory with a coiled platinum wire heated to incandescence by an electric current in nitrogen and other gases. Grove discovered the advantages of coiling to reduce heat loss and to improve luminous efficiency. He also discovered that nitrogen was the best gas to use since it had the least heat conductivity. Over fifty years later Langmuir continued this work in a comprehensive manner, which resulted in the commercial introduction of the gas-filled lamp.

In 1878 Swan exhibited electric lamps with carbon-rod filaments and a year later Edison made the first practical electric lamp consisting of an evacuated bulb with carbonized cotton thread as filament. Various methods of producing carbon filaments were then developed, and it was not until 1898 that the first metal filament was introduced. This was the osmium filament of Dr Welsbach of Germany. After the osmium filament came the metallized carbon filament marketed in 1905, and the tantalum filament in 1906. Tungsten, because of its high melting point (3655° K), low evaporation rate at high temperatures, and other characteristics, had long been recognized as a suitable material for lamp filaments. Its commercial use was delayed for several years because of the enormous metallurgical problems to be faced in drawing the long fine wire required for lamp

filaments. Squirted tungsten wire had been introduced in 1904, and in 1906 Dr Coolidge in the United States produced the first drawn tungsten wire which is the basis of the modern tungsten lamp. The Coolidge patent was the first to describe the production of ductile tungsten. This was achieved by the novel technique of reduction of the tungsten oxide to a fine metal powder, followed by compression into bars, sintering, swageing and wire drawing. These techniques formed the basis of the modern industry of powder metallurgy.

Such an advance was a major achievement, since it led to the production of a stronger wire which could be handled more readily, and made possible the shaping of the wire into coils of very small diameter. All previous tungsten filaments had proved to be far too fragile and brittle to wind into coils.

By 1909 the General Electric Co. had started production of 'Osram' lamps. The name 'Osram' is derived from the metals osmium and wolfram (tungsten), which at that time were the filament materials used in the lamps.

Filaments were made of long straight wires mounted under tension in the form of a cage and enclosed in an evacuated bulb. The operating temperature of the filament was limited by the high evaporation rate and eventual filament rupture due to the mounting tension.

Because of its higher melting point tungsten soon replaced osmium as a filament material. About this time Langmuir in the United States began his study of the effects on lamp efficiency and filament evaporation of coiling and gas filling, and in 1913 he patented the gas-filled lamp containing a coiled filament, upon which is based the tungsten lamp of today.

Langmuir found that the lighted filament of a gas-filled lamp is surrounded by a cylindrical layer of relatively stationary gas in fairly close proximity to the filament (Figure 1). A rapid convection stream occurs beyond this at some distance from the filament. The diameter of this 'Langmuir Layer' was found to be largely independent of the diameter of the incandescent cylinder so that, by coiling the filament, the length of the cylindrical layer was greatly reduced, but its diameter hardly affected.

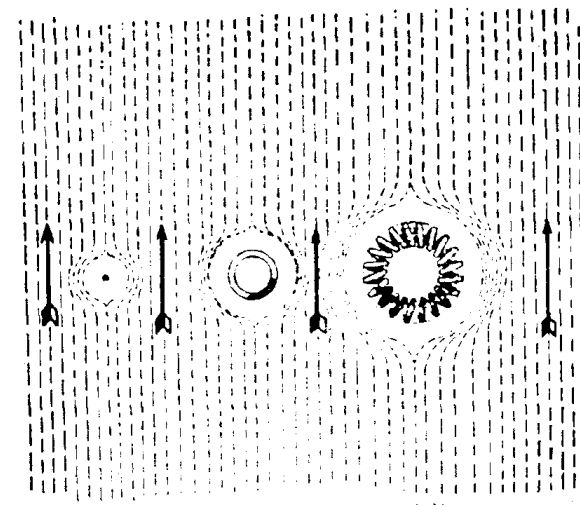


Fig. 1. Flow of gas around straight and coiled filaments seen in cross section.

Since heat losses and filament evaporation were found to be dependent on the dimensions of the layer, coiling the filament reduced the heat loss and hence raised the luminous efficiency of the lamp. The gas-filling helped to suppress evaporation of the hot filament and enabled the operating temperature to be increased without decreasing the life, thus contributing further to the increase in efficiency. Langmuir showed that other factors affecting the heat loss and evaporation were the molecular weight of the gas and its pressure in the lamp.

Coiling the filament introduced considerable problems, since the coil form was unstable, resulting in sagging and opening of the pitch of the coil. The catenary shape assumed by the filament (Inset 2) increased the heat loss and reduced the efficiency of the lamp. Another Coolidge patent of 1910 claimed the introduction of thoria as an additive during tungsten wire manufacture. This improved the ductility of the wire by creating a fine crystal structure which prevented brittle filament breaks at the grain boundaries. This 'offsetting' is due to adjacent crystals

slipping across one another on the face where they meet (Inset 3). Since this event can occur only where the crystal boundaries extend across the diameter of the wire, a finer crystal structure makes this possibility less likely. However, the addition of thoria did nothing to reduce the tendency of the filament to sag.

Up to about 1924 it appeared that the problem of sagging was incapable of solution in spite of intensive efforts by all the lamp manufacturers to overcome the defect. The solution of this problem was made more desirable by the discovery that even in vacuum lamps, coiling of the filament leads to greater efficiencies for the same lamp life when compared to the straight wire filament, because the rate of evaporation is much reduced due to tungsten condensation within the coil. This enables the operating temperature to be raised, giving a higher efficiency, which more than compensates for the poorer radiation characteristics of a coil compared with a straight filament at the same temperature. The resulting net gain thus arising from coiling is well worth while, with the added advantage of a more compact source.

Although in 1914 Langmuir had suggested coiled-coil filaments in gas-filled lamps to reduce further the cooling loss, the instability problem prevented their realization. Eventually a solution was found by the addition of traces of silica, alumina, and alkali oxides to the tungsten, combined with improved control of all the manufacturing processes (Inset 4). Problems of coil stability and uniformity of behaviour, however, remain today and are receiving continual attention.

The decade 1920–30 was noteworthy for the rapid development of manufacturing techniques by which high speed automatic machines replaced hand work and hand-controlled machine processes, and also for a revolutionary change in the concept of the function of inspection.

The main factor governing the introduction of automatic processes was the uniformity of the lamp components, particularly the glass components, such as the flange tube and the lamp bulb. With non-uniform parts each lamp needed individual treatment by hand. There was much progress made on the control of glass

composition which enabled even fluidity and stability to be maintained. High electrical resistance at elevated temperatures was also necessary. These requirements were met in glasses of low cost and automatic machines followed.

Thus, by 1930 the basic technical problems associated with the gas-filled tungsten filament lamp, its automatic production, and its close manufacturing control, were well established.

THE MODERN TUNGSTEN LAMP

The Filament

Single coil tungsten filaments are now used for the whole range of general service vacuum and gas-filled lamps from 15 watts to 1500 watts. They are made of tungsten containing traces of silica or alumina additives, usually termed 'dopes', which render the tungsten coils relatively stable.

Coiled-coil filaments were produced in the Research Laboratories of the General Electric Company Ltd in 1924 (Inset 5), but it was not until 1934 that high voltage general service coiled-coil lamps were first marketed. The delay was in part due to the development period required to design a successful automatic mounting machine able to handle the coiled-coil filaments without undue breakage, the design being partly dependent upon making the coils more robust by attention to coiling and heat treatment. The problem of arcing or 'flashing' within the lamp across the shorter coiled-coil was an additional delaying factor.

The modern coiled-coil filament is far more stable than it was twenty years ago, but further improvement is desirable. Wire preparation, coil treatment, mounting, lamp assembly, processing, and initial lighting are being studied for their relative effects on sag, both that which occurs in the first fractions of a second in which the lamp is lit, and that which occurs progressively during lamp life.

The Gas Filling

The requirements for the filling in gas-filled lamps are:

(*a*) it must be chemically stable and not react with the lamp components at high temperatures.

(*b*) its heat conductivity must be low,
(*c*) it must suppress evaporation efficiently,
(*d*) it must be available economically in large quantities, and
(*e*) its ionization potential must be high, to prevent 'flashing'.

Nitrogen was the only gas that came anywhere near to meeting these requirements until argon became commercially available. Langmuir's work showed that argon was far superior to nitrogen as a 'lamp gas', mainly due to its higher molecular weight and lower heat conductivity. It was also totally inert, whereas nitrogen slowly reacts with hot tungsten causing blackening of the bulb. It was found, however, that argon could not be used by itself in high voltage coiled filament lamps because of its low ionization potential, which led to serious 'flashing', particularly at switching-on. This flashing often destroyed the lamp because of the high current passed by the arc. Nitrogen mixed with the argon inhibited flashing and its use for this purpose became universal.

The present filling gas in general service lamps consists of 93 per cent argon and 7 per cent nitrogen at a pressure of about 700 mm Hg measured with the lamp cold. This gas composition and pressure have been reached following much experiment and development since the First World War. The first lamps were filled to limited gas pressures to reduce the heat loss which became serious owing to sagging of the single coil filaments and the progressive opening of the pitch of the coil as the lamp was burned. The nitrogen content of these early lamps was as high as 20 per cent to suppress the high tendency to 'flash' at the reduced filling pressures.

As the coiled filaments became more stable it was realized that the reduced evaporation, which occurred at higher pressures, outweighed the increases in heat loss and made possible higher efficiencies for the same life. The higher filling pressure reduced the tendency to flash, which allowed a decrease in the nitrogen percentage. The trend towards higher gas pressures continues and may make it possible to reduce further the nitrogen concentration with further gains in efficiency.

Krypton or xenon would be more advantageous than argon as a filling gas since they are heavier and poorer heat conductors. These gases were first separated commercially in 1935, some twenty years after the commercial extraction of argon, but they are still too expensive to use on a large scale in general service lamps. Krypton is used, however, in the miners' cap lamp, where high brightness and good efficiency are of first importance.

General service gas-filled lamps using coiled-coil filaments are made in wattage ratings from 40 W. to 100 W. The efficiencies are some 20 per cent higher than corresponding single coil lamps owing to the reduced energy losses by conduction and convection in the gas filling. Because of the very fine wire used in wattage ratings below 40 W. the cooling effects of the gas filling outweigh its advantages and these lamps are not gas-filled.

Lamp Fuses

At the moment of switching on a gas-filled lamp, there is a tendency for an arc to occur between the filament support leads, or, if the filament is defective, at a break in the coil. Unless suitable precautions are taken, this could lead to blown fuses or possibly even to a burst lamp if the mains fuses are unusually heavy. This phenomenon is known as 'flashing'. To avoid the inconvenience associated with blown fuses, the modern Osram lamp contains sealed glass-sheathed fuses in both leads fitted within the cap, which afford complete protection to the user (Inset 6).

As mentioned earlier, the problem of flashing accompanied the introduction of gas-filling. Apart from the effects of the nature of the gas and its pressure, much work was done about twenty-five years ago on the influence of lamp design and other aspects of lamp manufacture on the flashing problem. In 1930, 20 per cent or more of single coil lamps flashed either prematurely or at the end of life, but owing partly to greater control during processing as automatic manufacture grew, this figure was reduced to about 5 per cent in 1939 and is now less than 1 per cent.

Lamp fuses were first introduced in 1934 with the coiled-coil

lamp. These were essential since almost all coiled-coil lamps flash at the end of life. The first fuses were unsheathed, and were unsatisfactory because of their failure by oxidation and chemical attack from the capping material and soldering fluxes. Extensive trials were carried out on fuse materials, dimensions, and stability, resulting in the glass-sheathed fuse. Although they were expensive, sheathed fuses were immediately used in coiled-coil lamps just before the war and are now also used in all Osram single coil gas-filled lamps.

Precision in Lamp Processing

From what has been said it may be realized that the tungsten lamp is a sensitive and intricate device calling for extreme precision and cleanliness in all stages of manufacture. The tungstic oxide prepared from the ore has to be of the highest purity, with a specific particle size distribution. The oxide preparation and the reduction process to the metal powder have a large bearing on whether or not the filaments made from it will reach the standards demanded in modern lamps.

The wires most commonly used in lamps for general lighting service vary in diameter from 0·015 mm for the H.V. 15 W. lamp to 0·05 mm (Inset 7). These wires must be accurately round with local variations in diameter not greater than 1 per cent, i.e., less than 15 millionths of a centimetre. Extreme accuracy has therefore to be maintained in the diamond dies used to draw the wire. Filament size is usually controlled by weighing, on a high precision torsion balance, fixed lengths of wire cut in special gauges.

The wire is wound in the form of a fine helix about a core or mandrel of about 3 to 6 times the wire diameter. The tolerances for mandrel diameter are smaller even than for the wire, and the hardness of the mandrel metal must be constant so that the degree of 'biting in' of the wire as it is wound does not vary. The speed and tension of coiling is very accurately controlled, so that the total length of wire in every cut coil falls within very small limits. Tolerances on wire and mandrel diameter are matched to reduce the overall tolerance on the finished filament coil. The distance between turns of the coil is about 0·2 to 0·6

times the filament diameter and must not vary by more than a very small percentage. To prepare a coiled-coil filament the primary coil, whilst still on its mandrel, is coiled accurately a second time. The primary and secondary mandrels are eventually dissolved out with acid. The length of wire in a coiled-coil filament is about 100 cm. whilst the coil length is only 3 cm.

Such a fine wire is extremely sensitive to contaminants (Inset 8), particularly because of the cyclic chemical reactions that can occur over the very broad temperature range within a lamp, from about 50° C at the coolest part to about 2500° C at the filament. Less than ten parts per million of water vapour will cause an increase in filament sag and possible early failure due to filament oxidation. A similar quantity of hydrocarbon vapour will cause the filament to squirm and become distorted (Inset 9).

To render the lamp less sensitive to slight variations in pumping quality and assembly conditions, phosphorus getter is used to react initially with any contaminants present. The filament is dipped into a rotating pot containing phosphorus suspended in alcohol, before it is sealed into the lamp bulb. The amount and quality of the getter itself is very critical.

Lamp Design

To design a tungsten lamp of given voltage and wattage to give the maximum possible efficiency for a given life, is complicated, and does not easily lend itself to a solution on a theoretical basis. Almost all the parameters of lamp design (wire size, wire length, coil diameter, coil pitch, coil length, coil shape, nitrogen content of the gas, filling pressure, bulb volume, size and shape) are interdependent and each affects to a greater or lesser degree the initial efficiency, efficiency maintenance, and life of the lamp.

There are, however, empirical design methods for the filament which have been developed over the years, with limits on the coil diameter and pitch to maintain coil stability and including compensation procedures for the manufacturing tolerances. Theoretical work was carried out before the war on the inherent luminous efficiency of an incandescent tungsten coil and its

dependence on coil pitch and diameter, but this has not been related yet to the effects of filament sagging and heat loss, and hence to the lamp efficiency in lumens per watt of a practical gas-filled lamp.

Unlike that of the fluorescent lamp, the life of a tungsten lamp is largely determined by its initial efficiency. Throughout the last thirty years, the designed minimum average life of general service lamps has remained at about 1,000 hours and recent investigations have shown this value still to be the most advantageous to the user.

The internally frosted or pearl bulb was introduced in 1925 to diffuse the light from the bright filament. There is negligible light absorption in the diffusing bulb, the smooth outer surface of which does not readily collect dirt and is in any event easily cleaned.

Continuous improvements are being made in the stability of the lamp filaments, and lamp design has been modified to exploit these improvements to the fullest extent to achieve the maximum increase in lamp quality. The order of filament stability now possible has enabled two such design modifications to be made within the last few years:

(*a*) In Osram high voltage coiled-coil lamps the number of supports used to carry the filament between the lead-wires has been reduced from three to two. Since the filament coil length remains unaltered, there is an increase in filament weight between points of suspension of about 9 per cent, which calls for a high degree of coil stability. The reduction in the number of supports means that fewer cold points are introduced along the coil length, and worthwhile quality improvements have been gained.

(*b*) In the U.S.A. an improvement in quality has recently been obtained by mounting the filament vertically in the bulb. This localizes the evaporated tungsten to a small area of the bulb above the filament axis, resulting in a better efficiency maintenance throughout life, and a slightly longer life for the same efficiency if the lamp is burnt vertically with the cap uppermost. The single coil filament hitherto widely used in the States has at the same time been replaced by a double coil giving a further gain in efficiency.

Axial mounting of the filament is difficult to apply in Britain in ratings below 200 W. since the usual supply voltage is double that in the U.S.A., and necessitates the use of fine filaments which cannot be mounted vertically with a high degree of stability. There is little doubt that the horizontal arrangement of the filament so widely used in this country possesses special advantages in obtaining the best light distribution in the working plane.

Lamp Performance

Improvement in filament lamp quality is not so much the result of spectacular new developments, as of continual attention to detail and the introduction of such small design modifications as those described above.

The quality of the lamp can be expressed in a single figure called the 'efficiency percentage', which depends upon the initial filament wattage, the initial efficiency in lumens per watt, the lamp life, and the efficiency maintenance during life. Since the war there has been a general quality increase of about 6 per cent (5 units in efficiency percentage) for both single and coiled-coil lamps (Figure 2).

Apart from the rises in average level of lamp quality in terms of lamp efficiency and maintenance, increasing importance is being given to the concept of uniformity of performance between lamps. Statistical analyses of data show the magnitude of variability and how it is distributed within and between lamp batches on different occasions, and these analyses are essential in assisting the intensive programme now in hand in all aspects of lamp making to improve the already high level of uniformity. The task is complicated and intricate owing to the considerable inter-dependence of lamp parameters.

DEVELOPMENT TRENDS

The tungsten lamp industry remains a fully dynamic one, and there are many plans for expansion and development that are

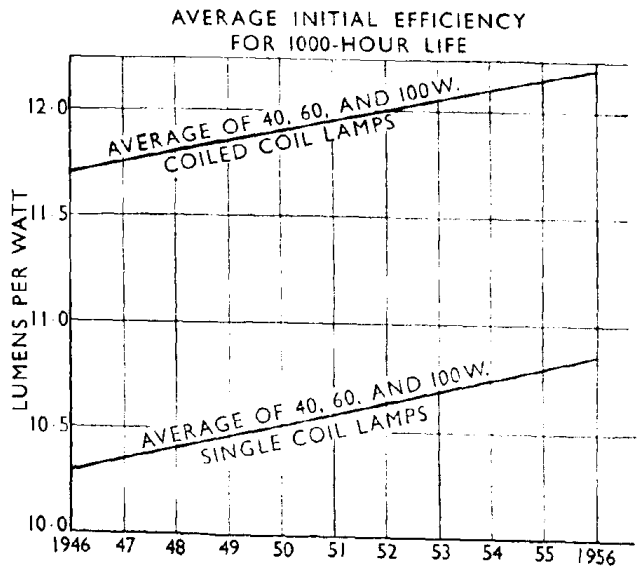


Fig. 2. Improvement in 'Osram' lamp quality since 1946.

being put into effect rapidly. Before concluding the present account it may be of interest to mention one or two of the present trends.

Performance gains may be achieved by increasing further the gas filling pressure and much work has been done to determine the limit to which lamps may, with full safety, be filled, and to develop techniques for sealing lamps filled to pressures in excess of one atmosphere.

Much thought is being given to the problem of reducing the bulb sizes of general service lamps and these trends are being internationally discussed with Continental lamp manufacturers with a view to European standardization. The technical problems involved include the effects on quality of bulb size reductions, the effect on lamp cap temperatures of bulb size changes, the standardization of lamp temperature measurement methods,

the effects on lamp fittings design (especially in optically designed fittings where light centre length is important), the effects on cable temperatures at or near the lamp terminals, and the investigation of the smallest bulb size technically possible and acceptable.

The present article has been concerned mainly with general lighting service lamps, but it should be mentioned that much important work is also proceeding on lamps for special purposes such as projection, automobile lighting, and other applications.

What of the Future?

Although much has been accomplished in the realm of electric incandescent light sources, much still remains to be achieved. The theoretical maximum efficiency for a 'full' radiator occurs at an operating temperature of around 6000° K, when the efficiency is 84 lumens per watt. The efficiency of a household general service lamp with a thousand hour life is less than 15 lumens per watt, and the maximum efficiency of any practical tungsten lamp is about 30 lumens per watt. The practical limits are set by the physical properties of the incandescent material; its melting point, its evaporation rate at high temperature, and its electrical resistivity. No satisfactory replacement for tungsten has been found over the last forty years, although the low resistivity of tungsten necessitates its use as a very fine wire to give the desired electrical characteristics. However, the technological environment of the lamp industry has advanced considerably in this time and much effort throughout the world is now being spent on developing stable high temperature materials in connexion with the nuclear energy and aeronautical programmes. Such developments are being closely watched, since they may have some bearing on the incandescent light source. Marked improvements in efficiencies may be possible, even with tungsten, by new methods of controlling the evaporation process.

The understanding of the basis of the properties of materials in the solid state has also developed very rapidly in the last decade. This understanding may help the lamp engineer to

fabricate refractory materials which emit, when heated, a higher proportion of radiation in the visible region, and therefore yield a higher luminous efficiency.

Thus, the incandescent lamp industry, which contributed so much to the establishment of the present technological era, may itself benefit from the fruits of technology in the industries to which it gave birth.

* * *

The author wishes gratefully to acknowledge the assistance given in the preparation of this article by many colleagues in the Research Laboratories and Osram Division of the General Electric Company Limited. In more extended form it was published in the *G.E.C. Journal* (1958 **25**).

THE SPINNING ELECTRON

P. S. FARAGÓ

INTRODUCTION

As is very well known, the electron is a particle of charge $e = -4{\cdot}8 \times 10^{-10}$ e.s.u. $= -1{\cdot}6 \times 10^{-19}$ coulomb, and of mass $m = 9 \times 10^{-28}$ g.

Both these quantities have been measured by various experiments carried out with free electrons, and also spectroscopically, that is by experiments with electrons bound to atoms. Both types of experiment give the same results within the limits of experimental error. The surprising assumption of the existence of 'electron waves', originally introduced for the interpretation of quantized states of electrons in atoms, was also justified by experiments on beams of free electrons. In fact electron diffraction became a phenomenon of great practical importance. The validity of the relation $\lambda = h/p$ between the wavelength λ and the momentum p of the electron (h being Planck's constant) has been justified experimentally with high accuracy. But in the data listed so far, all the quantities characterizing the electron have not been mentioned.

The detailed study of atomic spectra has shown that in order to interpret all the observations, such as the multiplet structure of spectral lines, splitting of spectral lines from sources in magnetic or electric fields (Zeeman-effect or Stark-effect), and so on, in terms of the quantum laws of electron motion around the nucleus, it has to be assumed that the electron has, independently of its orbital motion, an intrinsic angular momentum and a magnetic moment.[1] The hypothesis first introduced by Uhlenbeck and Goudsmit in 1925 stated that the angular momentum of the electron is $\frac{1}{2}(h/2\pi) \equiv \frac{1}{2}\hbar$ and it can have only two orientations relative to a given direction of reference:

parallel or antiparallel. The number $s=1/2$ giving the angular momentum in units of $\hbar$ is called the spin. The magnetic moment of the electron is $\mu=\frac{1}{2}(e\hbar/mc)\mu_0$: this quantity is called 1 Bohr-magneton. The ratio of the magnetic moment in units of Bohr-magnetons to the angular momentum in units of $\hbar$ is called the g-factor of the particle

$$g=\frac{\mu/\mu_0}{s}$$

Thus for electrons $g=2$.

The relation between the angular momentum and the magnetic moment is the same as that for a rotating sphere carrying a uniformly distributed charge on its surface (though it is not possible to apply consistently this simple model of the 'spinning electron' without getting contradictory results).

It was not altogether satisfactory that the concept of electron spin and magnetic momentum was justified solely by its success in the interpretation of previously unexplained features of atomic spectra; it was most desirable to have independent evidence. In what follows it will be seen that the problem is much more difficult than it appears at first sight, in particular if direct and highly accurate measurements on free electrons are wanted.

THE STERN-GERLACH EXPERIMENT

Both the old quantum theory due to Bohr and Sommerfeld, and wave mechanics teach us that the angular momentum of an atom due to the orbital motion of its electrons and also the corresponding magnetic moments vanish in certain electron configurations, the only angular and magnetic momenta left being that of one electron's intrinsic spin and magnetic moment (closed shells + one s-electron). This is the case, for instance, in alkali metals or noble metals. If such an atom is sent through a homogeneous magnetic field, it will experience only a torque, but no translatory force. A beam consisting of such atoms will not be deflected. Applying, however, an inhomogeneous field H there will be a translatory force acting on the atoms. Figure 1 shows schematically the experimental arrangement of Stern and

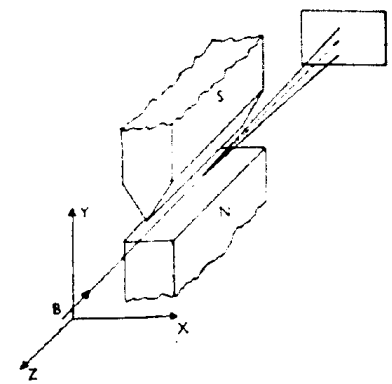


Fig. 1. Stern-Gerlach experiment. An atomic beam B passing through the inhomogeneous field of the magnet S—N is split into two.

Gerlach. A beam of atoms evaporating from a furnace and defined by slits passes through an inhomogeneous field in a narrow region around its plane of symmetry. If the magnetic moment μ could have any arbitrary orientation, a force

$$\mathbf{F}\sim F_y=|\mu|\frac{\delta H_y}{\delta y}\cos(\mu,H_y)$$

would spread the beam uniformly in the y direction. Instead of this, however, the beam produced two distinctly separate spots on the screen indicating that there are only two possible orientations of μ relative to the external field H: parallel or anti-parallel. Knowing the geometrical parameters of the experimental arrangement, and the magnetic field distribution, the observed beam splitting yielded a value for $|\mu|$ equal to 1 Bohr-magneton within the limits of experimental error.

This experiment is most convincing but, of course, it should not be considered as a direct proof of the existence of an intrinsic spin and magnetic moment of the electron. The experiment deals with electrons bound to atoms and the information obtained refers, strictly speaking, to certain features of the behaviour of electrons in atoms. An *intrinsic* property of the

electron should preferably be demonstrated by experiments on *free* electrons.

The obvious suggestion is to carry out a Stern-Gerlach type of experiment with free electrons. Unfortunately, however, this would not give the necessary information, as was pointed out by Bohr.[2] The argument is roughly the following.

Since the electron is a charged particle, when it moves in a magnetic field it experiences a force in a direction perpendicular to the plane defined by the directions of the magnetic field and of the velocity of the electron (Figure 2*a*), the magnitude of the force being proportional to the field intensity and to the electron velocity. Consider now electrons passing through the inhomogeneous magnetic field schematically shown in Figure 2*b* in a direction perpendicular to the plane of the drawing. Those entering the field in its symmetry plane will be deflected parallel

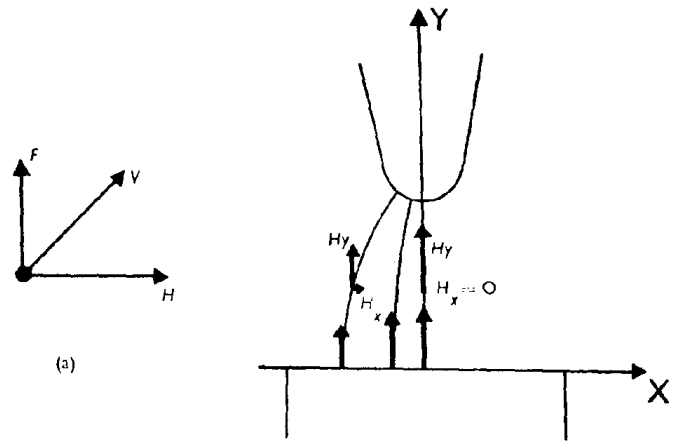


Fig. 2. (*a*) An electron of velocity v moving in the magnetic field H is deflected in the direction F. (*b*) The field of a 'Stern-Gerlach magnet' in the plane perpendicular to the beam. In the symmetry axis the field has only a y-component, but elsewhere there is an x-component also.

to the x-axis. This deflection is perpendicular to the deflection caused by the inhomogeneous field acting on the magnetic moment of the electron. Thus the two effects of different origin could be separated. However the field outside its symmetry plane has an x-component also. Consequently for an electron moving outside the symmetry plane, a component of the force due to its charge, and that due to its magnetic moment will be parallel, and the spread of the beam caused by the first may easily mask the splitting of the beam caused by the second. It is not difficult to show that in order to separate the two effects the width of the beam parallel to the x-axis has to be reduced by narrow slits to such an extent that electron diffraction will occur. As soon as this happens, the pattern observed on a screen cannot be interpreted in terms of the concepts of classical physics.

It is true quite generally that, due to the Uncertainty Principle, the intrinsic angular momentum of the electron cannot be separated from its orbital momentum and cannot be determined uniquely by an experiment that can be interpreted in terms of the *classical* concept of particle trajectories. Thus a very large and attractive category of experiments is ruled out.

ELECTRON POLARIZATION BY COULOMB SCATTERING

In search of experimental evidence of the spin of free electrons some suitable *quantum* phenomenon must be produced and detected. The first of such phenomena was predicted by Mott, applying Dirac's relativistic quantum theory to the problem of the scattering of electrons by heavy nuclei.

Before summarizing the results of these calculations, let us recall one of the most important features of Dirac's theory, as distinct from any earlier formulation of quantum theory.

The concept of electron spin and magnetic moment was introduced in the early stages of quantum theory as an *ad hoc* hypothesis supplementary to the rules of quantization, and it retained its character of an additional postulate even when the otherwise self-consistent theory of wave mechanics or matrix

mechanics was developed. Any attempt at a consistent theory of the spinning electron failed, leading to some contradiction or other with firmly established principles of relativity theory. If we add that there was then not a single phenomenon known in which the spin of free electrons played a part, one may doubt whether it was justified at all to consider spin as an intrinsic property of the electron.

It was the relativistic quantum mechanics of Dirac that changed the situation fundamentally. In this theory the spin and magnetic moment of the electron appears as a necessary consequence of the basic properties of mathematical relations expressing laws of nature and taking proper account of the Uncertainty Principle (relations between results of simultaneous measurement of physical quantities) and the Principle of Special Relativity (equivalence of inertial frames of reference).

Application of this theory to the problem of electron scattering by heavy nuclei led to the prediction of what is called electron polarization by Coulomb scattering.

In the vicinity of a nucleus the electric field varies very rapidly, and therefore the classical calculation of electron trajectories would make it necessary to define the initial conditions more precisely than is allowed by the Uncertainty Principle. The difficulty is avoided by considering the electron as a wave-train and applying the methods of wave mechanics, instead of trying to trace the trajectory of a point-particle. In addition to changing over to a quantum mechanical treatment, one has to take proper account of relativistic effects. If a magnetic dipole is at rest in an electrostatic field, there is no interaction between them. In the present problem, however, the electron is moving at a large velocity in the field of the nucleus and in this case – as taught by the special theory of relativity – the electron 'sees' not only an electric field of modified intensity but a magnetic field also. This magnetic field will interact with the magnetic moment of the electron; the effect is usually referred to as 'spin-orbit coupling'.

Applying Dirac's theory to the problem means taking a consistent account of all these effects automatically, and it leads to

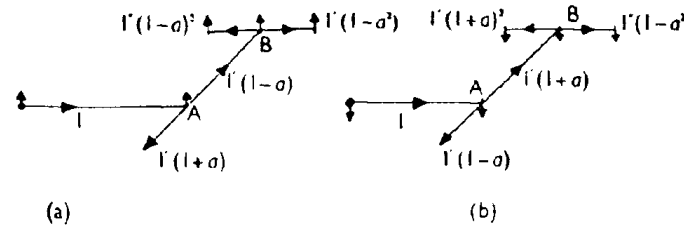


Fig. 3. A fast electron beam of intensity I is scattered at right angles by a heavy nucleus once at A and again at B. The beams considered are in one plane and the spin of the electrons is perpendicular to this plane. The intensities of the scattered beams depend on the spin orientation.

the following result.[3] Fast electrons of 0·1–0·2 MeV kinetic energy with their spins perpendicular to the direction of motion are scattered anisotropically by heavy nuclei. Figure 3 shows schematically what can be expected when scattering of a transversely polarized electron beam is observed in the plane perpendicular to the spin direction at a scattering angle of $\pm 90°$. Considering electrons with spin 'up' (Figure 3*a*) the scattered intensities are $I'(1-a)$ and $I'(1+a)$ respectively. A second scattering in the same plane again at $\pm 90°$ will give intensities $I''(1-a)^2$ and $I''(1-a^2)$ respectively. Electrons with spin 'down' (Figure 3*b*) behave in the opposite manner. Consequently if there is a beam consisting of equal numbers of electrons with spin 'up' and 'down', the intensities observed in opposite directions after the second scattering are

$$I_1 = 2\,I''(1+a^2)$$
$$I_2 = 2\,I''(1-a^2)$$

and the ratio of the two intensities is

$$I_1/I_2 = (1+a^2)/(1-a^2)$$

The asymmetry is usually characterized by the quantity

$$2a^2 = \frac{I_1 - I_2}{\frac{1}{2}(I_1 + I_2)}$$

Figure 4 shows the variation of the asymmetry (in percentage)

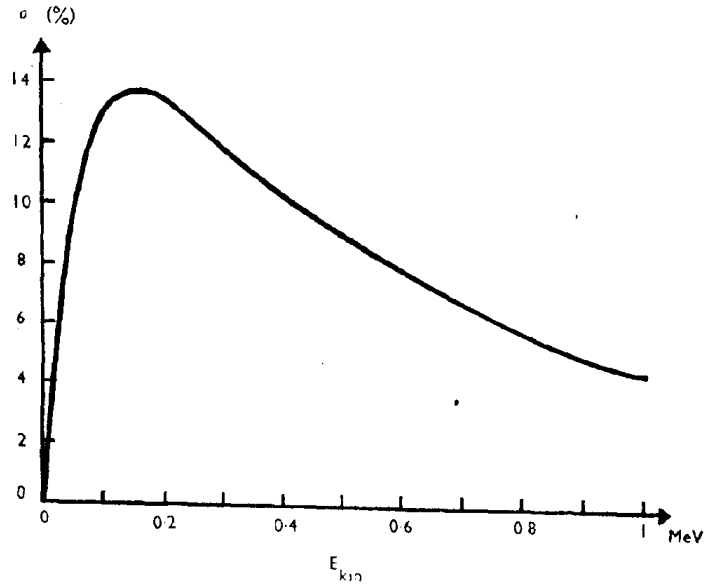


Fig. 4. Percentage asymmetry as a function of kinetic energy in a double scattering experiment with an unpolarized incident beam. Angle of scattering, 90°; scatterer: mercury.

as a function of the kinetic energy of the electrons if the scatterer is mercury (atomic number $Z=80$).

If the beam consists of electrons of random spin orientation, the first large-angle scattering will produce a transversely polarized beam, i.e., one in which one of the spin directions perpendicular to the direction of motion is dominant. The anisotropy of a second large-angle scattering gives a means for the detection of the polarization. This phenomenon was studied by a number of investigators, using beams of about 200 keV and, usually, thin gold foils as scatterers. The most recent experimental results are in fair agreement with theory. Therefore it can be said that there is at last a direct evidence of the existence of an intrinsic spin 1/2 of free electrons. The accuracy of the

polarization experiments, however, is far too small to give satisfactory information about the magnitude of the magnetic moment. Still, the overwhelming success of Dirac's theory in all the problems it was applied to made it most plausible that the magnetic moment of the electron is 1 Bohr-magneton.

'ANOMALOUS' MAGNETIC MOMENT OF THE ELECTRON

Within the last ten years or so, however, the situation has changed considerably. With the increase in accuracy of spectroscopic methods, the development by Rabi and his collaborators of atomic and molecular beam technique, and the advent of radio frequency and microwave spectroscopy, certain experimental results showed a small but genuine discrepancy when compared to the predictions of Dirac's theory. The first and certainly the most famous one was a shift of certain energy levels in the fine structure of the best known atomic spectrum, that of hydrogen. This phenomenon is usually referred to as the 'Lamb shift'. Measurements of the hyperfine structure of hydrogen, and the precise comparison of g-factors of one atom in various energy states, or those of different atoms in corresponding energy states, gave results consistent only with the assumption that the magnetic moment of the electron is not exactly 1 Bohr-magneton but is larger by about 0·1 per cent.

Soon after the first experimental evidence of this 'anomalous' magnetic moment, its theoretical interpretation was given by Schwinger. Using the methods of quantum electrodynamics a 'radiative correction' could be derived expressing the effect of the electromagnetic field of the electron upon itself. The g-factor of the electron could be determined theoretically and the 'anomaly' could be given by a series of terms proportional to ascending powers of the fine structure constant $\alpha=e^2/\hbar c \sim 1/137$. For the first two terms the proportionality factor was also calclulated. Values accepted at the present time are

$$\tfrac{1}{2}g = 1+(\alpha/2\pi)-0{\cdot}328(\alpha/\pi)^2+\ldots = 1{\cdot}0011596$$

The experiments referred to previously had determined g with an accuracy of 1 part in 100,000 and agreed with the above

value. But once again the highly accurate results have been obtained by experiments on electrons bound to atoms and the evaluation of the observed data, taking various corrections into account, involves considerable theoretical assumptions which themselves need justification. Therefore the importance of the direct measurements of the g-factor of free electrons became greater than ever before. On the other hand, of course, the conditions of such an experiment are extremely difficult since the accuracy of the measurements has to be comparable with that of the spectroscopic results, i.e., of the order of one part in a hundred thousand.

The extreme accuracy of microwave spectroscopy suggests that its methods should be applied to the present problem also. If a beam of free electrons passing through a homogeneous magnetic field H_0 is exposed to a radio frequency magnetic field perpendicular to H_0 and having a frequency

$$\omega_s = \frac{2\mu H_0}{\hbar} = \frac{g}{2}\,\frac{e}{m}\,\frac{H_0}{c}$$

the radio frequency field will induce transitions between the two possible orientations of the spin in the field H_0. If the reorientations of the spins as such, or the sharp increase of absorption of the radio frequency radiation, can be detected, one can measure ω_s in the given field H, and g can be calculated. Such an experiment, however, is extremely difficult, a few of the typical problems being the following.

(*a*) An electron beam as it is emitted by, say, a thermionic cathode is unpolarized. The number of electrons with spins parallel and antiparallel to the field is equal. The r.f. irradiation will produce no net effect, which could be detected by macroscopic methods, since equal numbers of transitions $s = -1/2 \longrightarrow +1/2$ and $s = +1/2 \longrightarrow -1/2$ will take place. Therefore an experiment of the kind outlined needs a polarized electron beam to start with.

(*b*) How sharp the resonance is, depends on the time a resonant system – oriented electrons at present – is exposed to the r.f. field, and on the damping of the system itself. To achieve a

resonant width $\Delta\omega/\omega \sim 10^{-5}$ the electrons have to spend about 10^5 cycles in the r.f. field and during this time their orientation must not be changed by perturbations, such as certain types of collisions. Consequently some device must be found to 'trap' the electrons in the field. Also the beam must be in an extremely good vacuum, or in a specially selected atmosphere, and the electron space charge must be low. The low electron concentration suggests that the detection of absorption of r.f. radiation by an increased loss in the r.f. circuit which is the usual method of detecting resonance absorption in r.f. and microwave spectroscopy is hardly feasible. Therefore some special method of detection has to be found.

COMPARISON OF THE g-FACTOR OF FREE ELECTRONS AND ELECTRONS BOUND TO ATOMS

Dehmelt[4] has compared the g-factor of the free electron with that of the valency electron of the sodium atom. The principle of the experiment can be summarized as follows. If there are free electrons in a gas atmosphere, the collision of slow electrons with atoms of opposite spin orientation may result in the 'exchange' of the spin orientation of both partners, with the conservation of the total angular momentum. Let us consider now a volume filled with a vapour consisting of oriented sodium atoms, i.e., in every atom the spin of the valency electron has the same orientation. Let us produce free electrons in the same volume by some

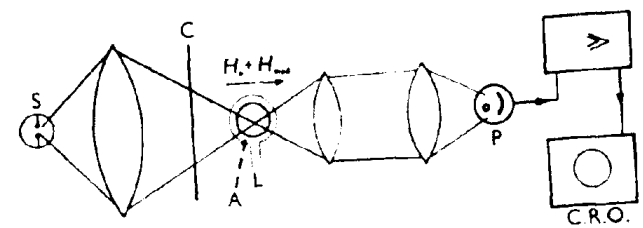


Fig. 5. Schematic diagram of the arrangement used by Dehmelt to compare the g-factor of free electrons with that of the ground state of sodium atoms.

means (e.g., photo-electrically); their spins will be oriented at random. The exchange collisions of electrons with sodium atoms will increase the number of free electrons having one particular spin orientation in return for a partial de-orientation of the sodium atoms themselves. After a certain period of time an equilibrium state will be established with partial orientation of both electrons and atoms. If this equilibrium state is disturbed by destroying the polarization of the electrons, it will be re-established again by a further loss of the orientation of the atoms. In short: free electrons can be polarized by exchange collision with oriented atoms and a change of the degree of polarization of the electrons can be transferred – by the same mechanism – to a change in the orientation of the atoms.

Now, it was shown by Kastler[5] that a partial orientation of atoms can be produced and detected by an optical method, called 'optical pumping'. This is based on the spectroscopic selection rules for the Zeeman-effect, i.e., the laws governing transitions induced by polarized light if the atoms are in a magnetic field. The technique of optical pumping was applied by Dehmelt, with the apparatus schematically shown in Figure 5. Saturated sodium vapour and an inert buffer gas are contained in the 'absorption bulb' (A) of volume 200–1000 cm³ situated in the uniform and constant magnetic field H_0. Free electrons are produced by photo-ionization using the radiation of a mercury arc (not shown). The orientation of the sodium atoms is achieved by illuminating the bulb with the light of a sodium arc (S) through a circular polarizer (C). The intensity of the transmitted light, as measured by the photo-cell (P), gives an indication of the degree of orientation: higher transmitted intensity corresponds to a higher degree of orientation. The loop (L) produces a r.f. magnetic field perpendicular to H_0, and with the aid of modulating coils a field H_{mod} parallel to H_0 is superimposed to H_0. Whenever $H_0 + H_{mod}$ reaches the resonance value corresponding to the frequency of the r.f. field, the absorption of r.f. energy by the partially oriented free electrons will reduce the populations of the two possible states of spin orientation practically to equality, i.e., it destroys the polarization of the electron gas.

This in turn – as explained above – reduces the degree of orientation of the sodium atoms and can be detected by the reduced transmission of the polarized sodium resonance radiation. The resonance curve can be displayed on a cathode ray oscilloscope.

To evaluate the g-factor of the electrons an accurate knowledge both of the resonance frequency ω_r and of the resonance field intensity H_r is necessary. To avoid the measurement of the latter, another type of resonance phenomenon is observed in the same field, namely the transitions between the 'hyperfine' levels of the ground state of the sodium atom, induced by radio frequencies. From these data and using certain results of atomic beam experiments carried out by other authors, Dehmelt found that the g-factors of the valency electron of sodium in its ground state and of the free electron are equal to an accuracy of 3 : 100,000.

This experiments yields some of the most important information available to date on the subject, but because of its complicated character more direct measurements are still desirable.

HIGH ACCURACY MEASUREMENTS ON FREE ELECTRONS

The first successful result of high accuracy obtained in experiments with free electrons has been reported recently from the University of Michigan; the basic idea was put forward a few years ago[6].

Let us consider a beam of polarized fast electrons in a homogeneous magnetic field H, moving in a plane perpendicular to H, their spins being also perpendicular to H. The electrons describe circular orbits with an angular velocity

$$\omega = \frac{e}{m}\,\frac{H}{c}$$

and the spins precess around the field direction with the Larmor frequency

$$\omega_r = \frac{g}{2}\,\frac{e}{m}\,\frac{H}{c}$$

Were the magnetic moment of the electron 1 Bohr-magneton

exactly, the two kinds of angular velocities would be equal to one another and the angle between the direction of motion (orbital momentum) and spin would be constant throughout the motion. If on the other hand there is a deviation between the actual magnetic moment and the Bohr-magneton, $g \neq 2$, the angle between the two vectors will change at a rate $\omega_s - \omega$, or after k complete orbital revolutions by an angle

$$\Phi = 2\pi k \frac{\omega_s - \omega}{\omega}$$

This means that in case of $g \neq 2$ the initial polarization of the beam is gradually changing during the motion.

Previously it was seen that Coulomb scattering shows an azimuthal asymmetry if the beam is transversely polarized, but such an asymmetry does not occur if the polarization is longitudinal. Therefore, if the beam is allowed to impinge on a scatterer and if we measure the azimuthal asymmetry of the intensity of the scattered beam as a function of the number of orbital revolutions we get a periodic curve (Figure 6). One whole period is covered if

$$2\pi k_0 \frac{\omega_s - \omega}{\omega} = 2\pi,$$

and hence

$$\frac{1}{k_0} = \frac{\omega_s - \omega}{\omega} = \frac{g}{2} - 1.$$

Relativistic corrections due to the high velocity of the electrons were neglected.

A determination of the number of orbital revolutions k_0 corresponding to one period in the variation of the scattering asymmetry gives directly the deviation of the g-factor from the value 2. This deviation is expected to be of the order of 0·001 and its measurement to an accuracy of only 1 per cent gives the g-factor itself to an accuracy of 1 : 100,000.

In the experiment carried out by Crane *et al.*[7] a pulsed beam of polarized electrons produced by Coulomb scattering, spirals

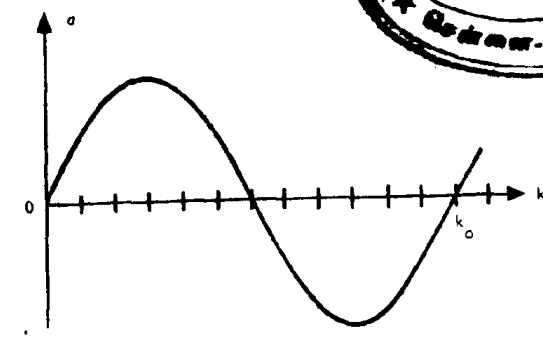


Fig. 6. The azimuthal asymmetry a of the scattering of a polarized beam depends on the number of orbital revolution k described in a magnetic field.

in the homogeneous field of a 20 feet long solenoid. Each 'turn' of the 'tightly wound' helical orbit corresponds essentially to one circular orbit described above. To produce a prescribed number of orbital revolutions the electrons are trapped in the following manner. Around the central region of the solenoid, an auxiliary coil modifies the homogeneous field; it is still axially symmetric but is radially decreasing (Figure 7). In this region the axial component of the velocity of the electrons is decreased by a voltage pulse applied to a system of coaxial electrodes (not shown in the figure). The result is a kind of electron motion well known in certain high-energy electron accelerators, in particular in betatrons: the particles, while describing approximately circular orbits, are also oscillating to and fro in a direction parallel to the solenoid axis z. This motion is maintained until a second voltage pulse increases the axial component of the electron velocities to such an extent that the electrons can leave the 'magnetic trap'.

The electrons, leaving the trap, hit a thin gold foil and the polarization of the beam is measured by detecting the asymmetry of the large angle Coulomb scattering. From the time spent between the two scatterers, and the magnetic field inten-

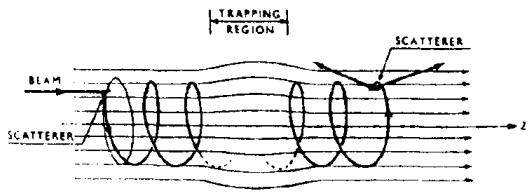


Fig. 7. Schematic diagram of the magnetic field distribution and electron orbits in the Michigan experiment.

sity in the solenoid (both quantities are measured directly), the number of orbital revolutions can be obtained. Plotting the variation of polarization as a function of the number of orbital revolutions, the quantity $\frac{1}{2}g-1$ is derived, as explained above.

The accuracy of the value of the g-factor itself as given by this experiment is claimed to be of the order of 1 : 1,000,000. Although some unexpected phenomena observed in the course of the experiment make further investigations necessary, there seems to be no serious discrepancy between the experimental results and the theoretical predictions.

Two further suggestions for direct measurements will be outlined briefly; each is the basis of an experiment in progress.

(*a*) An experiment is being carried out in Cambridge by Frisch, based essentially on a suggestion by Bloch.[8]

Thermal electrons entering a constant and homogeneous magnetic field with velocities very nearly parallel to the field describe spiral orbits. If an electrostatic field of properly chosen potential distribution is set up suddenly, some of the electrons can be 'trapped'. Figure 8 shows schematically an electrode system and the potential distribution along its axis. Electrons in region D with energies smaller than the height of the potential barrier will be trapped, swinging to and fro, the radial spreading of the beam being prevented by the magnetic field.

The total energy of the electrons can be written in the form:

$$E=\frac{p^2}{2m}+[(2l+1)+gs]\mu_0 H,$$

where $p^2/2m$ is the kinetic energy of the electrons; $l=0, 1, 2, \ldots$

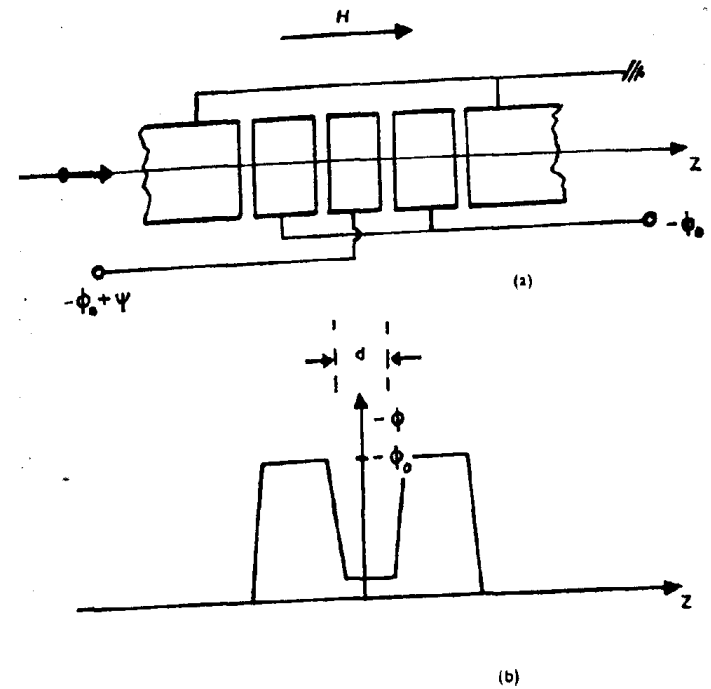


Fig. 8. Electrons moving along the z-axis are trapped in the region 'd' of the electrode system (a) if their kinetic energies are lower than the height of the potential barrier (b).

and $s=\pm 1/2$ are 'quantum numbers' characterizing the orbital motion and the spin orientation of the electrons, respectively. The term $(2l+1)\mu_0 H$ is the energy due to the orbital motion in a plane perpendicular to the magnetic field; in classical terms one would say that this is the energy of the magnetic dipole equivalent to the circular current produced by the orbital motion of the electron. The energy of the intrinsic magnetic dipole moment is given by $\pm g s \mu_0 H$. Figure 9 schematically shows the energy levels $E(l,s)$ occupied by electrons of kinetic

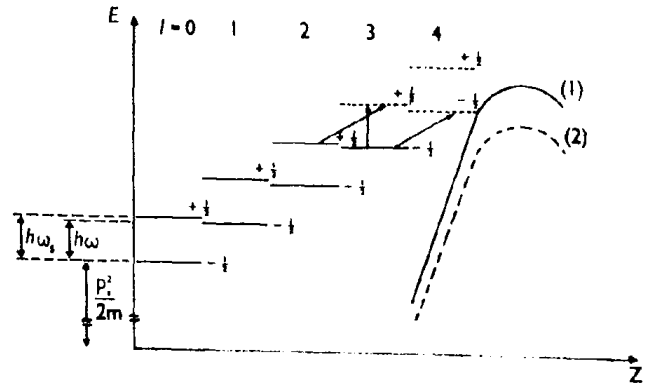


Fig. 9. Energy levels populated by electrons moving in a homogeneous magnetic field. Curve (1) shows an electrostatic potential barrier; an inhomogeneity parallel to the main field tilts the barrier (1) into (2).

energy $p^2_z/2M$. The solid curve on the right-hand side shows the potential barrier. In the particular example shown electrons with quantum numbers $l \leqslant 3$ are all trapped and so are electrons with $l=4$ and $s=-1/2$; those with $l=4$, $s=+1/2$ are not trapped.

It can be shown that in an inhomogeneous field with a gradient parallel to the z-axis the electrons are accelerated. The effect of the inhomogeneous field is in fact equivalent to tilting the potential barrier. In the example shown in Figure 9 the application of the inhomogeneity has the effect of replacing the potential barrier (1) by (2). Thus the application of the magnetic field inhomogeneity 'blows out' some of the electrons previously trapped, namely those populating the top levels. When the inhomogeneity is removed the potential barrier (2) regains its original form (1), and a repeated application of the same inhomogeneity will not liberate any more electrons, because the top levels are already empty, and the rest remain below the rim of the barrier.

If a r.f. field is now applied, there are two frequencies at

which transitions to higher levels are rendered possible by absorption of r.f. energy. One of these, given by the relation

$$\hbar\omega = 2\mu_e H$$

increases the energy of the orbital rotation (i.e., the quantum number l); the other, given by the relation

$$\hbar\omega_s = g\mu_e H$$

produces 'spin flipping' (i.e., the quantum number $s=-1/2$ changes into $s=+1/2$). If any one of the two processes takes place, an application of the inhomogeneity will again allow the observation of escaping electrons by virtue of this repopulation by r.f. energy absorption of the previously empty top levels. By determining the two frequencies one can calculate

$$g/2 = \omega_s/\omega$$

As frequencies can be measured with extremely high accuracy, it is hoped that the value g can be determined to an accuracy of 1 part in 100,000.

An interesting feature of the scheme just outlined is that the experimental system as a whole is an artificially produced atom of macroscopic scale. The electrons are bound to it in quantized states and the transitions between these states are induced by macroscopic oscillators. It is remarkable that such a system will manifest physical properties usually associated only with phenomena on an atomic scale.

(*b*) The approach suggested by Crane *et al.*, in a much simplified version, is being followed in Edinburgh.[9] As in the experiment described previously, the purpose is to determine the variation of the polarization of electrons as a function of the number of orbital revolutions in a magnetic field.

To carry out the measurements we are developing a device schematically shown in Figure 10. It has been shown recently that electrons emitted in beta-decay of radioactive nuclei are longitudinally polarized. We put a sample (S) in a crossed homogeneous electric and magnetic field; E H. The sample holder is relatively thick, so that electrons are emitted in one hemisphere only. It is easy to show that if the electric field is weak enough, the electron orbits can be considered as circles drifting slowly in the direction of the x-axis. (The same type of motion occurs

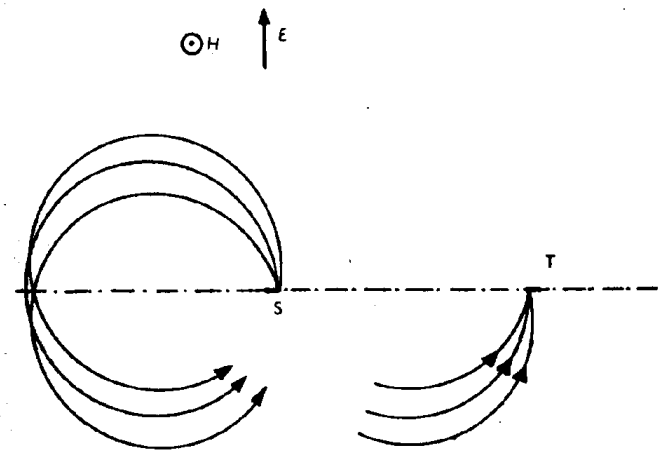


Fig. 10. In a crossed homogeneous magnetic (H) and electric (E) field, electrons emitted by source S reach target T after a large number of orbital revolutions.

in magnetrons.) The number k of orbital revolutions necessary to bring an electron from the source (S) to the target (T) can be varied in a wide range by varying E at a constant value of H. To detect the variation of polarization, electrons scattered from a thin gold foil into the two opposite directions perpendicular to the plan of the orbits are observed by scintillation counters.

By plotting the variation of polarization as a function of the number of orbital revolutions, the quantity $\frac{1}{2}g-1$ is obtained in the same way as in the case of the Michigan experiment.

* * *

It may be asked why all these efforts are made for the measurement of a small correction to a quantity that is altogether irrelevant in all practical applications of electrons.

The solution of fundamental problems of nuclear physics depends on how far we can explore the nature of elementary particles. The theory of the elementary particles and their interactions is, for the time being, very far from complete. Evidence concerning existing theories and hints for developing future

ones come most convincingly from experiments that yield clear-cut and highly accurate information of interaction phenomena of elementary particles. The direct determination of the magnetic momentum of elementary particles certainly belongs to this category.

REFERENCES

1. Born, M.: *Atomic Physics* (London, Blackie, 1953), Chap. 6.
2. Bohr, N.: *Quantentheorie und Naturbeschreibung* (Berlin, Springer, 1931).
3. Tolhoek, H. A.: 'Electron Polarization, Theory and Experiment', *Rev. Mod. Phys.*, 1956 **28** 277.
4. Dehmelt, H. G.: *Phys. Rev.*, 1958 **109** 381.
5. Kastler, A.: *Proc. Phys. Soc. A.*, 1954 **67** 853.
6. Crane, H. R., Pidd, R. W., and Louisell, W. H.: *Phys. Rev.*, 1953 **91** 457.
7. Crane, H. R., Pidd, R. W., Louisell, W. H.: *Bull. Am. Phys. Soc.*, 1958 **3** 369: Schupp, A., Pidd, R. W., Crane, H. R.: *Progress Report*, 1959.
8. Bloch, F.: *Physica*, 1953 **19** 821; Frisch, O. R., private communication.
9. Faragó, P. S.: *Proc. Phys. Soc.*, 1958 **72** 891.

RESEARCH REPORT

A. W. HASLETT

ROYAL SOCIETY EXHIBITS

THE annual conversazione held by the Royal Society on May 14, at Burlington House, London, provided its usual embarrassment of material and the usual difficulty that there is never enough time to look at it. A remarkable combined effort, needing time for study and a free use of illustration to convey the results, was a display of photographs illustrating the observational study of dislocations in crystals.

Beginning in 1949 with the discovery of spiral growth patterns in agreement with a theory originated by F. C. Frank and others at Bristol,* it carried the story through to the development of transmission electron microscopy to show dislocations in metal foils thinner than cigarette paper. The exhibits were the joint work of the Associated Electrical Industries Research Laboratory at Aldermaston, the H. H. Wills Physics Laboratory at Bristol University, the Crystallography Laboratory and the Department of Metallurgy at Cambridge, the Metallurgy Division of the National Physical Laboratory at Teddington, and Tube Investments Research Laboratory at Cambridge. Apart from transmission electron microscopy, the other principal development of recent years has been the use of X-ray microscopy to show dislocations in specimens which, with this method, can be somewhat thicker. As the programme summed it up: 'Direct observations of many phenomena such as dislocations in motion, cross slip, stacking faults, and dislocation loops produced by the condensation of vacancies during quenching have been made for the first time.' The bringing together of such an

* See, for example, the article 'Theory of Crystal Growth' by W. K. Burton in *Science News*, 1951 **21** 26.

exhibition for an all too brief appearance at Burlington House encourages the hope that some other body, perhaps the Institute of Physics or the Institute of Metals, might sponsor a repeat production with more space and a longer run.

Structure and Function in Bacteriophage

In the four-week interval between last year's May conversazione and its more social repetition in June, a group drawn from the Medical Research Council Unit for Molecular Biology at the Cavendish Laboratory, Cambridge, and the Electron Microscope Section of the Laboratory assembled an exhibit to illustrate some very recent work on the structure of bacteriophage, the bacterial viruses described by J. E. Hotchin in *Science News 33*.* The results of this research have still not been published, but they have been carried in the interval to the point when a plausible picture can be given of the mechanism by which infection is accomplished. This, as previously suggested, is by injection into the bacterium of nucleic acid from the bacteriophage. The exhibitors on the present occasion were Dr S. Brenner and Mrs L. Barnett of the M.R.C. Unit, and R. W. Horne of the Electron Microscope Section.

The bacteriophages in question are three genetically and structurally related viruses which grow on the bacterium *Escherischia coli* (formerly known as *Bacillus coli*), and are known respectively as T2, T4, and T6. They have 'molecular weights' of about 200 million, and contain roughly equal amounts of deoxyribonucleic acid and protein. Of these components, deoxyribonucleic acid is thought to serve the function in living cells of acting as a master template for the synthesis of proteins. Since infection of a susceptible strain of bacteria by one of these viruses consists in a usurpation by the virus of the synthetic mechanisms of the cell – the virus being unable to multiply on its own – it is evident that the entry of virus nucleic acid into the bacterium is the essential step in infection.

* J. E. Hotchin: 'Recent Work on Bacteriophage', *Science News*, 1954 **33** 89.

At the stage when Dr Hotchin wrote the article referred to, it was known that a bacteriophage particle possessed a head and a tail; that the deoxyribonucleic acid was in the head; that it was surrounded by a protein coat, and that the substance of the tail, too, was protein; and, finally, that it was the tail that adhered to the cell wall of a bacterium that was in course of being infected. It was therefore possible to imagine that, in some way, the tail penetrated the cell wall, and that some form of syringe action followed. But beyond this, knowledge of the structure of bacteriophage was rudimentary.

Last year's exhibit showed the complexity of the protein coat. It consists of a hexagonal head, enclosing the nucleic acid, and a tail which consists in turn of a core, a sheath, and tail fibres. This year's exhibit carried the matter a stage further – from structure to function. In the first place, the core of the tail is hollow and terminates in a hexagonal plate, to the six corners of which six V-shaped tail fibres are attached. Secondly, and more important, attachment of the bacteriophage particle to a bacterium is accompanied by a contraction of the sheath towards the head – thus propelling the tail fibres, as it is supposed, into the cell wall.

The nature of this contractile movement has been established. In isolated bacteriophage particles, not attached to a bacterium, exposure to hydrogen peroxide causes a similar movement of contraction. It appears, therefore, that contraction is brought about by a chemical stimulus, and the effect of this is to break the bonding through hydrogen atoms that, in the isolated bacteriophage, holds the sheath and core of the tail together. The identity of the natural stimulus remains a matter of speculation, as also is the mechanism by which the deoxyribonucleic acid is propelled from the head of the bacteriophage. What appears most remarkable at the present stage is the existence of a contractile mechanism, reminiscent of muscle, in a particle which, although organized in considerable degree, is still below the level of organization needed to be independently self-replicating and thus an organism. More than ever, the boundary between living and non-living is shown to be nebulous.

Poliomyelitis Immunization

A method of assessing the degree of immunity induced by inoculation against poliomyelitis was demonstrated by Dr D. G. Evans and Dr F. T. Perkins of the National Institute for Medical Research, Mill Hill. It consists essentially in observing the extent to which blood serum from the inoculated individual can be diluted and, having been mixed with virus and incubated, will prevent the destruction of monkey kidney cells, grown in test-tubes, when brought in contact with the mixture. Although more or less standard in its general approach, such a method has not before been developed for poliomyelitis.

Apart from the general desirability of being able to measure the immunity conferred in different conditions of inoculation, there was also a particular incentive to the research. Some years ago when the so-called triple inoculation was being developed – that against whooping cough, diphtheria, and tetanus – it was noticed that in a small proportion of cases symptoms of paralysis followed. The nature of this paralysis is still not understood. But, as a possible approach, it was thought that it might be an advantage to effect inoculation against poliomyelitis before the triple inoculation. Because of heavy infant mortality from whooping cough, this meant at as early an age as possible. In the interval, the material used in the triple inoculation has been improved; and, although there are still a few cases of paralysis, the risk remaining is small compared with the value of the protection conferred. Early immunization against poliomyelitis has, however, been found to be impracticable as a standard procedure. The reason is that, where passive immunity has been acquired from the mother before birth, effective inoculation is impossible until this temporary immunity has disappeared. Since two inoculations are needed to give a satisfactory degree of immunity (and passive immunity may last until about six months) it follows that, if immunization against poliomyelitis were given first, the child would be approaching eighteen months of age before the triple inoculation could be given. From this point of view, the most promising solution is prob-

ably to combine inoculation against poliomyelitis with the other three, and research with this object in view is being done in a number of laboratories.

A positive point that has emerged is that children or adults who have had a sub-clinical infection with poliomyelitis, and have therefore developed antibodies against the virus, respond particularly well to inoculation. Since, in times of epidemics, many more people carry the virus than show recognizable symptoms, this is an important fact to have established.

Wheat-Rye Hybrids

Ability to make use of wheat-rye hybridization to improve the hardiness of cultivated wheats and their resistance to disease has, up to now, been limited by inability to bring about recombination between their respective chromosomes. Research at Zurich, and by Dr R. Riley at the Plant Breeding Institute, Cambridge, has led independently to the discovery that one pair of wheat chromosomes – out of 21 pairs – act in some way as a barrier preventing association of wheat and rye chromosomes. If this pair of chromosomes is not present (and wheat stocks are available that lack it), association occurs, and there is thus a new possibility: that of introducing rye genes into wheat without having to be content with intact rye chromosomes, incorporated on an all-or-none basis. In effect, this means that rye can be used to improve the qualities of wheat with as much facility as if it were another wheat variety, but with the advantage that characters not previously present in any wheat can now be put there at will. The biggest prospect of improvement probably lies in introducing qualities of disease resistance of a different order from those found in existing cultivated wheat stocks. In this sense, the discovery may be correctly described as a break-through. But a long period of years will be necessary to exploit it, as with any major programme in plant breeding.

Super-graphite

Possible new materials based on graphite, and possessed of unusual electrical properties, were shown by Professor A. R. Ubbelohde, of the Department of Chemical Engineering, the Imperial College of Science and Technology, and his colleagues Dr L. Blackman and J. F. Mathews. The aim of their research is to realize the potentialities of a material which, in its individual crystallites, is at the same time a metallic conductor and a semi-conductor, according to the direction in which it is viewed. This property – mainly a theoretical suggestion up to the present time – arises from the structure of graphite. Any one crystallite consists of a series of parallel planes, each occupied by an extended hexagonal network of carbon atoms. Each atom in such a network is equivalent to any other, and within the plane of the network electrons are comparatively free to move. The spacing between planes is about 2·4 times as great as that between the carbon atoms within them, and the forces that hold one plane to another are correspondingly weaker. In a direction perpendicular to the carbon atom networks, the crystallite is a semi-conductor. While this is the theory of the matter, the extent to which this difference in properties with direction can be realized in actual graphite depends on the extent to which the individual crystallites composing it are mutually aligned. Most synthetic graphite shows little alignment. Natural graphite, selected from Ceylon deposits, is better. Synthetic graphite, made at the Imperial College by decomposing methane on a hot carbon filament, has been produced in films up to several millimetres thick, which show good alignment. This has been confirmed by X-ray diffraction pictures; and, in terms of electrical properties, resistance measured in a direction perpendicular to the planes of carbon atoms is 10,000 times greater than resistance along the planes.

Since oriented and normal graphite are equivalent to different elements in respect of their electrical properties, it is possible to make thermocouples incorporating junctions between them, and with these to measure temperature. Measurements up to 1,000° C were demonstrated, and curves of thermoelectric power have been obtained in the laboratory covering temperatures up to 2,100° C.

A further consequence of the layered structure of graphite is

that various compounds are formed in which atoms of a different element, for example potassium, are introduced between successive planes of carbon atoms. In the case of potassium, compounds with the otherwise incomprehensible formulae C_8K and $C_{16}K$ are known. Similar compounds are formed with rubidium and caesium which, like potassium, readily relinquish an electron; but compounds are also formed with, for example, bromine, which readily accepts an electron. Seemingly, the intervening graphite planes can either accept or donate electrons.

In order to study the electrical properties of such compounds, it is necessary to begin with oriented graphite. Indeed, it was largely for the study of these compounds that the Imperial College graphite was made. Some of the results obtained are rewarding. For example, the specific resistance of C_2K, in the direction of the graphite planes is actually less than that of nickel. The electrical resistance of graphite is also markedly reduced in compounds in which bromine atoms are introduced; and a demonstration was given of a cyclic change in resistance as bromine atoms were alternatively introduced into the graphite and removed from it.

HIGH PRESSURE CHEMISTRY

The success already achieved in the General Electric Laboratories, Schenectady, New York, in applying extreme pressures, combined with high temperatures, to make synthetic diamonds of industrial grade has been extended by H. M. Strong to a purely scientific study of the melting point of iron under a pressure within one order of magnitude of that estimated for the boundary of the Earth's core. The highest pressure used was 96,000 atmospheres, and at this pressure iron of 99.94 per cent purity melted at a temperature of 1740° C, with a probable error of 15° C. The pressure at the boundary between the mantle of the Earth and the core – the regions at about 1900 miles depth, at which earthquake records show an apparently sharp transition from a solid to a liquid or quasi-liquid state – is estimated to be 1,400,000 atmospheres, or about 15 times the highest pressure at which the melting point of iron was measured.

Extrapolation to this point was effected by making use of an equation published in 1953 by the late Sir Francis Simon. In publishing his equation, Simon pointed out that, while it had worked well enough when applied to substances of low melting point, notably helium, knowledge of the melting curves of metals was practically limited to those of the alkali metals – sodium, potassium, rubidium, and caesium – whereas even the heat of fusion and the volume change of iron at its normal melting point had only lately been measured reliably. Use of the Simon equation depends on fixing two constants, one of which at least is affected by the constitution of the substance in question. Simon applied his equation to iron, extrapolating from atmospheric pressure to a pressure more than a million times greater because of the interest of such an estimate for theories of the structure of the Earth. He wrote that the best way to arrive at a more secure prediction would be to carry the melting curve of iron to sufficiently high values to show the shape of the curve and to get an estimate of its curvature. This, he pointed out, would mean making measurements up to about 40,000 atmospheres. In spite of the pioneer researches of Professor P. W. Bridgman, at Harvard, he thought it evident that no such experiments could be expected 'in the near future'. Whether an interval of six years could be so described must be a matter of opinion. In any case, the upper limit of pressure has been extended to a level nearly two and a half times the figure that he asked for.

Through the melting point at atmospheric pressure, and from 10,000 atmospheres to 96,000 atmospheres, the curve follows a smooth course which, with a suitable choice of constants, fits well enough to the Simon equation. For iron similar to that used in the experiments, the extrapolated melting point at a pressure of 1,400,000 atmospheres comes out at 2,340° C, with a probable error of 50° C. Applied to the boundary layer of the Earth's core, and assuming this to consist of iron, Dr Strong increases the probable error to 150° C to allow for the effects of dissolved impurities and for a possible drift of the extrapolated curve from the true one. On this basis, the temperature at the boun-

dary of the Earth's core should be somewhere between 2140° C and 2540° C, provided that it is indeed iron that becomes molten at the boundary.

The arguments in favour of the core of the Earth consisting mainly of iron were summarized some years ago by Professor A. Holmes as follows: 'Partly because such a composition matches the high density of the Earth and the velocity of the P (seismic) waves through the core; partly because it corresponds with the composition of iron meteorites, which are probably samples from the interior of a disrupted planetary body; and partly because the presence of abundant iron oxides in the crustal rocks implies a high concentration of iron in the interior, just as in a blast furnace a little oxidized iron remains in the slag, while most of the iron sinks to the bottom.' Despite some more recent theories, this is still in essentials the most widely held view. On the other hand, it has been argued that, while the outer part of the core must be molten – because of the evidence from earthquake waves – the inner part of the core, under still higher pressure, may again be solid. Notwithstanding the geophysical interest of iron, it seems a reasonable guess that the investigation now reported is part of a wider programme in which measurements on other metals will be made.

F. E. Simon, *Nature*, 1953 **172** 746.
H. M. Strong, *Nature*, 1959 **183** 1383.

RADIATION HAZARDS

Hard on the heels of reports that the level of fall-out from nuclear test explosions had doubled within a year, the International Commission for Radiological Protection has presented its long-awaited recommendations on the 'maximum permissible genetic dose' – from all sources – applicable to large populations. Formally, there is no connexion. The basis of the Commission's report is an attempt to balance benefits against penalties; it is therefore irrelevant to the question of fall-out, except for those who see a benefit in test explosions, and the Commission, taking a long view, does not consider it specifically. Practically, there are good reasons for looking at fall-out

and total genetic dose together – and they are not the ones that are most often mentioned in public. Whereas there are overwhelming arguments, apart from fall-out, for wanting to put an end to test explosions, the argument from radiation damage points mainly at hospitals. This is well illustrated by a dilemma in which the members of the International Commission found themselves.

Starting from the position that any increased source of genetic dosage should be kept to as low a level as possible, consistent with benefits considered to balance the increase, the Commission would obviously have preferred to have arrived at an overall limit, applicable to all man-made sources of radiation and all countries, independently of the level of background radiation, which varies locally. They were prevented from doing so by two facts. First, they considered it their duty to make reasonable allowance for dosage resulting from an eventual big expansion in the peaceful application of atomic energy, both in electricity generation and through the use of radio-isotopes in industry. Secondly, they were aware that 'in some countries' – meaning, in particular, the United States – the increased genetic dose resulting from the diagnostic use of X-rays was estimated to be about 4·5 rems; that is, at as high a level as most geneticists would think should be accepted, or higher.

In illustration of this point, the Medical Research Council committee on the hazards of nuclear and allied radiations to man had suggested that such an overall limit – when arrived at – was unlikely to be greater than the natural background dose, and might be less; whereas the dose from background radiation in the United States is in the neighbourhood of 4 rems. Failing a figure in this neighbourhood, the Commission would evidently have preferred to take 6 rems as the absolute limit. But in that case 'the contribution from all sources other than medical procedures would be limited to 1·5 rems in these countries', which would 'impose unacceptable restrictions' on them. Baulked at this point, they fell back on the proposition that 'a genetic dose of 10 rems from all man-made sources is regarded by most geneticists as the absolute maximum and all would prefer a lower

dose'; and, for the purpose of recommendation, they decided to separate medical and other uses, a measure of long-term damage should be given by the sum of the individual dose multiplied by the number of individuals receiving it, reckoned up for different groups of the population. It is therefore necessary for practical planning to make some kind of apportionment between one type of exposure and another, while hoping that all sources of exposure will be kept to as low a level as reasonably possible. The Commission refrains from making any general recommendation, holding that individual nations are entitled to their own views on benefits. They make only an 'illustrative' apportionment. On this basis, they allow 1 rem for occupational exposure, 0·5 rem for special groups (for example, those living in the neighbourhood of nuclear plants who may be exposed to a higher than average dose), and 2·0 rems for the population at large, with a balance of 1·5 rems left in reserve. The allowance for occupational exposure is on the basis that as many as 1·7 per cent of the whole population might be in this group – or, considerably more, in effect, since, if individual limits are observed, which the Commission assumes, the average must be well below the limit, and in a well-run organization only a small fraction of it. The occupational allowance should therefore be conservative, and so, it may be hoped, are the other figures.

Radiation Protection: Recommendations of the International Commission on Radiological Protection. Pergamon Press.

DEW IN AGRICULTURE

Recent attempts to develop agriculture in very low rainfall regions have stimulated research on the water requirements of crops, and an interesting offshoot of this work is a renewed interest in dew, renewed because dew has been little studied by meteorologists since the classical work of Wells and Aitken in the nineteenth century. Agriculturalists want to know whether dew is a significant factor in the water economy of desert plants, and, if so, whether it can be exploited. Some workers claim that it can, and they produce striking figures for crop yields increased

by dew. But most plant physiologists, and also meteorologists, are sceptical of these claims.

If dew did encourage plant growth, one would expect to find that the amount of moisture condensing on leaves through the night would be comparable with the loss by evaporation on the following day. In England it is found that on a hot summer day a crop will transpire the equivalent of one to two tenths of an inch of rain; in a very dry climate this can be much more. But the maximum amounts of dew reported from many parts of the world, including the arid zones, are one to two hundredths of an inch per night, and this means that in fine weather dew evaporates within a few hours of sunrise. So, unless rainfall or irrigation has provided adequate soil moisture reserves, crops must endure severe water stress for the rest of the day.

Apart from the relative magnitude of dew and transpiration, it is important to know where the dew is coming from. Many people are inclined to treat with flippancy the question whether dew rises or falls, but, where growth is limited by inadequate soil moisture, it is a serious question for agriculture. If dew rises from soil beneath plants, soil moisture is depleted and the plants are no better off with water lying on their leaves as dew instead of wetting the soil round their roots. Only when there is a downward flow of water vapour from the atmosphere can dew make a real contribution to a plant's water supply.

Some years ago, an investigation on the dew that formed on a cricket ground in Middlesex showed that on typical dewy nights, calm and cloudless, there was so little air movement over the grass that virtually no dew fell from the atmosphere. But soon after sunset the grass was often covered with dew drops because water vapour was diffusing upwards from the soil. The relative importance of rising and falling dew may be different for other crops and climates, but this point has largely been overlooked in recent arid zone studies.

There is evidence that plants can absorb dew forming on their leaves; and it has even been shown that pine seedlings can absorb moisture from an atmosphere that is almost but not quite saturated – with no visible condensation on the leaf. There is

nothing biologically magical about this: from what is known of the physics of the system, it is what would be expected. The important point is that here again the rate of absorption is much less than the normal transpiration rate.

To sum up: if small amounts of dew can increase growth or even keep plants alive during a long rainless spell in an arid climate, the physiological mechanism has yet to be discovered. But further study may produce surprises, and few people today would care to repeat the assertion made by John Aitken in 1885, when he said: 'The immense amount that has been written on the subject of dew renders it extremely difficult for one to state anything regarding it which has not been previously expressed in some form.'

B.B.C. *Science Review* 1 July 1959.

SOME BOOKS RECEIVED

AN OUTLINE OF HUMAN RELATIONS by E. Chesser (London, Heinemann, 1959), pp. 446, 25s.

A century ago, western industrial man was safely installed as the lord of creation – the chosen of God's creatures. Man did not fit *into* Nature, he was its overlord. White-man's dominance over the coloured races was hardly in question, and the unfortunate conditions of many of his fellow-whites – in Manchester and Glasgow, in Rotterdam and Hamburg, could be attributed to the lack of abstinence, economic and moral, of these miserable creatures. Since Darwin's day, much of this has altered. Technological change has transformed the Earth, but catastrophic wars and continued threats of war emphasize man's continuing close relationship with the lesser brutes of the Earth; subject races see their future in their own hands rather than in those of a beneficent overlord; by comparison with present-day statistics, Victorian child mortality figures bring a blush of shame, yet stable human relationships in the form of Christian marriage are, it often seems, in jeopardy; psychological medicine is within sight of offering help to the mentally sick and the sexually aberrant.

But man, having made a valiant attempt to conquer the Earth, and essaying even space beyond, has still much to do in the conquest of man. If he is to survive he must make himself the object of his study just as much as he has turned his attention to the material and biological world in which he lives. This study is the topic of Dr Chesser's important synthesis – the relationship of man to his fellows, black, white, and khaki, to society in general, and also, philosophically, to the forces of the whole Universe.

Time is short. Only when mankind comprehends his own nature and achieves a sympathetic understanding of the problems and aspirations of his fellow men, can he rightly use the devastating powers that twentieth-century science and technology have put at his disposal. Towards the study of mankind by man, Dr Chesser's book is a significant step. If it were included among every twenty-year-old's birthday presents, it would contribute materially to an increased hope of a happier and more stable future for mankind.

THE NEW MATHEMATICS by I. Adler (London, Dobson, 1959), pp. 187, 18s.

There must be many of us who have been able to take our mathematical studies so far, but who have utterly failed to advance beyond that point. Sometimes the fault is in ourselves – there are convincing arguments that mathematical ability is a matter of inborn genes – one either 'sees' a mathematical argument or one doesn't – but, at the same time, a serious charge can not infrequently be laid at the door of poor, unimaginative, and often just downright bad, teaching.

It is, therefore, a pleasure to be able to welcome *The New Mathematics*. To anyone familiar with the elementary algebra and plane geometry taught at school, it will reveal that in what is already known by most adults who have had a grammar-school education lie the concepts of advanced aspects of the subject: group theory, ring fields, vector spaces, homo-, iso-, and homeo-morphisms.

Anyone prepared to work through the volume, and in the reviewer's opinion this isn't difficult, can enter this fascinating world where two and two sometimes make, and sometimes do not make, four. Sixth-formers could use it as a bridge to the systematic study of advanced texts, and teachers could use it to give a fuller flavour to the curriculum. At its level, *The New Mathematics* is one of the most stimulating books that has come to the notice of the reviewer for some time.

WHAT IS CYBERNETICS? by G. T. Guilbaud (London, Heinemann, 1959), pp. 126, 10s 6d.

G. T. Guilbaud is Director of Studies at L'École Pratique des Hautes Études in Paris, and is well known on the Continent as an authority on Cybernetics. Now, in the short space of 120 pages, he has produced a remarkably clear introduction to his subject, intended, in the main, for non-specialists in the subject – the treatment is largely devoid of mathematics – but, at the same time, its crystal-clear logic will make it a valued guide even to those professionally connected with the many ramifications of this fascinating subject. He begins by clearing away not a few misunderstandings about the origin of the word itself, showing that, in his *Philosophy of Science*, Ampère used the word Cybernetique a good century before its renaissance on the other side of the Atlantic. Successive chapters deal with circuits and networks, feedback and purposive activity, the measurement of information, imitative gibberish, information and probability, and communication. The concluding part is called 'towards unifications'.

The translation is by Valerie MacKay, whose husband Dr D. M. MacKay, a leading exponent of the subject in Britain, supplies a preface. Dr MacKay's assessment is worth quoting. He says: 'M. Guilbaud's little book has one exceptional merit. It is a genuine effort to portray Cybernetics as a whole – as a phenomenon, one might say, and not just as a single technique or even a collection of techniques. . . . Cybernetics, as he uses the term, is essentially a new way of thinking about the organization of action – a new inter-disciplinary language stationed at the cross-roads of the older sciences and capable of performing unique, if modest, services there.'

T. H. HUXLEY: SCIENTIST, HUMANIST, AND EDUCATOR by C. Bibby (London, Watts, 1959), pp. 330, 25s.

Anyone wandering around South Kensington today cannot but notice the surge of construction and reconstruction that has been necessitated by the fresh realization of the vital importance to Britain's survival of sound technological education. But many whose knowledge of the history of technical education is less complete than it might be will perhaps fail to see, in all this building, phase two, as it were, of schemes set afoot a hundred years ago by one of the most remarkable of Victorian pioneers, T. H. Huxley, the subject of Dr Bibby's new biography.

It is more than half a century since we had an appraisal of Huxley's status, and as the years have passed he has increasingly been seen in the excessively narrow context of the champion of evolutionary theory – 'Darwin's bulldog' – rather than as one of the really remarkable minds of his century. Not only is it healthy to have this imbalance redressed in the centenary year of the *Origin of Species*, but one cannot fail to sense in reading these pages that where Huxley was a hundred years ago, we are again, only more so. Huxley's tiny School of Mines became the Royal College of Science, and today we are witnessing the greatest-ever expansion of the institution that has stemmed from it. Huxley saw the foundation of the School Boards. These came and went, but is all well with their successor? Do not we need a new ethos in education? His activities during his incumbency of the Lord Rectorship of the University of Aberdeen – an honorary post of considerable influence in the patronage of the

student body – helped to establish the teaching of science in the Scottish Universities. Today we can see rifts in the walls of the humanists' preserves. Where is the Huxley who will stimulate a new flowering of British genius? Cast in the role of Darwin's bulldog, much of Huxley's philosophy of life has been obscured in recent years. We are, therefore, in debt to Dr Bibby for this scholarly essay, so much of which is pertinent to mid-twentieth-century problems. The volume includes a select list of Huxley's publications, a tabular conspectus of his life and times, extensive references, and an adequate index.

THE STRUCTURE AND EVOLUTION OF THE UNIVERSE by G. J. Whitrow (London, Hutchinson, 1959), pp. 212, 21s.

It is just ten years since Dr Whitrow wrote his *Structure of the Universe*, but so much has happened in the ensuing decade in the astronomical and cosmological fields that, when the customary new edition was due, it became evident that it was not a new edition that was called for but substantially a new book.

On the observational side, not only has the great 200-inch telescope at Palomar Mountain been commissioned, but a whole new science, radio astronomy, has come into being. And so far as theory is concerned, with advancing knowledge, attention has been directed to the origin of the elements and to problems of world evolution. These changes account for the modified title of this volume which now covers the observations that take us into the depths of the universe, space time and relativity, world-models, the structure of galaxies, and the evolution of the universe.

The new work is written with the penetrating insight that we have come to expect from Dr Whitrow, and those who seek a perspective on cosmological themes will find that the historical aspects of the subject are given adequate treatment. A short appendix discusses why physical space has three dimensions, and there is a useful bibliography and index.

THE SIXTH SENSE – AN INQUIRY INTO EXTRA-SENSORY PERCEPTION by Rosalind Heywood (London, Chatto and Windus, 1959), pp. 224, 21s.

All of us have had reports of phenomena that cannot be explained in terms of the accepted laws of science; some of us, indeed, may have

direct experience of them. So far as the second-hand reports are concerned, there is always in the mind of the sceptical scientist the feeling that, were all the facts known, the 'inexplicable' would take its place in the accepted scheme of things. It is, therefore, useful to have from Mrs Heywood what amounts to a concise history of the activities of the Society for Psychical Research. Here in readily accessible form are the experiments of Sidgwick and Myers, the telepathic investigations of Gilbert Murray, accounts of mediumship subjected to stringent test conditions, and the experiments in extra-sensory perception associated with the names of Rhine, Soal, and others.

Where the 'explanation' of these phenomena lies is a matter of speculation, but a dispassionate perusal of the author's pages leaves no doubt that here is something that merits the most careful consideration.

ABOUT OUR CONTRIBUTORS

J. M. BARRY read chemistry at Oxford and biochemistry at Cambridge, finishing in 1948. Since then he has done four years of biochemical research at Chicago, California, and Johns Hopkins Universities, and is now a lecturer in biochemistry in the Department of Agriculture of Oxford University. His research has been mainly on protein biosynthesis, particularly in the mammary gland. He has previously contributed to Penguin *New Biology 24*.

P. COOK is a graduate of Glasgow University. For six years he was with I.C.I. (Explosives) Ltd at Ardeer, Ayrshire, where he helped in the manufacture of carbon compounds with co-valent bonds which readjusted themselves with extraordinary rapidity at suitable moments. After the war he turned to teaching and is now a science master at Aberdeen Grammar School. He has written a number of feature programmes for the B.B.C., generally of a scientific nature; his latest script – intended to mark the 'launching' of the fast reactor at Dounreay – comments on the origins and development of work on atomic energy.

P. S. FARAGÓ was educated at Budapest University, Hungary, and obtained the Dr Phil degree in 1940. He started research at the Tungsram Research Laboratory, Unjpest (nr Budapest). In 1948 he became reader in Physics at the Budapest University, and from 1951 he was also the head of the Department for Electromagnetic Waves in the Central Research Institute of Physics of the Hungarian Academy of Sciences. Since 1957 he has been senior lecturer in Natural Philosophy at the University of Edinburgh. His field of research is electron physics.

J. MOIR, M.I.E.E., was till recently Head of the Communication Section in the Electronic Engineering Department of the British Thomson Houston Company, Rugby. He is now Technical Director of Goodmans Industries, Wembley. He is chairman of the Audio Engineering Panel of the Institution of Electrical Engineers and author of the standard textbook on sound reproduction, *High Quality Sound Reproduction*, published by Chapman and Hall.

J. A. MOORE, who is 33 and married, began his career in lamp research in 1943 as a student assistant in the Research Laboratories of the General Electric Co. Ltd. After three years' part-time study

and work on the development of the fluorescent lamp, he was called-up and spent his period of service as a radar instructor in the R.A.F. On an F.E.T. Award he then studied full-time in London University at Sir John Cass College, where he served as President of the Students' Union from 1950–51, graduating B.Sc. with honours in mathematics in 1952. He then rejoined the lamp research group at the G.E.C. Research Laboratories, where he has been investigating incandescent light sources, including the many and varied problems connected with the continual improvement of the tungsten lamp.

J. R. P. MONKMAN was born in 1920 in Yorkshire. He is married with one infant daughter. Educated at Retford school (Notts) and London University, he joined the London laboratory of one of the railway companies in 1938. During the war, from 1940, he served as surveyor with Royal Artillery (Field), later in North Africa and Italy. In 1943 he transferred to the R.A.S.C. as assistant War Department Analyst and remained there till demobilization in 1946. He then rejoined British Railways in Yorkshire as a water treatment chemist. Subsequently he spent eight years as an analyst with a Yorkshire firm of textile dyers and finishers, based in Lancashire, and recently has been textile chemist with a Textile Research Association. Interests: motoring and photography.

C. H. WADDINGTON, F.R.S., is Professor of Animal Genetics in the University of Edinburgh. He began his scientific career with a study of evolution among fossils, but soon passed from geology to biology, becoming particularly interested in the physiological mechanisms of embryonic development and their relation to genetics. He is the author of several books, including the Pelican *The Scientific Attitude, Principles of Embryology, Strategy of the Genes*, etc.

CONTENTS OF RECENT ISSUES