THE RADIO AND ELECTRONIC ENGINEER

The Journal of the British Institution of Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

VOLUME 27 JANUARY 1964 NUMBER 1

The Presidential Address of John Langham Thompson

Delivered after the Annual General Meeting of the Institution on Wednesday, 27th November 1963

The growth and stature of the Institution owe much to the leadership of the men who have been my predecessors in this office. I am therefore very conscious of the responsibility that now rests upon me to continue their work—the more so in view of the distinguished Presidency that I am to follow.

The Institution now enters its 39th year and its own history reflects the history and growth of one of the world's most important technologies. When we were founded in 1925 the British Broadcasting *Corporation* was not to come into existence for another year, and the industry whose science and technology we represent, the industry we now call the Electronics Industry, was not an industry at all. It numbered barely as many companies (large and small) as the then new Institution had members.

Today, the electronics industry is named by the National Economic Development Council as being one of those industries whose continued growth is vital to the development of the national economy. The annual value of its production is approaching £400 million and its contribution to our export trade some £80–90 million. A recently-published White Paper[1] states that the electronics industry employs not less than 300,000 people, or 3·7% of the total labour force employed in the manufacturing industries of this country. Our industry is shown to employ 7,500 qualified scientists and technologists—equal to 9·4% of all the scientific and technological manpower engaged in the manufacturing industries of this country; and to have forecast that, by 1965, it will need to increase the total number employed from 7500 to 10 500.

The Institution and the electronics industry have grown up together. Each has made a contribution to the other. The professional electronic and radio engineer gives to industry the technological support essential for the industry's well-being, and the electronics industry gives to our members the facility for the exercise of their intellectual capacity. We are interdependent, one upon the other, and the skill and resource of the one reflects the skill and resource of the other.

We both owe much to the vision of the scientists on whose work our present technology is founded; in particular Clerk Maxwell. The great contribution of Clerk Maxwell to radio and electronic science is honoured in the Institution's armorial bearings and by the inauguration by the Institution of the Clerk Maxwell Memorial Lectures.

We also both owe much to the imagination, courage and determination of the relatively few men—industrialists and technologists—who brought about our separate but interdependent existence.

Pledged, as our founders were, to the principle that education must be kept abreast of new techniques and research, we have provided through our meetings and publications a platform for the dissemination of information, and throughout our history we have landmarks showing the Institution's influence in bringing into focus matters demanding urgent attention. An example of our forward thinking is the recent production of the report by our Research Committee. This is no longer the responsibility of the Institution and I will therefore not expand on the importance of the National Electronics Research Council. I welcome, however, this opportunity to echo and convey on behalf of our members, firstly our thanks to our retiring President for his initiative in starting this enterprise, and secondly to express good wishes for the success of the National Electronics Research Council under his chairmanship.

Electronics Research

Lord Mountbatten's conception of bringing together Government interests, education and industrial enterprise into one Council has reminded me of a couplet from Robert Browning's *Andrea del Sarto:*

> . . . a man's reach should exceed his grasp,
> Or what's a Heaven for? . . .

The Council of N.E.R.C. has already formulated its main object, and it makes stirring reading:

> 'In the public interest, to carry out, encourage and co-ordinate research in the fields of radio and electronic science and engineering.'

By the term 'co-ordinate' I hope that the Research Council will not seek to interpret this as the need to avoid *all* duplication of research effort. No one person, company or even country can have a monopoly of intelligence. It is in the nature of things that only with hindsight can we perceive in an individual that extraordinary capacity for imaginative creation, original thought, invention or discovery that we call genius—and genius does what it must, not what it is co-ordinated to do.

Now it could be argued that we are an Institution of engineers not of research scientists; that we should concern ourselves more with the *application* of new knowledge than with the *acquisition* of new knowledge; that research needs to be *exploited* not *expanded.*

But would this be true?

Before we can answer this question we must first define what we mean when we use the word 'research'.

Research is a generic sort of word the meaning of which changes according to the context in which it is used or according to how the person using it wishes to be understood. Quite often, and quite improperly, the word 'research' is used as a synonym for development.

The 'Trend' Report[2] defines scientific research carried out in universities as:

> '. . . the pursuit of knowledge without any necessary reference to practical application or utility',

whereas the Finance Act in specifying the depreciation that is allowable, on capital equipment to be used for scientific research, says:

> 'Scientific research related to a trade includes any scientific research which may lead to, or facilitate, an extension of that trade.'

There is no sharp dividing line between pure basic research carried out solely in order to increase our knowledge of the material world; and applied research which has as its object the obtaining of a practical and material goal which can be reasonably well specified. Each of them is involved to some degree in development, if only in the building of prototype plant and equipment.

For example, to enable the next significant step forward to be taken in exploring the structure of matter, the European High Energy Physics programme has proposed that a new proton synchrotron should be built. This is estimated to cost £120 million over an $8\frac{1}{2}$-year construction programme. The design study is estimated to cost £320,000 and the complementary equipment a further £21 million!

The 'Zuckerman' Report[3] whilst emphasizing that there is, and can be, no clear line of demarcation between one form of research and another, defines five categories of activity normally included under the portmanteau phrase 'research and development'.

Pure Basic Research—carried out solely to increase scientific knowledge.

Objective Basic Research—carried out in fields of recognized potential technological importance but only in those areas where existing scientific knowledge is seriously insufficient.

Applied Project Research—which has as its object the attainment of a practical goal which can be fairly precisely defined.

Applied Operational Research—directed to improve the use of an existing process or piece of equipment.

Development—which bridges the gap between research and production.

Using these definitions of research there surely can be no doubting that the Institution must be *very much* concerned with research, particularly in the fields of applied research and development and in the field of objective research. For these are the fields of technology and advanced technology.

The Importance of Technology

We live in a technological age and whether we like it or not there is virtually no escape. The everyday life of each one of us is encompassed by an artificial technological world on which we depend for our very existence.

Could any of the population of the developed countries of the world exist for very long without gas, electricity, drinking water, sewage disposal, lighting, heating and refrigeration, and without its present systems of transportation, and communications, or without the essential drugs, antibiotics, vaccines, medicines and the modern methods of surgery?

We do not express wonderment that these services exist; only horror and fear if they are absent, or apprehension and anger if they are even temporarily suspended.

Collectively, all these services and things form an essential part of what we call our 'standard of living'. The standard of living in any country, however, is determined by the state of that country's national economy. Our own standard will be determined by the effort that we make to increase our productivity, for only by increased productivity will we be able to afford to enjoy any improvement in our standard of living.

There is no lack of technological invention to hinder increased growth; what has been lacking has been the acceptance by industry of the contribution that technology can offer and the realization of the potential magnitude of that contribution.

This unawareness of the importance of technology has extended also into the fields of higher education. From the 'Trend' Report[2] we learn that at Oxford and Cambridge Universities all but 14% of the students were studying arts or pure science; that six established universities have no faculty of technology, and that taking universities as a whole, 25% of the students were reading pure science and only 15% technology.

Only in Britain do we find this emphasis on pure science and lack of emphasis on technology. In all the other great industrial nations of the world the position is the reverse, more students qualify in technology than in pure science.

Equally disturbing are the published findings[4] showing that apart from a lower number of entrants, the intellectual quality, judged on G.C.E. 'A' level results, of students entering for technology is lower than those entering for pure science.

The combination of all these factors has combined to produce a serious shortage of trained engineers in all disciplines,[1] and especially in our own field of electronics, which will certainly persist until the 1970's.

We must not be surprised at this reaction in favour of science and against technology. We use the word 'science' too much and the word 'technology' too little. In the last decade almost every major technical achievement has been popularly hailed as a triumph of science—space probes, manned satellites, communication satellites, nuclear power and nuclear submarines.

These achievements fired the imagination of the whole world. How much more so must they have fired the imagination of those in the fifth and sixth forms who are now the science students of today. If this was science it was exciting, adventurous and satisfying.

There was no authoritative body that could emphasize to them that this was more truly technology; that technology, bounded by the natural laws formulated by the scientist, had provided this physical reality.

The truth is that science and technology are part of the same spectrum of activity. On the one hand, pure science, the search for the acquisition of new knowledge and on the other, technology, the application and the exploitation of that new knowledge.

Science and technology are, and always will be, interdependent one upon the other. As the search for new knowledge progresses the equipment needed to conduct that research becomes more and more complex and the scientist must turn to technology to provide him with that equipment. The new proton synchrotron that I mentioned earlier is an example.

Action must therefore be taken quickly, to impress upon those who will be the students of tomorrow, and upon their parents, that technology can offer work that is as exciting, as rewarding and as challenging as any work offered by pure science, and that in advanced technology, such work often approaches close to the very frontiers of existing knowledge.

The Profession of Engineering

Even more important is the urgent necessity to enhance the status and the public image of technology and the technologist, as a profession, for the teacher, the parent and the friend exert considerable influence on a student's final choice of career. This is really the key to all the recruitment problems that now face us.

By himself the individual professional engineer can do little to bring this about, but his professional institution, voicing the collective views of its membership, can do much. It can command attention because it speaks with professional authority: the authority of its membership.

Our own Institution has always both accepted and discharged this responsibility to its membership: it has played a prominent part in the establishment of courses suitable for the proper education and training of the professional electronic and radio engineer; its vigorous proposals for the establishment of departments of electronic engineering in more of our universities is well known; to each of the many committees and commissions that, both during and since the war, have been set up to investigate and report upon the education, training and utilization of the scientific and technological manpower in this Country, our Institution has made valuable contributions. The Institution has also played a prominent part in fostering these same ends in the Commonwealth countries.

In the interests of the disciplines that they represent the professional engineering institutions must continue to act independently. But in finding the solution to their common problem they must, in the public interest, now all act in unison. And I emphasize *act* and *now*.

Consultation between the separate institutions will never achieve the positive results that are now required. It is therefore essential that the individual institutions be part of an overall organization in which we may all discuss our common problems and pool our ideas. Such a part is now given to us with the formation of E.I.J.C.—the Engineering Institutions Joint Council, shortly to become incorporated as an independent body with a title more suited to its declared objects.

Agreement has already been reached on a common Part 1 Entrance Examination and discussions are now centred on a common standard for Part 2. I emphasize a common *standard* of examination, not a common examination. With thirteen diverse engineering disciplines I do not believe that it would be possible to agree a common examination.

I believe there is now the urgent need for the Council of this new body to agree a common policy and initiate immediate action—*to enhance the public image of technology as a profession.*

There is the urgent need for an agreed approach to all schools at sixth-form level and to technical colleges, on the many different disciplines that technology embraces and of the scope and breadth of each of these individual areas. It is equally important that these facts should be made known to 'careers' masters.

A 'Careers Guidance' handbook, attractively presenting this information, must be produced and speakers of the different disciplines should be encouraged to lecture upon their particular speciality.

In addition there is the need for an approach to the universities to seek their help in the encouragement of post-graduate research in the fields of technology. Post-graduate research of this nature could well set a pattern that other students would wish to follow.

The Christmas lectures that have been started by one or two institutions could not only be more widely copied but shared in by all institutions. Our Institution in particular could make a spectacular contribution to such an occasion!

There are other and equally important tasks awaiting action by this new body:

It must concern itself with persuading industry to give more encouragement for part-time day course education; and in allowing more of their suitably qualified and brightest young men to be released for further education by way of sandwich courses. And not only for education but also for re-education. All too often the senior engineer is expected to keep abreast of all new ideas, and the application of these new ideas. We all know that it is difficult enough to keep abreast of events that happen in our particular speciality and we all know that all too often we find ourselves blissfully unaware of a significant development in another speciality. I believe that short courses, lasting perhaps a few days or at the most a week, could do a great deal to keep the senior engineer up to date with current developments and ideas, their fields of application and their potential.

There is the need also for both education and re-education in the science of management, in particular for the many scientists and technologists who are employed in management, for good scientists and engineers do not necessarily make good managers.

The authority to manage is quite distinct from the ability to manage. Management is an art as much as a science—it is human beings that have to be managed—and both the art and the science need to be learned; experience is a poor substitute for proper instruction, for not all of us benefit from our 'experience'.

For those experienced in the art and science of management, there are new scientific techniques of management and control to be learned; critical path analysis, network analysis and the other offsprings of PERT—programme evaluation review technique.

I believe there is also the need for E.I.J.C. to maintain a central register of *all* professional engineers, but a register on a more informative basis than the old Central Register once maintained by the Ministry of Labour.

This new body could also do much to promote better understanding between the members of the thirteen Institutions, of the trends of progress and invention in disciplines other than their own, and in areas of common interest.

Co-operation with the Institution of Electrical Engineers

I have in mind an extension of the collaboration that now exists between our Institution and the Institution of Electrical Engineers: collaboration that I welcome and hope to see extended.

Both Institutions have their individual fields of main interest but the boundaries are not clearly defined and there are many overlapping areas of common interest. It is encouraging to see that in these areas there is full recognition by both Institutions that in the public interest it is their duty to work together. To this end the two Institutions collaborated with the interested American Societies in the organization of two very successful International Conventions recently held in London: on television and telemetry.

The Medical Electronics Groups of the two Institutions now hold their discussions and meetings jointly and it is very encouraging to see at these meetings representatives of each of the branches of the science and practice of medicine.

The advances in pure science can be accepted or disbelieved; ignored or not even understood. But the advances in technology are not abstract reasoning. They are real; they exist; they are there for everyone to see and not everyone is pleased.

Every major advance in technology has meant change; and change from the established order arouses hostility.

Always has this been so and our recent history abounds with examples, the technical advances that brought about the industrial revolution; the coming of the railway, the motor car, the aeroplane; radio broadcasting, talking pictures and television. All these things and a thousand and one others are today a familiar part of our everyday pattern of living. But whilst they were *becoming* our pattern of living they aroused hostility from those who felt that their security had been threatened, and scepticism, honest doubt or disbelief from those who through ignorance or self-opinionation, did not understand their application and potential.

Is there any reason to believe that the new changes we are about to experience will not arouse these same human emotions?

We are all models of objectivity when change affects the other man's job or his pattern of living. But when change affects our own job or pattern of living, or we think it will, objectivity is immediately replaced by extreme subjectivity; we think only of how this change will affect us, and we give no thought to whether this change is essential or necessary or desirable.

Here then is another field of endeavour that the new E.I.J.C. must explore; to find ways and means of communicating with the people who will use these new machines, equipments, methods and techniques; for stripping away the mystery of these 'new developments' and for explaining in simple terms their operation, application and potential. I believe this to be the responsibility of the new body, for some of the first people to be convinced of the use and application of these new developments will be the professional engineers of disciplines other than the disciplines inventing them. Our own Institution can make an important contribution in this endeavour, for the application of electronic devices, systems and controls will be entering into almost all of the disciplines of technology.

Commonwealth Cooperation

There is the need also for concentrated action in the Commonwealth. Individually, the Chartered Institutions are not able to do a great deal. Working together through E.I.J.C. might well make possible the setting-up of the necessary machinery in the various centres of the Commonwealth to stimulate professional activity and allow for the reading of papers and the holding of discussions.

Particularly is this important in the developing and emerging countries of the Commonwealth. Only by the knowledge and application of technology, in all its disciplines, can their economic growth be assured. National prestige may well call for a national airline, but economic prestige, the prestige arising from a thriving national economy, from the improved standard of living of their people, will call for technology, the technologist and the technician.

The Engineer and the Technician

Professional engineers rely to an enormous extent upon the technician. The two fill complementary roles. The technician may sometimes think that he can 'get on' just as well without the professional engineer, but the professional engineer *knows* that he could never 'get on' without the technician.

What do we really mean by the word 'technician'. Much has been written on this subject: the 'Crowther' Report[5] defined some of his duties; the Commonwealth Engineering Conference defined others and introduced the term 'engineering technician'. Our Institution has written on the subject and so have other Institutions. Much has been written—but not all of what has been written has been read!

Let us simplify the problem and so make the remembering easy. Less than ten years ago, this Institution, and I think I am right in saying the I.E.E., still regarded the Higher National Certificate as a sufficient educational test for Associate Membership. Since then, the standard of the H.N.C. has not appreciably altered, yet it is now thought of as a suitable qualifying examination for the technician grade.

This is the standard we hope to see achieved by all technicians. But between that goal and the achievement of that goal there must be at least one, possibly two, intermediate grades. And this is one of the problems. We talk of the 'technician grade' but we do not distinguish between the levels of academic achievement in that grade.

Our own Institution has always supported the technicians of our own discipline. The help and support it has given to the Radio Trades Examination Board is an example. As its name implies R.T.E.B. is not a Technician Association, it is a qualifying body.

The Certificate that it awards to students who satisfy its examination requirements is nationally recognized. It has done much in formulating standards of competence and in the establishment of educational courses leading to the achievement of those standards. In so doing it has provided a valuable service. But a valuable service only to a small sector of the electronics industry. All its energies and abilities have been confined to the establishment of a recognized qualification for those engaged in the servicing of domestic radio and television equipment and, more recently, the maintaining of electronic equipment other than radio and television.

I am glad to report that recently our Institution and the Institution of Electrical Engineers exchanged views on the formation of an association to look after the wider interests of the engineering technicians employed in the electrical, radio and electronics industries.

But there is now the need also for the formation of a central body representing the interests of all the engineering technicians: an independent body representative of the whole field of engineering technical endeavour, with a policy-forming Council consisting mainly of technicians and financially supported by the various Technician Associations that have already been formed or are to be formed: a central body with a common policy that can speak with authority on behalf of all the engineering technicians.

Now I would like to voice a personal opinion on the aims and objects of such a body. I emphasize personal opinion because I would not want it to be thought that I am expressing the Institution's views, nor the view of the E.I.J.C. And I realize also that it will be the technicians themselves that will frame their own aims and objects.

I think that the main function of this Council of Engineering Technicians should be:

(i) To act as an examining and certifying body with grades to indicate the several levels of educational attainment.

(ii) To act with educational authorities, to stimulate, organize and constantly review, suitable courses in technical colleges in preparation for its examinations.

(iii) To co-operate with industry in bringing about suitable training schemes in industry. And to convince industrial and educational authorities of the need for organized refresher courses at, say, five-year intervals.

We must resist any tendency to wish to see this new body closely following the pattern of a professional institution. Once formed, it must be allowed to evolve in its own way, but our own Institution could offer much help; particularly in advising on educational matters in the field of electronic and radio engineering.

One thing is already obvious. Namely, that technicians will be required to understand more of physics. Think of a technician trained, say, in 1940 having to maintain, service, or play a part in the production of, or modification to, lasers and masers!

Electronic and Radio Engineering Tomorrow

Perhaps we could give some guidance by looking into our own future; the future trends of Electronic and Radio Engineering.

The rapid progress in solid-state devices, microcircuits, and solid-state modules, foreshadows the time when we will cease to think in terms of independent lumped constants (L, C, R and other parameters) but will come to accept complete circuit modules. The line of demarcation between the active component and the passive component is already vanishing and the functional module is now in common use. By microcircuits I mean only thin-film or semiconductor integrated circuit techniques.

Micro-circuits should allow of much greater flexibility in the design of special purpose equipment and give to the designer the opportunity to optimize his design unencumbered by the limitation of standard components. Their use should allow him to reduce both size and weight and, most important of all, to allow of greatly increased reliability. Not only are the number of soldered joints greatly reduced, but the reliability of a complete integrated circuit approaches the reliability of a single transistor. Because of these factors of reduced weight and size, their adoption also allows much greater latitude in the use of redundancy techniques, so increasing reliability still further.

It would seem, therefore, that circuit designers will be divided into those who evolve these circuit units and those who apply them to meet their particular requirements. The equipment designer will be able to concentrate more of his attention on improving the specified performance, and the time lag between satisfactory functioning of the laboratory model and the commencement of production should be appreciably shortened.

Another important field is data recording and data processing. Until recent years the only way of preserving data was by pen and ink type records or by photographic film. Unfortunately neither of these methods was suitable for subsequent rapid data processes. I say 'unfortunately', because a piece of photographic film is really quite a remarkable information store: 500 million bits—in a nanosecond—with an entirely unskilled programmer—for sixpence!

In a short span of time very considerable progress has been made in the development of other forms of storage circuits. In the attempt to reconcile the mutually incompatible parameters of cost, speed, access time and volume—and possibly patents—the current designs of storage circuits display a remarkable admixture of statics, magnetics, electrics, electronics, pneumatics, mechanics, refrigeration, imagination and determination.

We now see a great number of 'instant stores' all capable of yielding their data back into electrical time, and many able to be addressed and interrogated at will. All this has greatly increased the scope of dealing with data. Not only in the obvious application to computers, but in an almost infinite variety of industrial, commercial, military, scientific, medical and communication applications.

Information can be recorded as soon as it arrives. Processed into a form suitable for further study and then compared with some previous standard, or with some data associated with another time, place or parameter. If the result is significant it can be subjected to further processing or used for some other purpose. If the data is judged, however, to be of no value, it can be discarded. A surprising amount of recorded data is, after comparison, found to be redundant and the facility to discard this redundant data increases the speed of access by reducing the volume of the store.

Data recording and processing will find increasing application in complex installations such as large industrial and communication networks. Allied with computers they will also enable machines to 'learn' quickly, to profit by their mistakes, and to be able to forecast the situations they will have to deal with. This 'control' function is what we really mean when we talk of automation — for automation is not mechanization—automation is control.

One of the situations that we as engineers have to deal with is how to keep abreast of the innumerable technical developments that are daily being reported in the technical and commercial journals of the world —and here we are not in control.

In any new development project it is always the intent to 'start off' where the other man 'left off', but because of the difficulty of finding out where all the various workers in a particular field 'left off', and of work done in peripheral fields, this intent is more often honoured in the breach than in the observance. As the speed of development increases so will this problem increase. A recent White Paper[6] refers to this problem and to the preliminary steps that have already been taken to resolve it.

I believe that improved methods of documentation *must* be introduced and that there is a long overdue, and now urgent, need for a Government operated data processing centre to provide at least in abstract form and in English, the main contents of the technical publications of the world. Of course there are many problems and very considerable capital expenditure would be called for, but I believe that, if such a service were available to industry on a rental or a fee paying basis, the return would justify this. Certainly the need justifies it!

In the computer field an analogous position exists. There is no lack of sophisticated ideas and equipment, only a lack of people who can decide *what information* to feed in and *how* to feed it in; of people who know how to plan and use programming. And, I might say, a lack also of people who really know what they want *out*. For on what is wanted out depends what must be put in and how it is put in.

We as an Institution must do all that we can to stimulate interest in systems and control engineering papers and discussions to endeavour to attract more engineers into this field.

Nevertheless, in the next generation of computers faster speeds will be required and the computer designer will face many problems. Will size, weight, reliability and cost dictate the use of micro-circuits or will higher speed, but with equally high reliability, dictate the use of other techniques, for example, tunnel diodes?

The increasing demand for higher switching speeds will eventually bring us to the limit of semiconductor applications. We shall be approaching the limit of the mobility of holes and electrons. Probably not later than 1975, the use of integrated circuits or tunnel diodes will be replaced by light-operated circuits.

Light is our most fundamental parameter and it is interesting to speculate on the new fields of technology, techniques and terminology that will then arise. Using fibre optics we shall have to interconnect our various 'light circuits' with 'light' pipes. For easy servicing we shall need to interconnect with 'light' plugs and sockets. A foot of light equals a nanosecond. Shall we be specifying the length of our 'light' pipes in nanoseconds, or inches, or 'f.o.l.'s'?

Light will certainly play an important part in many of the techniques of the future. In very recent years, a form of solid-state electronics has emerged which exploits the quantum energies of atoms and molecules. From this has been developed the LASER and the MASER. The LASER, short for Light Amplification by Stimulated Emission of Radiation, provides us with a very intense, coherent beam of light of controlled divergence at a frequency of infra-red or visible light. Already it has been shown possible to concentrate laser light to a point source and achieve an energy level far exceeding that from any other source including the sun. This technique will eventually find many applications in industry, particularly in the machining of metals—and, in modified form, in surgery. The advantages of an optical fret-saw cum drill cum milling cutter, a small fraction of a millimetre in diameter, are obvious!

The laser has obvious potential in communications but much further research remains to be carried out before we can assess the magnitude of that potential. Certainly it is not the answer to *all* our communication problems, at least so far as communications on earth, or from earth, are concerned. It is true that we can modulate a laser beam, and that at these extraordinary high frequencies an enormous number of separate channels of communication may be accommodated. But because it is a beam of light it is sensitive to interference by mist, cloud or fog, and further the noise level for a given bandwidth is higher than is the case using maser techniques.

The development of the MASER, short for Microwave Amplification by Stimulated Emission of Radiation, preceded the laser. Because of its nature the maser principle makes very low noise figures possible and operating frequencies are not limited by electron transit-time effects. Already masers have led to striking progress in the communications field and satellite communication relies greatly on maser techniques. *Telstar* is an example. In addition, these have provided vastly improved sources of stable frequencies, which are leading to a continued change in outlook on communications, precision timing and on standards work.

The overall advances in electronics are having a striking effect on world communications. The much increased reliability arising from solid state devices, coupled with lower power demands, reduced weight and size has greatly accelerated long-range communications.

Within a decade trans-oceanic cables and communication satellites will increase in number and complexity, and this, coupled with automatic dialling, higher channel utilization and possibly speech compression, will lead to a relatively cheap and comprehensive world-wide communication network.

Within this time scale automatic dialling between all of the major world centres will be possible and a global television link-up to cover events of world interest can be achieved.

The ways in which electronics can be applied to industry are already extremely varied, but progress has been inhibited by the impression that electronics is essentially complex and unreliable. The subject is now taking on the mantle of engineering respectability. Complexity continues to grow, but so has reliability and, to the user, electronic devices are becoming simpler. I emphasize, *to the user!*

With this, the confidence of the industrial engineer has increased markedly in the past ten years and he is now prepared to grant electronic devices a major role in many branches of industry. This trend will continue and will probably accelerate to the point where every branch of industry will be using electronics to control its processes, maintain its balance of stocks, supervise its distribution and sales, and calculate its costs, salaries, returns and profits.

There is likely to be a growing demand for small special-purpose computers which can be synthesized through a limited range of standard 'bricks'. This, in turn, will call for a measure of standardization in such things as recording systems, inputs, logic, and print-out units, so that all manufacturers' products are compatible. Each section of the Industry will have its own specialized engineers, who will be trained to assemble the type of computer or control unit they require in much the same way that, at the present time, they now assemble production lines and services.

Conclusion

This then is the broad picture of the part that the Institution has played and has still to play: of the responsibilities that we have discharged and the responsibilities we still carry.

I have had the privilege of being concerned with 'R and D' projects in radio and electronics for the past thirty-six years. I only wish that I could look forward to working in this same field for the next thirty-six years for these will be years of great endeavour and of great opportunity.

In the fulfilment of our aims I hope that our judgment will keep pace with our scientific knowledge and that the members of this Institution will always bear in mind the words inscribed on our Grant of Arms—

'Science for the benefit of man.'

I thank you for the honour you have conferred upon me in electing me to be your President and I pledge myself to do all in my power to further the aims and objects of the Institution.

References

1. "Scientific and Technological Manpower in Great Britain, 1962." Cmd. 2146. October 1963. H.M.S.O.
2. "Committee of Enquiry into the Organization of Civil Science." Cmd. 2171. October 1963. H.M.S.O. (The 'Trend' Report).
3. "The Management and Control of Research and Development." 1961. H.M.S.O. (The 'Zuckerman' Report).
4. "Technology and the Sixth Form Boy." 1961. Oxford University, Department of Education.
5. "15–18." Report of the Central Advisory Council, for Education—England. Vol. 1. 1959. H.M.S.O. (The 'Crowther' Report).
6. "Annual Report of the Advisory Council on Scientific Policy, 1962–63." Cmd. 2163. H.M.S.O. (The 'Todd' Report).

(*Address No. 33*)

U.D.C. 621.395.625.3

A Novel Type of Magnetic Recording Head

By
J. C. BARTON, Ph.D.†
AND
C. T. STOCKEL, B.Sc.†

Presented at a meeting of the Electro-Acoustics Group in London on 9th October 1963.

Summary: Conventional magnetic heads are not suitable when a large number of narrow tracks have to be recorded simultaneously. Attention has therefore been directed to types of head in which the magnetic field is produced by a relatively large current in short sections of a small diameter wire, without using any magnetic circuit. In this way it has been possible to record 40 tracks simultaneously on ¼ in tape. As a variant of this technique the wire is replaced by an evaporated strip of silver, placed transversely and edge-on to the tape. This provides a head which has intrinsically good 'wear' properties and also enables a pair of heads to record on interleaved tracks. A number of heads of this type have been made which provide 60 tracks on ¼ in tape.

The design of these heads is discussed and it is shown that they are suitable for data recording and that they have a far smaller time constant than conventional heads. Their use for normal analogue recording is not yet feasible as the power dissipation is too high. However it is suggested that re-introducing a micro-magnetic circuit, by an evaporation technique, would overcome this problem while retaining the other advantages.

1. Introduction

The magnetic heads which are used in tape recorders today are basically the same as those used when magnetic tape was first introduced. The greatly improved quality which is now achieved is the result chiefly of more precise construction and to a lesser degree of the use of improved materials. The very success of the standard technique seems to have discouraged any attempt to reconsider the problem of head design. The limitations of the established design become clear when an attempt is made to use very narrow track recording. For data work, and possibly in other applications too, there is an obvious interest in such attempts, but it is extremely difficult to make conventional heads for more than six or eight tracks on ¼ in (6·35 mm) tape. Interleaving of tracks is possible but this introduces additional problems.

On basic physical grounds there seems no reason why it should not be possible to use tracks at least an order of magnitude narrower than those commonly used. In fact if one considers the shape of the magnetized region on a tape, when recording short pulses, the conventional system leads to the worst possible self-demagnetization effect. One might therefore expect that the actual density of information storage, in the sense in which the terms are used in communication theory, will be greater if narrower tracks are used, even though their individual signal to noise ratio will be worse.

† Northern Polytechnic, Department of Physics, London, N.7.

Our own interest in narrow track recording arose from a very specialized requirement in cosmic ray research. We have needed to get the maximum possible numbers of tracks on standard ¼ in tape but have not had to worry much about the playback problem, as it has been sufficient to make the recordings visible by using a simple magnetic powder technique. In doing this we have found it necessary to use a radically different approach to head design.

A brief report on this work, including a survey of possible playback techniques, has already been given.[1] The present paper will describe the design and construction of the new types of heads and report on their performance for data recording. Possible ways in which they could be used for more general applications will then be discussed but no experimental work has been done in this direction.

2. Design and Construction of Heads

The general principle of these recording heads is to dispense altogether with the conventional magnetic circuit and to use the magnetic field around a current carrying conductor. The case generally given for using a magnetic circuit is that it is thought to lead to greater efficiency, but a little consideration shows that this is not necessarily true. This can be seen by reference to Fig. 1(*a*) which shows, approximately to scale, a section through a normal head. The most striking feature is that a relatively vast magnetic circuit is used to produce a field which is only useful

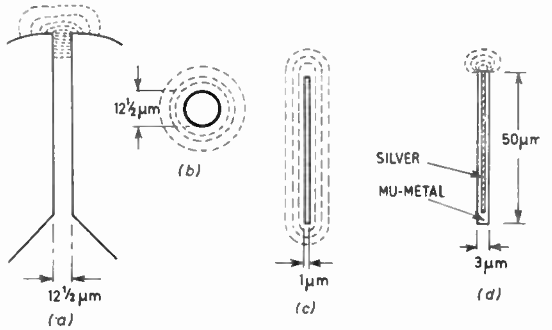


Fig. 1. Field distributions at recording heads

(*a*) Conventional (*c*) Evaporated strip
(*b*) Wire (*d*) Proposed improvement

in a very small region; moreover it is only the fringing field which is used instead of the stronger field in the gap. Figure 1(*b*) shows how a recording field of comparable extension can be produced by a current carrying wire. It is immediately clear that the second arrangement is likely to be more efficient in the sense that a greater proportion of the magnetic field energy is in the useful region.

The great simplicity of this arrangement makes it easy to use tracks of any desired width and, what is more important, to have a large number side by side. This has been realized using the construction shown in Fig. 2. A stack of alternate pieces of mica and silver are clamped together, their ends ground to a suitable contour and the wire soldered across. Then any section of the wire can be used by passing a current down one silver strip and up the next, each pair of strips being connected to their own transformer. The thicknesses of the mica silver strips for any one head, were all the same, from 75μm to 125μm according to the number of tracks required.

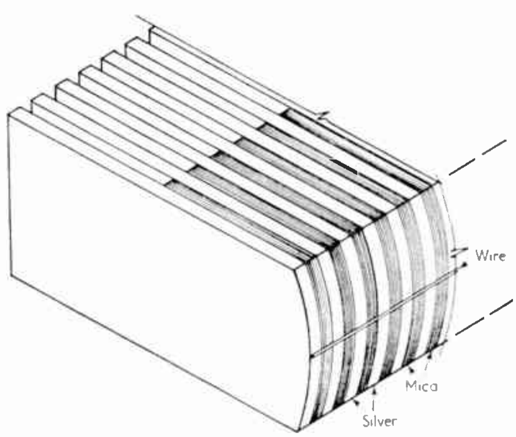


Fig. 2. Multi-track wire head.

For heads recording 24 or 40 tracks on $\frac{1}{4}$ in tape this method has proved satisfactory, using a copper plated tungsten wire of 15μm diameter. For saturation recording $H \simeq 2i/r \simeq 300$ oersteds, so that the current required is about 1·5 amps. The transformers used for pulses of 5 μs duration were 12 mm ferrite toroids with 40-turn primary windings and single turn secondaries. A 24-track head of this type has been successfully used over a period of some months in an experiment to measure the intensity of the cosmic radiation deep underground.[2]

The main problems encountered were to keep the wire straight during soldering and to protect the wire

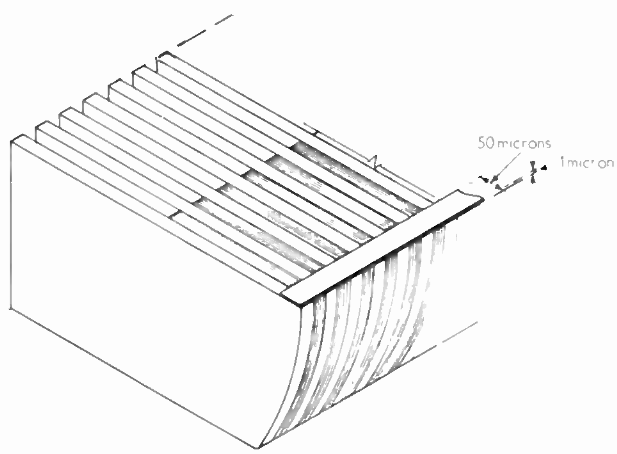


Fig. 3. Multi-track evaporated strip head.

from abrasion by the tape. The latter was solved by fairing in the wire to the surface of the head with a cold-setting epoxy resin. However it seems improbable that the 'wear' properties of these heads will ever be very good. A more radical solution is to replace the wire by a fine strip, set edge-on to the tape as indicated in Fig. 1(*c*), and requiring the construction for a multi-track head shown in Fig. 3. The latter is made by grinding flat the sides of the stack of mica and silver strips and then depositing the strip, usually of silver, by evaporation. The accuracy of track alignment is then limited only by the accuracy with which the surface is made flat; in our heads the flatness was established optically to be within 1μm. The electrical characteristics of this type of head should only change slowly as the head wears.

Calculating the recording current for this type of head involves regarding the strip as composed of parallel filaments and integrating their effects. Under similar operating conditions the current required is found to be a few times larger than for the wire heads. Since the evaporated strip is so thin the resolution along the tape is exceptionally good.

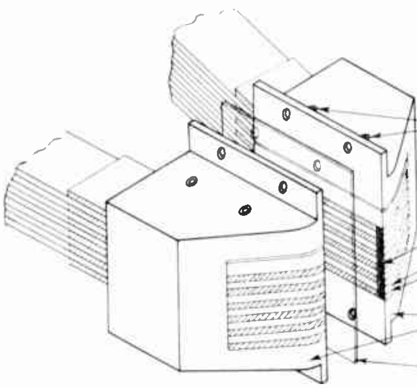


Fig. 4. Exploded view of staggered track head. (For clarity the diagram shows only a small number of tracks.)

An attractive feature of the evaporated strip heads is that they are well adapted to the use of staggered tracks. Two half heads can be placed adjacent with only a sheet of mica between them, as shown in Fig. 4, thus avoiding the large and possibly inconstant time difference which is normally encountered when using staggered tracks.

Several 60-track heads of this latter type have been made. The evaporation has so far been carried out under rather poorly controlled conditions so that there is an appreciable spread in the characteristics of individual sections of the heads. Figure 5 shows the histograms of the resistances of the head elements for two halves of one head. It will be seen that one side of the head shows a much larger spread than the other; the very high value for one section may be due to mechanical damage during final assembly. With greater care in the alignment of the mask and with improved evaporating technique we believe that results as good as for the better half this head could be achieved consistently. An example of the quality of the records

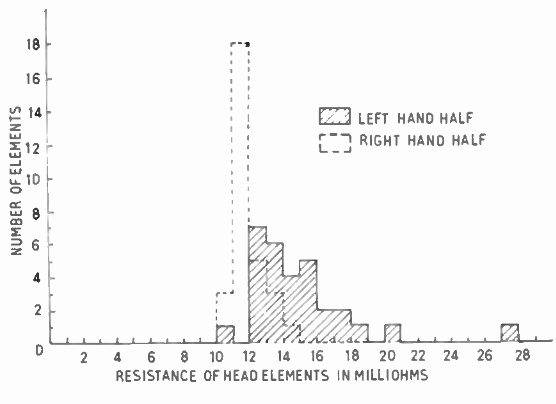


Fig. 5. Histograms of resistances of individual head elements for staggered-track head.

at present obtained is shown in Fig. 6. Examination of the powder pattern itself indicates that the resolution along the tape is certainly better than 25μm. It is also apparent that the magnetized region is somewhat wider than the mica thickness; this suggests that it might be advantageous to use thinner mica and thicker silver strips.

The inductance of the head elements is extremely small, of the order of 1 nH, and we have not succeeded in measuring it. In practice the lead inductances, which are about 10 nH, and the leakage inductance of the driving transformer will be much more significant in determining the high frequency performance. Using the 12 mm toroidal cores mentioned previously the time constant is about a microsecond but with improved transformer design it should be possible to reduce this figure considerably.

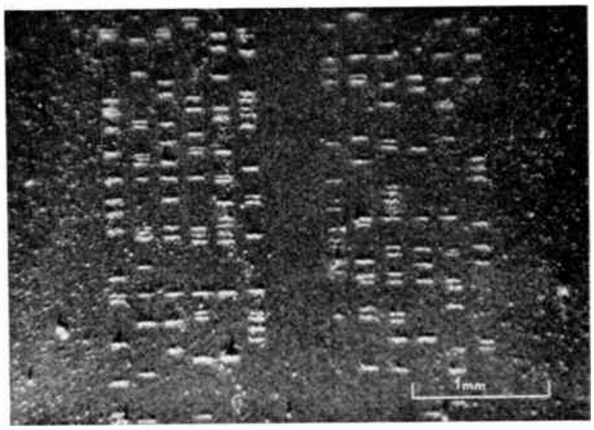


Fig. 6. Magnetic powder pattern showing recorded pulses from elements of one half of a 60-track head. Only 12 of the tracks were used for this picture. Oblique illumination was used to show the pattern more clearly. The various large dots are due to dust particles which have settled on the tape after applying the magnetic powder.

The main disadvantage of these heads is that the power dissipation is relatively high. For both the soldered wire and the evaporated strip types the resistance of each head element is usually something like 20 mΩ. This means that for a sinusoidal signal the power dissipation is approximately 30 mW per element for the wire head and over 150 mW for the strip head. The former figure is just about tolerable if only a single element is energized but precludes the use of several tracks simultaneously. The strip heads are considerably worse but would be comparable in performance if a thicker strip could be used. Unfortunately it is not easy to obtain satisfactory evaporated layers much more than 1μm thick.

The comparative performance of different types of heads are shown in the Table.

Table 1

Characteristics of individual head elements for different types of recording head.

Type of head	Track width	Resistance	Inductance	Time constant	Recording current	Stored magnetic energy	Recording power level
Commercial 4-track head—typical values	0·030 in	11Ω	5 mH	450 μs	0·01 A	$2{\cdot}5 \times 10^{-7}$J	1 mW
Special 8-track head—conventional design[3]	0·015 in	3Ω	0·6 mH	200 μs	0·02 A	$1{\cdot}2 \times 10^{-7}$J	1 mW
40-track wire head	0·003 in	0·025Ω	0·01 μH†	0·4 μs	1·5 A	$2{\cdot}5 \times 10^{-8}$J	60 mW
60-track staggered evaporated strip head	0·004 in	0·02Ω	0·01 μH†	0·5 μs	4 A	$1{\cdot}6 \times 10^{-7}$J	300 mW

† Inductance mainly due to leads.

3. Use for General Purpose Work

As stated in the introduction we have not made any detailed study of the performance of these heads under normal play-back conditions, though preliminary experiments showed that the signal level from a 40-track wire head at 15 in/s was about 0·4 μV. Since this signal appears across a resistance of only 20 mΩ the signal/noise ratio is still acceptable; the r.m.s.-Johnson noise would be about 40 dB down for 50 kc/s bandwidth. Again the strip geometry will be less effective than the wire one, but the actual response can be calculated rather accurately. No attempt has been made to measure cross-talk. The electrical coupling between channels is certainly small but tolerances on lateral movement of the tape will probably be more significant.

Daniel and Axon[4] proposed the use of this geometrical arrangement for the absolute measurement of the recorded signal level on magnetic tape. The use of an evaporated multi-track head would appear advantageous in providing greater precision than the single strip of foil used by these authors and also in enabling the lateral extension of the recorded region to be monitored. The possibility of using this technique for investigating the problems of track alignment and of tape skewing, etc. may be worth further attention.

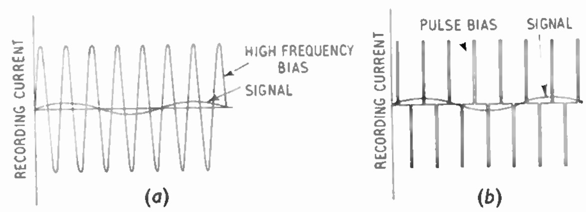


Fig. 7. Comparison of conventional high frequency bias and suggested bi-polar pulse bias.

Because of the excessive power dissipation the heads are not directly suitable for normal recording using high frequency bias. A possible way of overcoming this limitation is to replace the sinusoidal bias current (Fig. 7(*a*)), by a bi-polar pulse one, (Fig. 7(*b*)). This ought to have the same linearizing effect as the usual bias but with a greatly reduced dissipation. The more complicated circuit needed to generate this waveform could be justified if recording was required simultaneously on a large number of tracks. Due to the very low time constant of the heads it should be possible to reduce the duty cycle to less than one tenth.

An alternative solution is to reduce the power dissipation at the head by re-introducing a magnetic circuit on a micro scale. A possible arrangement is shown in Fig. 1(*d*) where it is envisaged that layers of mu-metal are evaporated on either side of the silver layer. We have not ourselves any experience of evaporating mu-metal but reports in the literature suggest that no insuperable difficulties need be anticipated.[5]

4. Conclusions

The heads described in this paper are suitable for recording pulses but more development is necessary before they can be applied to general purpose recording. The great simplicity and cheapness of the heads suggests that their application even to wide track recording is worth further consideration.

The use of extremely narrow tracks has been shown to be feasible. The limitations to the use of narrower tracks in data processing systems are set by the engineering difficulties of tape transport and by the relatively large tolerance on the tape width. It is to be hoped that the tape manufacturers will soon be able to provide tape with tolerances less than the customary 100μm, so that a study of the remaining difficulties can be more readily undertaken.

5. References

1. J. C. Barton, "Multi-track recording on magnetic tape", "Proceedings of the Symposium on Nuclear Instruments, Rutherford Jubilee Conference", Ed. J. B. Birks, p. 182, (Heywood, London, 1962).
2. J. C. Barton, "Intensity of the cosmic radiation at great depths underground", *Phil. Mag.*, **6**, p. 1271, October 1961.
3. J. C. Barton, "Eight-channel recording on ½ in magnetic tape with a stationary-tape playback", *J. Sci. Instrum.*, **33**, p. 415, November 1956.
4. E. D. Daniel and P. E. Axon, "The reproduction of signals recorded on magnetic tape", *Proc. Instn Elect. Engrs*, **100**, Part III, p. 157, May 1953.
5. F. B. Humphrey, *et al.*, "Introduction to magnetic thin films", "Fifth National Symposium on Vacuum Technology Transactions", p. 204. (Pergamon Press, New York, 1958.)

Manuscript received by the Institution on 30th August 1963. (Paper No. 873/EA11).

DISCUSSION

Under the chairmanship of Mr. C. T. Chapman

Mr. A. V. Davies: As a designer of conventional heads I have followed this work with great interest, and would like to congratulate Dr. Barton and his colleague on producing a new design which adequately meets his requirements in a far simpler way than could be achieved by the usual techniques.

I would like, however, to discuss in more detail the comparisons made in the efficiency and inductance of the new head and conventional designs, and the possible use in other applications, data processing in particular. The cosmic ray application has basically the sole requirement of recording with a high track density; the head is not required to replay, erase or overwrite previous history, or simultaneously record pulses at high packing density on all tracks, while in most pulse recording applications these requirements are all-important.

1. Let us first consider the comparisons between the efficiency of conventional heads and the new design. In this and an earlier paper† it has been stated that the volume of magnetic material is enormous compared with the usefully employed region, and assumed that the stored energy in the cores is therefore large compared with the energy in the useful region, and the head has unnecessarily high inductance which imposes a limit on the highest frequency which can be used. No calculations are made and the diagram of field distributions in Fig. 1 and comparisons of Table 1 can be misleading without deeper examination. In a conventional head, the inductance is approximately proportional to the square of the number of turns, n, and to the track width, w.

$$L = K_1 n^2 w \qquad \ldots(1)$$

The recording field is proportional to the ampere-turns but independent of track width, hence for a particular recording field, H, the current i is inversely proportional to the number of turns,

$$i = \frac{K_2 H}{n} \qquad \ldots(2)$$

In the replay process, the replayed voltage V, is proportional to the track width, the number of turns, and the tape speed, v,

$$V = K_3 nwv \qquad \ldots(3)$$

The inductance of a head can be 'tailored' to the user's requirement by adjustment of the number of turns, as in transformer design, a lower inductance requiring a higher record current and giving a lower output. For example, in Table 1, the first head, having 0·030 in track width and 5 mH inductance, probably has approximately 250 turns. A reduction to 25 turns would reduce the inductance by a factor of 100 to 50 μH, and the record current for the same field would be increased from 10 to 100 mA, which is not excessive. The output from the 5 mH head at 15 in/s would be something like 7 mV peak to peak from a saturated tape, and with the lower inductance would be 700 µV; allowing for the track width being an order of magnitude higher than with the new type of head, the replay output per unit track width is still much higher with a conventional design of low inductance. I have successfully used conventional heads of 20 to 30 µH inductance for high frequency applications; the limitation in frequency response of a conventional head is not the inductance but the lag in flux rise time due to eddy-current and, to a lesser extent, hysteresis effects in the cores. Here, an efficient magnetic circuit design, minimum work hardening in assembly, and the use of thin laminations minimize these losses.

A considerable amount of work has been done in improved core materials for high frequency and fast rise time applications, for example, the use of alfenol laminations,‡ which have 2½ times the resistivity of mumetal, and the development of dense ferrite materials suitable for fine gapped applications. § The comparisons in inductance and stored energy of Table 1 are not indications of efficiency; the 'special' conventional head of Table 1 has half the track width of the 'commercial' head and requires double the record current, indicating a halving of the number of turns. Thus, the inductance can be expected to reduce by a factor of 8, which is the observed reduction. The stored energy is ½ Li^2, which from eqns. (1) and (2) becomes

$$E = \tfrac{1}{2} Li^2 = \tfrac{1}{2} K_1 K_2^2 H^2 w \qquad \ldots(4)$$

It is seen that the energy for a particular recording field is independent of the number of turns and is proportional

† See ref. 1.

‡ K. M. Rose and G. C. Glen, "Recent developments in soft magnetic alloys", *Electrical Manufacturer*, June 1958.

§ S. Duinker, "Durable high-resolution ferrite transducer heads employing bonding glass spacers", *Philips Res. Rep.*, **15**, p. 342, August 1960.

A. L. Stuyts, J. Verweel and H. P. Peloscheck, "Dense ferrites and their applications', I.E.E.E. International Conference on Non-Linear Magnetics, Washington, 1963, Session 1.

to track width. The stored energy for the 'special' head is seen to be half that for the 'commercial' head, so the efficiency of the two heads is probably much the same. The efficiency will be governed by the constants K_1 and K_2, which I intend treating in more detail in a published paper in the near future.

For the stored energy figures to be comparable as an indication of magnetic circuit efficiency, eqn. (4) shows that the energy should be expressed in terms of energy per unit track width for the same recording field. Making this adjustment, and assuming the commercial four-track head to have a track width an order of magnitude larger than the new designs, the conventional designs appear some fifty times more efficient than the wire head and three hundred times more efficient than the strip head (there appears to be a slight error in the stored energy calculations for the new design—if defined as $\frac{1}{2}Li^2$ they should be divided by two).

In a well-designed conventional head the massiveness of the magnetic circuit reduces rather than increases the stored energy in the cores. The volume of magnetic material is used in making the lamination sufficiently wide for the core reluctance to be small compared with the gap reluctance, the path length being as small as winding space will allow. This being the case, and with a lamination geometry which minimizes flux leakage, the majority of the energy is stored within the gap. Admittedly, the stray field records the tape, and the recording field is only equal to the gap field at the contact surface. Gap reluctance, however, is made as high as possible by reducing the pole tip depth, and hence energy wastage within the gap, to as small a value as reasonable head life will allow.

2. The recording current comparisons are also misleading. The figures for the new designs produce a field of 300 oersteds at the surface of the wire or strip, which is approximately the coercivity of a normal oxide coating. In Fig. 1(*a*) and (*b*) the gap length of the conventional head and the wire radius are both 12·5 microns, the same order as the thickness of the oxide coating on a tape. It can be shown that for a well-designed head, the ampere-turns necessary to saturate fully the coating thickness are approximately proportional to the product of the coating thickness and oxide coercivity, and typical figures are 0·5 ampere-turn per 0·0001 in of coating. The recording currents quoted for the two conventional heads correspond to the region of 2·5 ampere-turns, which will more than saturate a 12·5 micron coating thickness. The field at a gap length separation will be approximately one-third of the field at the surface, and in the case of the wire, the field at a wire diameter separation will have also reduced by this amount. The saturation field for a tape is approximately treble the coercivity, hence the field at a gap length separation will be about 1000 oersted, and at the contact surface of the tape will be some 3000 oersted, an order of magnitude higher than the field produced by the new design for the quoted current.

The paper has stated that excessive power dissipation prevents the use of the new head for a high frequency sinusoidally biased application. This limitation also applies to the majority of pulse recording applications, since it is usually necessary to saturate fully the coating at some stage, either in erasing or overwriting the previous history of individual tracks, there being comparatively few recording systems where full tape width or bulk erasure may be used. There is no difficulty in pulse recording below saturation, as Dr. Barton's powder patterns have shown; in fact it can be advantageous only to penetrate partially the coating, leading to better resolution, and there is a trend towards working at a high pulse packing density with a 0·0002 in coating thickness rather than the usual 0·0004 in, giving something like a factor of two improvement in pulse resolution. For good pulse resolution, however, it is still necessary for the recording field at the contact surface of the tape to approach saturation, which is approximately treble the quoted field for the new design.

3. The time-constant figures of Table 1 are misleading, since they are referred to the head resistance, which is dependent on the gauge and number of turns of wire used. The time-constant in the recording process which governs the rise-time of the current is the ratio of head inductance to source resistance, and this can be reduced to negligible proportions at the expense of driving power. If the source resistance is R, the time-constant T will be L/R, the head resistance usually being negligible. Thus, from eqns. (1) and (2), it can be shown that the driving power P is

$$P = i^2 R = \frac{K_2^2 H^2}{n^2} \times \frac{K_1 n^2 w}{T}$$

$$= \frac{K_2^2 K_1 w H^2}{T} \qquad \ldots(5)$$

Thus, for a given recording field, the driving power to give a time constant T is independent of the number of turns but inversely proportional to the time constant and proportional to track width. For example, if a driving power of 1 watt is applied to the four-track head of Table 1, the source resistance would be 10 000 ohms and the time constant will be 0·5 μs.

The power required is admittedly very large compared with the power dissipated in the head, but is dissipated within the driving circuit where it can do no damage. Bearing in mind that this is producing a recording field which is an order of magnitude larger to a track an order of magnitude wider than the new heads, in terms of driving power per unit track width for the same recording field and time constant, the conventional head again appears to be, from eqn. (5), fifty times more efficient than the wire head and three hundred times more efficient than the strip head.

4. Let us now consider the replay performance. It is possible in theory to step-up a 0·4 μV signal from a 20-milliohm resistance so that a reasonable value of signal/noise ratio is maintained before amplification, but in practice the resistance and inductance of leads to transformers in a high track density head is likely to be a serious limitation, since each track will require an individual transformer. A conventional digital head is normally wound with sufficient turns to render replay transformers unnecessary. Signal/noise ratio will become more critical at high track densities since, while replay signal is proportional to track width, tape noise is proportional to its

square root owing to random phase distribution of the noise.†

5. The suggestion of saving power in a biased system by pulsed bias is most interesting; there have been many papers written on the subject of a.c. bias, and it is doubtful whether any theory fully explains all aspects of the process. The most commonly held explanation, which was developed by Westmijze‡ and has recently been extended by Eldridge and Daniel,§ describes the linear recording process in terms of the anhysteretic magnetization curve. Doust|| and later, Sebestyen and Takacs¶ observed that it is possible to obtain a distortionless recording when any one element of tape is in the recording region for less than the periodic time of the bias. The Sebestyen and Takacs process would definitely require a sinusoidal waveform, since the mark/space ratio of a rectangular waveform would be unaffected by the superposition of the audio field.

6. Finally, after challenging the efficiency of the magnetic circuit in conventional heads, the paper proposes its re-introduction by vacuum deposition of mumetal round the strip. This would undoubtedly effect an improvement, and the head would then become a single-turn version of the conventional design. It is probable, however, that a considerable thickness would be necessary to approach the desired magnetic circuit efficiency and the limitations imposed by lead inductance and resistance would still apply.

High track density recording, when tape guidance problems are sufficiently solved, is a problem which will face the head designer in the none-too-distant future, and experimental conventional heads have been made for studies at 200 to 500 tracks per inch.†† The limitations of Dr. Barton's design would appear too severe for its general acceptance, and the conventional head becomes very difficult to make in its present form.

Mr. J. F. Doust: My paper|| to which Mr. Davies referred was an examination of the performance of magnetic tape in terms of the properties of the individual particles rather than of those of the bulk material. On this basis the efficacy of a pulse bias would need some consideration but it seems likely that it would be successful.

† D. F. Eldridge and A. Baaba, "The effects of track width in magnetic recording", *Trans. Inst. Radio Engrs, (Audio)*, **AU-9**, p. 10, January–February 1961.

‡ W. K. Westmijze, "Studies in magnetic recording", *Philips Res. Rep.*, **8**, pp. 148 and 161, April and June 1953.

§ D. F. Eldridge and E. D. Daniel, "New approaches to ac-biased magnetic recording", *Trans. Inst. Radio Engrs (Audio)*, **AU-10**, May-June, 1962.

|| J. F. Doust, "The behaviour of magnetic particles during the recording process", *Sound Recording and Reproduction*, **5**, p. 61, 1956.

¶ L. G. Sebestyen and J. Takacs, "Magnetic recording—theory of tape magnetization", *Electronic Technology*, **38**, p. 274, 1961.

†† L. F. Shew, "High density magnetic head design for non-contact recording", *Trans. Inst. Radio Engrs (Electronic Computers)*, **EC-11**, p. 764, December 1962.
Eldridge and Baaba, *loc. cit.*

I am very interested in the use of the evaporated strip conductor which Dr. Barton has described and I should like to ask him whether there exists an analysis of the working field at the edge of the strip in terms of its thickness, its width, and the frequency content of the recording current. I should also like to know whether a mumetal layer on the strip is likely to give any great advantage in view of the large leakage fields likely to occur as a result of the thinness of the conductor.

Mr. R. G. T. Warren: With regard to the difficulties mentioned in applying a h.f. bias to the type of head under consideration, I should like to suggest that it is possible to record analogue information without the use of conventional electrical h.f. bias by application of ultrasonic vibration to the head assembly. My own observations show that this is possible but experiments carried out so far have been too elementary to indicate whether this form of physical bias application is comparable with the normal application of electrical energy.

In connection with the previous mention of saturated tape I should say that this value of magnetization is of no real consequence since it is the *remanent magnetization* which is of interest to us in that it is this value which will eventually produce the playback e.m.f. I should also like to suggest that the resolution of a recording medium is a function of a perpendicular magnetization of the medium due to the curvature and non-uniformity of the recording field. Information which is recorded longitudinally with such a recording field will always impart a certain perpendicular magnetic component the stray fields of which will depart from the longitudinal and give rise to induced e.m.f. in adjacent tracks. This effect is also influenced by the domain structure of the recording medium since it is obvious that domains can have a preferential axis of magnetization set in a longitudinal mode which can mitigate the effect of the undesired perpendicular component.

Mr. H. H. Findeisen (*Communicated*): One application of Dr. Barton's head with which I have been experimenting was a single conductor head made by evaporating a U-shaped conductor on a simple glass plate. After evaporating, this conductor was plated up to approximately 0·0005 in and then a second glass plate placed over it for protection. The horizontal connecting bar of the U was then ground back to leave a conductor width of about 0·002 in.

This conductor was used to write a single digit across a 1-in wide tape and the digit thus written was used to check gap scatter and skew of a 16-track reading head. The application is interesting because one can calibrate skew and gap scatter of recording and reading heads independently.

Experiments which I made but never fully concluded and which were concerned with the reading with a narrow track single wire head such as used by Dr. Barton for recording, have so far proved unsuccessful and I feel that because of the very low outputs obtainable, there is little hope of ever using such a head for reading.

REPLY BY THE AUTHORS

The question asked by Mr. Doust is partially answered in a paper by D. J. Sansom,† in which the field due to a long thin strip is determined. Calculations for the short lengths of strip used in these heads would be very difficult but the theory mentioned should provide an adequate guide. We agree that the extent of the improvement which would result from mu-metal layers is uncertain, but suggest that the leakage is determined by the ratio of the length of magnetic flux path to the thickness of the layer and that this could still be small enough to result in a substantial improvement.

We are very interested to hear that Mr. Findeisen has already started experiments with evaporated strip heads for calibration purposes. We agree that the problems of play-back are severe but believe it is the mechanical difficulties which really limit the performance. However it has been shown by Eldridge and Baaba‡ that satisfactory play-back, using a conventional type of head, is possible with a track width of only 50μm. The additional problems introduced by the proposed types of head are only those associated with the matching transformer and should not be insuparable.

It would be interesting to have further details of the vibration technique used by Mr. Warren. However this could not help to reduce the power dissipated in the head sections during recording, so would not solve the same problem for which we propose using pulse bias. We agree with Mr. Warren's comments on the importance of the remanent magnetization and on the three-dimensional character of the recorded pattern. For very high resolution recording there would be considerable advantages in using tape with a thinner active layer.

The detailed comments and criticism of Mr. Davies are extremely valuable and we look forward to his future paper in which the topics are to be dealt with in detail. Meanwhile we make the following replies to some of his criticisms; we have already referred to one point, concerning the use of a mumetal layer, in our reply to Mr. Doust above.

1. We agree that we were in error in referring to 300 oersteds as a saturating field and that our tapes are not fully saturated throughout the thickness of the active layer. However, we do not think that a recording field needs to be as large as three times the minimum field, although this may be necessary for demagnetizing. Indeed we have observed that the sharpness of the powder patterns, which is an indication of the resolution, is better when a smaller recording field is used, as would be expected. We also agree that thinner tape is necessary for the highest resolution.

The figures given in our table were intended to be comparative ones and refer to the recording current that we used rather than that necessary for demagnetization. It is possible that some commercial heads are more efficient than others so that for the four-track head the current given may be rather greater than the minimum recording current.

2. We accept the general validity of his analysis regarding the relative magnitudes of the stored energy in each type of head but are doubtful whether eqns. (1) and (4) will still be accurate for very small values of w. The assumption implicit in these equations is that there is no leakage flux. Even with normal track widths it is known that the amount of flux leakage is already considerable, as is shown by the amount of cross-talk which occurs unless magnetic screens are interposed between adjacent head sections. With thinner laminations it becomes increasingly difficult to avoid a deterioration of the magnetic properties during assembly, so that the leakage problem will become very severe for narrow track heads.

The difference between the ratios of stored energies, as estimated by Mr. Davies and ourselves, is thus attributable partly to the differing track widths, partly to uncertainties about how deeply the magnetization penetrates and partly to differing estimates of the field necessary for recording. Further comparative measurements are necessary to clarify these points and decide which estimates are more reliable. In any case the relative advantage of the new types of head will become greater as the track width is reduced.

3. Regarding the time constants of the heads, it can easily be shown that for any magnetic circuit device the time constant is independent of the gauge of winding wire, provided only that the available winding space is fully used. Mr. Davies's first assertion is therefore incorrect. His further point that the time constant can always be improved by driving from a high resistance source is of course true but the accompanying waste of power could become serious in a system using a large number of channels.

4. It is suggested that lead impedance will be a serious limitation on playback performance. In the system we use for recording the lead resistance is only of the same order as the head element resistance so that this difficulty is not insuperable. It is true that the conventional type of head itself also performs the function of acting as a transformer, but we believe that by separating the functions of head and transformer each can be designed for optimum performance. This is also advantageous economically as the cost of the head is reduced and is the only part which will ever need replacement.

5. His suggestions that the use of pulse biasing might throw some light on the validity of the various theories of biasing is a very interesting one which ought to be investigated.

6. Mr. Davies agrees that higher track density systems must certainly be envisaged. In view of the admitted mechanical difficulties of conventional heads for this purpose we believe that the type of head discussed in our paper is more likely to be capable of development to overcome its remaining disadvantages.

† "The magnetic field arising from a steady current flowing in a long straight conductor of rectangular cross-section", *Contemporary Physics*, 4, p. 459, 1963.

‡ *Loc. cit.*

Reports of the Special and Annual General Meetings

A Special General Meeting of the Institution was held in London on 27th November 1963, attended by 110 Corporate Members, for the purpose of approving the Resolution for changing the name of the Institution to "The Institution of Electronic and Radio Engineers". The chair was taken by Admiral of the Fleet the Earl Mountbatten of Burma, K.G. (Retiring President) who formally moved the first Resolution of the Council which effected the necessary amendments to the Charter. The two Resolutions were as shown in the Notice of the Meeting printed in the November 1963 issue of *The Radio and Electronic Engineer*.

Following a brief discussion to clarify certain points, the Resolution was put to the meeting and approved by 706 votes to 5.

The Chairman then formally moved the second Resolution of the Council which made amendments to Bye-laws 1 and 9 dealing primarily with designations of membership.

This Resolution was passed unanimously.

Lord Mountbatten then commented: 'We have, by a very large majority, now agreed to seek the sanction of Her Majesty in Council to changing our name. I believe this opens a new chapter for the Institution which started with wireless technology, then changed to radio engineering, and is now concerned with electronic engineering.'

The Institution's second Annual General Meeting since incorporation by Royal Charter followed the Special General Meeting. Moving the adoption of the Annual Report of the Council for the year ended 31st March 1963, Lord Mountbatten said he was particularly pleased to do so because it gave him a chance to express his personal thanks to all the members who had served the Institution during his Presidency. No less than 344 Corporate Members had served the Institution on committees and by representation on outside bodies.

He continued: 'For the first time all members have received the Annual Report through our newest publication—the *Proceedings*. This, I suggest, is a further example of the way in which the Council is consistently considering increases in the services which the Institution can give to its members. Whilst the *Proceedings* were originally planned for the benefit of members in Great Britain, the idea has been taken up most enthusiastically in our Commonwealth Divisions. Although it is an event which occurred after the close of our year, I must refer to the enthusiasm of Indian members in producing similar *Proceedings* to reflect Institution activities in that country.

'During my visit to India in April I was able to meet and to explain to members of the Institution in Delhi the Council's plan for granting a greater measure of autonomy to our Commonwealth Divisions. In consequence, we now have an administrative arrangement in India which satisfies the natural aspirations of Indian members to further the Institution's work in their own country without the impediment of tight control from London. This, I can assure you, has aroused noticeable goodwill towards the Institution as a whole and will, I am sure, provide a pattern for the ultimate development of not only our own Institution but, I hope, of others in contributing to goodwill throughout the Commonwealth.

'I would particularly like to draw your attention to the contribution that the Institution has made towards securing unification in the engineering profession. All the Engineering Institutions have worked together to secure this unification under the Engineering Institutions Joint Council. If we are to take full advantage of scientific achievement we must increase our engineering ability at a far quicker pace. I believe that all our engineering institutions must expand in membership because we need more engineers.

'Before leaving this important need for the engineering institutions to work more closely together I welcome this opportunity of my last appearance as President of the Institution to express genuine pleasure as well as very great satisfaction at the good relations which we now enjoy with the Institution of Electrical Engineers. There are so many ways in which the two Institutions can collaborate'.

The Chairman also paid tribute to the work of the various Standing Committees.

The Report was approved unanimously.

The Chairman then called upon Mr. A. A. Dyson, a Vice-President, and Chairman of the Finance Committee, to move the adoption of the Accounts for 1962–63.

Mr. Dyson said he was in the happy position of not having to defend an excess of expenditure. He continued: 'However, the satisfactory position of our Income and Expenditure Account does not allow for complacency, because the margin of excess of income for the twelve months is inadequate for the forward financing of the Institution's development and capital expenses. In brief, the task of expanding the Institution's services to members and our general activities is still too dependent upon current income'.

Much had been done in improving the Institution's headquarters to meet the problem of inadequate accommodation and consequent limitation of staff.

The acquisition of No. 8 in addition to No. 9 Bedford Square had removed these problems, and had enabled the Council to undertake further services.

'We are not, however, devoting all our resources to the provision of amenities and services solely for members in Great Britain', said Mr. Dyson, 'In India and Canada—as a start—Divisions have been established with their own headquarters and staff and responsibility for publishing national *Proceedings* on similar lines to the *Proceedings* published for members in Great Britain. The Council intends to develop this Divisional structure of the Institution in other countries of the Commonwealth. All this will obviously involve considerable expenditure for initial capital outlay.

'The Finance Committee has, of course, the greatest difficulty in working to a budget because of the constantly increasing costs with which we all have to contend personally, but which become major matters in running the financial affairs of the Institution. Nevertheless, the Finance Committee is continuing to control expenditure to a budget which will enable the Institution to build up reserves, the need for which I have already mentioned'.

The accounts were approved unanimously.

The next item concerned the election of the Officers and the Council for 1963–64. No nominations other than those of the Council had been received and Lord Mountbatten therefore declared the following elected:

President: John Langham Thompson. *Vice-Presidents:* L. H. Bedford, C.B.E., M.A., B.Sc., F.C.G.I.; A. A. Dyson, O.B.E.; I. Maddock, O.B.E., B.Sc., W. E. Miller, M.A.; Colonel G. W. Raby, C.B.E.; Professor E. E. Zepler, Ph.D. *Ordinary Members of Council:* G. Wooldridge, Ph.D., M.Sc. (Member) and Rear Admiral John Grant, C.B., D.S.O. (Companion). *Honorary Treasurer:* G. A. Taylor.

Together with the remaining members of the Council who will continue to serve in accordance with the periods laid down in Bye-law 42, these members will constitute the 1963–64 Council of the Institution.

After the presentation of Premiums and Awards, and Examination Prizes, a vote of thanks to the retiring President was moved by Professor Emrys Williams in the following terms:

'I speak on behalf not only of the members present, but also of members all over the world who are unable to be with us this evening. I rise under the item "Any other business" not, in fact, to discuss business, but to comment on a very strong feeling existent among our members throughout the world. I refer to the wish of those members that we incorporate in the Minutes of today's business deep appreciation and gratitude of members of the Institution to our retiring President. We wish to see in our records reference to the man who nearly 30 years ago sought membership of a very small Institution because he believed that it deserved and needed encouragement.

'What he has done in our Institution during those intervening years shows in almost every facet of our activities. In education, he has not lost sight of the need for the Institution to give its help to the technician and craftsman. In another and wider field, he has ensured the Institution's participation in the important matters of co-ordinating the engineering profession as a whole. He has made several contributions to our proceedings and I need only cite his papers on the future of computer technology and on satellite communication which were written at a time when these subjects were not so popular as they are today. They are examples of his forward looking scientific outlook.

'As our Charter President and as the only man to be twice President of our Institution, Lord Mountbatten has a unique niche in history and it is not an exaggeration to say that he also has a unique niche in the hearts of those who share his enthusiasm for the Institution even if they sometimes cannot keep pace with it. We are probably the only Institution which thinks quickest at the top!'

Mr. M. A. Goutama (Associate Member) expressed the appreciation of the members of the Indian Division to Lord Mountbatten.

After thanking Professor Williams and Mr. Goutama for their tributes Lord Mountbatten said that he had the utmost faith in the future of the Institution and added:

'This gives me a chance to add one more word—one more admonition. The removal of "British" from the title means that we have to redouble our effort in the Commonwealth—it is not a bit of good being insular. I hope that Mr. Thompson and his successors will keep well to the forefront that it is the Engineering Institutions which can bind together the Commonwealth with all the links that are not written in the Statutes of Westminster'.

The President then closed the Annual General Meeting and left the hall to a standing ovation.

After a brief interval Mr. J. L. Thompson, the newly elected President, gave his Presidential Address, published on pages 3–10 of this issue of *The Radio and Electronic Engineer*.

A full report of the business of the Special and Annual General Meetings appears in the January issue of the *Proceedings*.

U.D.C. 621.396.677.45

The Scalar and Vector Potentials of the Infinite Sheath Helical Aerial

By

T. S. M. MACLEAN†

Summary: From consideration of the surface charge density and the linear current density on an infinitely long sheath helix, expressions are derived for the scalar and vector potentials, from which in turn the field quantities are evaluated. By considering one further boundary condition the characteristic equation for the helix is derived.

1. Introduction

The propagation of electromagnetic waves along an infinite helix is usually analysed in terms of an anisotropically conducting cylindrical sheet in which current flow takes place only in the helical direction. The replacement of the physical helix by this sheath model allows the wave equation to be solved subject to the appropriate boundary conditions, and the resulting equations are available in text book form.[1, 2] An alternative method of solution in many aerial problems is to focus attention on the charges and currents flowing in the aerial conductor and from these to derive scalar and vector potentials from which the electric and magnetic fields may be derived directly.[3]

A recent publication[4] has solved this problem for the travelling-wave tube in order to determine the circuit constants of inductance and capacitance for the single tube and for coaxial coupled tubes. In this paper the scalar and vector potential solutions for the single helix are extended to cover the case of the helical aerial where an angular variation of the fields of the form $e^{-j\theta}$ exists. This solution is of practical value for the end-fire aerial, since the radiation from two diametrically opposite points is then in phase in that direction. Experimentally this is found to occur when the circumference is of the order of one wavelength, but the lower and upper limits may be determined quite accurately[5] from the characteristic equation developed in Section 5. Thus a new method of solution is used which is related firmly to physical charges.

2. The Scalar Potential

Consider an infinitely long sheath helix with a surface charge density q coulombs/metre^2 distributed on its cylindrical surface of radius r_1. The wave associated with this charge travels in the positive z direction and follows the helical pitch angle as $e^{-j\beta z}\, e^{-j\theta}$. The scalar potential $d\phi$ at a point $P(r, \theta, z)$, due to an infinitely long narrow blade $r_1 d\theta_1$ of this helix is

$$d\phi(r,\theta,z) = \frac{1}{4\pi\varepsilon}\int_{-\infty}^{+\infty} q r_1\, d\theta_1 \exp\left[\frac{j(\omega t-\beta z_1-\theta_1-kR)}{R}\right] dz_1 \quad \text{......(1)}$$

where k is the free space phase constant, and R is the distance between the point P and a point $Q(r_1\theta_1 z_1)$ on the narrow blade as shown in Fig. 1.

Then

$$R^2 = (z-z_1)^2 + r^2 + r_1^2 - 2rr_1\cos(\theta-\theta_1) \quad \text{......(2)}$$

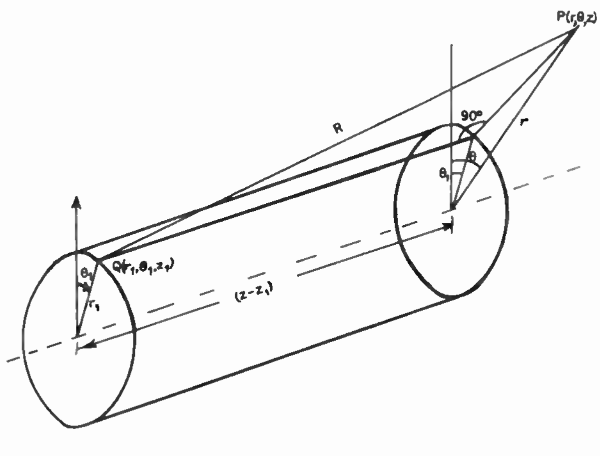


Fig. 1. Cylindrical co-ordinate system for sheath helix.

It is shown in Appendix 9.1 that the integral in eqn. (1) may be evaluated to give

$$d\phi(r,\theta,z) = \frac{qr_1}{2\pi\varepsilon} K_0(a\sqrt{\beta^2-k^2})\exp[j(\omega t-\beta z-\theta)]\, d\theta_1 \quad \text{......(3)}$$

and that integration of the blades round the circumference yields the result

† Electronic and Electrical Engineering Department, University of Birmingham.

$$\phi(r,\theta,z) = \frac{qr_1}{\varepsilon} \begin{matrix} K_1(\gamma r_1)I_1(\gamma r) & r \leqslant r_1 \\ & \\ I_1(\gamma r_1)K_1(\gamma r) & r_1 \leqslant r \end{matrix} \exp[j(\omega t - \beta z - \theta)] \qquad \ldots\ldots(4)$$

In the case of the travelling-wave tube it has been found by Kino and Paik[4] that a transmission line capacitance per unit length can be defined by making use of the equation of continuity $\frac{\partial Jz}{\partial z} + \frac{\partial q}{\partial t} = 0$, but in the case of the helical aerial this is not possible because of the variation of the current in the circumferential as well as the axial direction.

3. The Vector Potentials

Let the linear current density travelling in the helical direction be denoted by $J_{\|}$ amperes/metre. The circumferential and axial components of this current are then

$$J_\theta = J_{\|} \cos\psi \qquad \ldots\ldots(5)$$

and

$$J_z = J_{\|} \sin\psi \qquad \ldots\ldots(6)$$

where ψ is the helical pitch angle which the direction of current flow makes with the plane normal to the axis.

3.1. *The Axial Component A_z of the Vector Potential*

Due to the axial linear current density J_z, there will be an axial component of vector potential dA_z at the point P in Fig. 1 due to the infinite blade of current on which Q lies. This will be given by

$$dA_z = \frac{\mu J_z r_1 \, d\theta_1}{4\pi} \int_{-\infty}^{+\infty} \frac{\exp[j(\omega t - \beta z_1 - \theta_1 - kR)]}{R} dz_1 \qquad \ldots\ldots(7)$$

Equation (7) is of precisely the same form as (1) and consequently the resultant vector potential is

$$A_z = \mu J_z r_1 \begin{matrix} K_1(\gamma r_1)I_1(\gamma r) & r \leqslant r_1 \\ & \\ I_1(\gamma r_1)K_1(\gamma r) & r_1 \leqslant r \end{matrix} \exp[-j\theta] \qquad \ldots\ldots(8)$$

3.2. *The Circumferential Component A_θ and the Radial Component A_r of the Vector Potential*

The circumferential linear current density J_θ flowing in the current element $r_1 d\theta_1$ at the point Q produces an element of vector potential at P in the direction of the unit θ vector at Q. This element may be resolved into a component parallel to the unit $\boldsymbol{\theta}$ vector at P by multiplying it by $\cos(\theta-\theta_1)$, and into another component parallel to the unit $\boldsymbol{r}$ vector at P by multiplying by $-\sin(\theta-\theta_1)$. Thus due to the infinite blade of current on which Q lies

$$dA_\theta = \frac{\mu J_\theta r_1 \, d\theta_1}{4\pi} \cos(\theta-\theta_1) \int_{-\infty}^{+\infty} \frac{\exp[j(\omega t - \beta z_1 - \theta_1 - kR)]}{R} dz_1 \qquad \ldots\ldots(9)$$

and

$$dA_r = -\frac{\mu J_\theta r_1 \, d\theta_1}{4\pi} \sin(\theta-\theta_1) \int_{-\infty}^{+\infty} \frac{\exp[j(\omega t - \beta z_1 - \theta_1 - kR)]}{R} dz_1 \qquad \ldots\ldots(10)$$

These integrals are identical with those which occur with the scalar potential as far as the integration with respect to z_1 is concerned, but the second integrations involving θ become

$$A_\theta = \frac{\mu J_\theta r_1}{2\pi} \exp[j(\omega t - \beta z)] \int_0^{2\pi} K_0[\gamma(r^2 + r_1^2 - 2rr_1\cos(\theta-\theta_1))^{\frac{1}{2}}] \left[\frac{e^{j(\theta-\theta_1)} + e^{-j(\theta-\theta_1)}}{2}\right] e^{-j\theta_1} d\theta_1 \qquad \ldots\ldots(11)$$

and

$$A_r = \frac{j\mu J_\theta r_1}{2\pi} \exp[j(\omega t - \beta z)] \int_0^{2\pi} K_0[\gamma(r^2 + r_1^2 - 2rr_1\cos(\theta-\theta_1))^{\frac{1}{2}}] \left[\frac{e^{+j(\theta-\theta_1)} - e^{-j(\theta-\theta_1)}}{2}\right] e^{-j\theta_1} d\theta \qquad \ldots\ldots(12)$$

Using the addition theorem from reference 7 these give

$$A_\theta = \frac{\mu J_\theta r_1}{2} \begin{bmatrix} K_0(\gamma r_1)I_0(\gamma r) & & K_2(\gamma r_1)I_2(\gamma r) \\ & + & \\ I_0(\gamma r_1)K_0(\gamma r) & & I_2(\gamma r_1)K_2(\gamma r) \end{bmatrix} e^{j(\omega t - \beta z - \theta)} \begin{matrix} r \leqslant r_1 \\ \\ r_1 \leqslant r \end{matrix} \qquad \ldots\ldots(13)$$

and

$$A_r = j\mu \frac{J_\theta r_1}{2} \begin{bmatrix} K_0(\gamma r_1)I_0(\gamma r) & & K_2(\gamma r_1)I_2(\gamma r) \\ & - & \\ I_0(\gamma r_1)K_0(\gamma r) & & I_2(\gamma r_1)K_2(\gamma r) \end{bmatrix} e^{j(\omega t - \beta z - \theta)} \begin{matrix} r \leqslant r_1 \\ \\ r_1 \leqslant r \end{matrix} \qquad \ldots\ldots(14)$$

The result obtained by Kino and Paik for the travelling-wave tube[4] enabled a transmission line inductance per unit length to be defined which causes a fall of potential in the transmission line. This is clearly not possible for the helical aerial since the axial components of current at two diametrically opposite points on the sheath helix cancel, because of the $e^{-j\theta}$ variation in the fields, and consequently there is no resultant axial current in the sheath helix to cause such a fall of potential.

4. Electronic and Magnetic Field Components

The magnetic field components are derived from the vector potentials alone since

$$B = \operatorname{curl} A \qquad \ldots\ldots(15)$$

In terms of cylindrical coordinates these field components are derived in Appendix 9.2 for a given pitch angle ψ and specified linear current density $J_{\parallel}$. The results are in agreement with those which have been derived previously by the wave equation and boundary condition method.

The electric field components require the use of both the scalar and vector potentials, in the form

$$\boldsymbol{E} = -\operatorname{grad} \phi - \frac{\partial A}{\partial t} \qquad \ldots\ldots(16)$$

These separate components for the cylindrical coordinate system are also derived in Appendix 9.2.

5. Derivation of the Characteristic Equation

Although the concept of an equivalent transmission line with known values of L and C has turned out not to be possible for the helical aerial, the phase velocity can still be determined from the characteristic equation. To derive this, use is made of one boundary condition only instead of the four which are required explicitly by the wave equation method. Since the electric field at the helix is zero in the direction of current flow, for a perfect conductor,

$$\boldsymbol{E}_{zr_1} \sin\psi + \boldsymbol{E}_{\theta r_1} \cos\psi = 0 \qquad \ldots\ldots(17)$$

If these two components of electric field are derived from the scalar and vector potentials directly, an additional relationship between these scalar and vector potentials is also required since the phase velocity must be independent of both charge and current. This relationship is given by

$$\operatorname{div} A = -\mu\varepsilon \frac{\partial \phi}{\partial t} \qquad \ldots\ldots(18)$$

which from equations (8), (13), (14) and (4) results in

$$q = \frac{(\beta r_1 + \cot\psi)}{\omega r_1} J_z \qquad \ldots\ldots(19)$$

Making use of this result in eqn. (17) and simplifying algebraically gives the characteristic equation

$$\frac{(\gamma^2 r_1^2 + \beta r_1 \cot\psi)^2}{k^2 r_1^2 \gamma^2 r_1^2 \cot^2\psi} = -\frac{I_1'(\gamma r_1) K_1'(\gamma r_1)}{I_1(\gamma r_1) K_1(\gamma r_1)} \qquad \ldots\ldots(20)$$

From this equation the phase velocity of the different possible waves[5] which travel along the sheath helix can be evaluated. These waves are either four or two in number depending on the circumference of the helix, and the usefulness of the helical aerial as a circularly polarized device begins at the transition between them. Although two waves exist above this point, consideration of the power carried by them[5] shows that only one of them is significant, thus corroborating experimental measurements. The transition point may be shown to be given to a useful approximation by[8]

$$\begin{array}{c}\text{circumference in wavelengths}\\ \text{(lower frequency)}\end{array} = \frac{\cos\psi}{1 + \sin\psi}$$

where ψ is the pitch angle. There is also an upper frequency limit imposed either by pattern break-up or the existence of other modes, but for aerials of small size, not greater than, say, two wavelengths in axial length, it is given approximately by

$$\begin{array}{c}\text{circumference in wavelengths}\\ \text{(upper frequency)}\end{array} = \frac{\cos\psi}{1 - \sin\psi}$$

These frequencies then define the useful working range of the aerial, which increases with pitch angle up to a maximum of about 15 deg imposed by array factor considerations.

6. Conclusions

An alternative to the wave equation method of examining the helical aerial has been derived, which concentrates attention on the physical charges and currents, rather than on the mathematical solution of the wave equation.

7. Acknowledgments

Conversations with Mr. J. Layton and Dr. V. G. Welsby contributed to the development of this paper and grateful acknowledgment is made to them.

8. References

1. D. A. Watkins, "Topics in Electromagnetic Theory" (Wiley, New York, 1958).
2. R. E. Collin, "Field Theory of Guided Waves" (McGraw-Hill, New York, 1960).
3. E. B. Moullin, "Radio Aerials" (Oxford, Clarendon Press, 1949).
4. G. S. Kino and S. F. Paik, "Circuit theory of coupled transmission systems", *J. Appl. Phys.*, 33, p. 3002, October 1962.
5. T. S. M. Maclean and W. E. J. Farvis, "The sheath-helix approach to the helical aerial", *Proc. Instn Elect. Engrs*, **109**, Part C, p. 548, September 1962 (I.E.E. Monograph No. 519E).
6. W. Magnus and F. Oberhettinger, "Formulas and Theorems for the Functions of Mathematical Physics", p. 118 (Chelsea, 1954).

7. A. Erdelyi, "High Transcendental Functions", Vol. 2, p. 102 (McGraw-Hill, New York, 1953).

8. T. S. M. Maclean, "An engineering study of the helical aerial", *Proc. Instn Elect. Engrs*, **110**, p. 112, January 1963 (I.E.E. Paper No. 4029E).

9. Appendix

9.1. *Evaluation of Integrals*

It is required to evaluate first

$$d\phi(r,\theta,z) = \frac{1}{4\pi\varepsilon}\int_{-\infty}^{+\infty} qr_1\, d\theta_1 \frac{\exp j\{\omega t-\beta z_1-\theta_1-k[(z-z_1)^2+r^2+r_1^2-2rr_1\cos(\theta-\theta_1)]^{\frac{1}{2}}\}}{[(z-z_1)^2+r^2+r_1^2-2rr_1\cos(\theta-\theta_1)]^{\frac{1}{2}}}\, dz_1 \quad \ldots\ldots(21)$$

Then changing the variable by letting $z' = z-z_1$

$$d\phi(r,\theta,z) = -\frac{\exp[-j(\beta z+\theta_1)]}{4\pi\varepsilon}\int_{-\infty}^{+\infty} qr_1\, d\theta_1 \frac{\exp j\{\omega t+\beta z'-k[z'^2+r^2+r_1^2-2rr_1\cos(\theta-\theta_1)]^{\frac{1}{2}}\}}{[z'^2+r^2+r_1^2-2rr_1\cos(\theta-\theta_1)]^{\frac{1}{2}}}\, dz' \quad \ldots\ldots(22)$$

$$= -\frac{qr_1\, d\theta_1 \exp j(\omega t-\beta z-\theta_1)}{4\pi\varepsilon}\int_{-\infty}^{+\infty}\frac{\exp j\{\beta z'-k[z'^2+a^2]^{\frac{1}{2}}\}}{(z'^2+a^2)^{\frac{1}{2}}}\, dz' \quad \ldots\ldots(23)$$

where $\quad a^2 = r^2+r_1^2-2rr_1\cos(\theta-\theta_1)$

This is a Fourier Transform which is tabulated by Magnus and Oberhettinger.[6] The result is

$$d\phi(r,\theta,z) = \frac{qr_1}{4\pi\varepsilon}\pi i H_0'[a\sqrt{k^2-\beta^2}] \times$$
$$\times \exp[j(\omega t-\beta z-\theta_1)]\, d\theta_1 \quad \ldots\ldots(24)$$

For a slow wave solution which is required for end-fire aerials, the phase constant β must be greater than k, so that the potential may be written

$$d\phi(r,\theta,z) = \frac{2qr_1}{4\pi\varepsilon}K_0(a\sqrt{\beta^2-k^2}) \times$$
$$\times \exp[j(\omega t-\beta z-\theta_1)]\, d\theta_1 \quad \ldots\ldots(25)$$

Integrating over the circumference of the helix gives

$$\phi(r,\theta,z) = \int_0^{2\pi}\frac{2qr_1}{4\pi\varepsilon}K_0[\gamma\{r^2+r_1^2-2rr_1\cos(\theta-\theta_1)\}^{\frac{1}{2}}] \times$$
$$\times \exp[j(\omega t-\beta z-\theta_1)]\, d\theta_1 \quad \ldots\ldots(26)$$

where $\qquad \gamma^2 = \beta^2-k^2$

Equation (26) may be evaluated using the addition theorem for the modified Bessel Function[7]

$$K_0[\gamma\{r^2+r_1^2-2rr_1\cos(\theta-\theta_1)\}^{\frac{1}{2}}]$$
$$= \sum_{-\infty}^{+\infty} K_n(\gamma r_1)I_n(\gamma r)\exp[jn(\theta-\theta_1)], \quad r \leqslant r_1 \quad \ldots\ldots(27)$$

and

$$\sum_{-\infty}^{+\infty} I_n(\gamma r_1)K_n(\gamma r)\exp[jn(\theta-\theta_1)], \quad r_1 \leqslant r \quad \ldots\ldots(28)$$

Then after integration only the terms with $(n+1)$ equal to zero remain, since the limits are zero and 2π, so that

$$\phi(r,\theta,z) = \frac{qr_1}{\varepsilon}\begin{cases} K_1(\gamma r_1)I_1(\gamma r) & r \leqslant r_1 \ldots\ldots(29) \\ & \exp[j(\omega t-\beta z-\theta)r] \\ I_1(\gamma r_1)K_1(\gamma r) & r_1 \leqslant r \ldots\ldots(30)\end{cases}$$

9.2. *Derivation of Electric and Magnetic Fields*

From eqn. (15)

$$B = \text{curl}\, A$$

the following components in cylindrical coordinates may be derived

$$B_\theta = \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r} \quad \ldots\ldots(31)$$

$$B_r = \frac{1}{r}\frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z} \quad \ldots\ldots(32)$$

and

$$B_z = \frac{1}{r}\left[\frac{\partial(rA_\theta)}{\partial r} - \frac{\partial A_r}{\partial \theta}\right] \quad \ldots\ldots(33)$$

Substituting for the vector potentials from eqns. (8), (13) and (14) and simplifying algebraically leads to the following expressions for the internal fields

$$B_\theta = -\mu\left[\frac{\gamma^2 r_1^2+\beta r_1\cot\psi}{\gamma r_1\cot\psi}K_1(\gamma r_1)I_1'(\gamma r) + \right.$$
$$\left. + \frac{\beta r_1}{\gamma r}K_1'(\gamma r_1)I_1(\gamma r)\right] J_{\parallel}\cos\psi \quad \ldots\ldots(34)$$

$$B_r = -j\mu\left[\frac{\gamma^2 r_1^2+\beta r_1\cot\psi}{\gamma^2 r_1^2\cot\psi}K_1(\gamma r_1)I_1(\gamma r) + \right.$$
$$\left. + \beta r_1 K_1'(\gamma r_1)I_1'(\gamma r)\right] J_{\parallel}\cos\psi \quad \ldots\ldots(35)$$

and

$$B_z = -\mu\gamma r_1 K_1'(\gamma r_1)I_1(\gamma r)J_{\parallel}\cos\psi \quad \ldots\ldots(36)$$

The fields external to the helix are of identical form with the I and K functions interchanged.

Similarly from eqn. (16)

$$E = -\operatorname{grad}\phi - \frac{\partial A}{\partial t}$$

the electric field vectors may be written

$$E_\theta = -\frac{1}{r}\frac{\partial \phi}{\partial \theta} - \frac{\partial A_\theta}{\partial t} \qquad \ldots\ldots(37)$$

$$E_r = -\frac{\partial \phi}{\partial r} - \frac{\partial A_r}{\partial t} \qquad \ldots\ldots(38)$$

and

$$E_z = -\frac{\partial \phi}{\partial z} - \frac{\partial A_z}{\partial t} \qquad \ldots\ldots(39)$$

Hence making use of eqn. (18) in cylindrical co-ordinate form

$$\frac{1}{r}\frac{\partial}{\partial r}(rA_r) + \frac{1}{r}\frac{\partial A_\theta}{\partial \theta} + \frac{\partial A_z}{\partial z} = -\varepsilon\mu\frac{\partial \phi}{\partial t} \qquad \ldots\ldots(40)$$

the relation between the charge density q and the linear current density J_z given by eqn. (19) may be established and the following equations obtained for the internal fields:

$$E_\theta = \mathrm{j}\omega\mu r_1\left[\frac{\beta r_1}{r}\frac{(\beta r_1\cot\psi + \gamma^2 r_1^2)}{k^2 r_1^2 \gamma^2 r_1^2 \cot\psi} K_1(\gamma r_1) I_1(\gamma r) + K_1'(\gamma r_1) I_1'(\gamma r)\right] J_\parallel \cos\psi \qquad \ldots\ldots(41)$$

$$E_r = \frac{-\omega\mu}{\gamma}\left[\frac{\beta r_1(\gamma^2 r_1^2 + \beta r_1 \cot\psi)}{k^2 r_1^2 \cot\psi} K_1(\gamma r_1) I_1'(\gamma r) + \frac{r_1}{r} K_1'(\gamma r_1) I_1(\gamma r)\right] J_\parallel \cos\psi \qquad \ldots\ldots(42)$$

and

$$E_z = \mathrm{j}\omega\mu r_1 \frac{\gamma^2 r_1^2 + \beta r_1 \cot\psi}{k^2 r_1^2} K_1(\gamma r_1) I_1(\gamma r) J_\parallel \sin\psi \qquad \ldots\ldots(43)$$

The external fields follow as before by interchanging the I and K functions.

Manuscript first received by the Institution on 24th May 1963 and in final form on 16th August 1963. (Paper No. 874.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for

December 1963

1963 December	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	1963 December	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.
1	− 130·2	− 131·0	+ 15	17	− 129·3	− 129·8	+ 22
2	− 130·0	− 130·0	+ 15	18	− 130·5	− 130·8	+ 21
3	− 130·1	—	+ 14	19	− 129·5	− 129·6	+ 23
4	− 130·2	− 130·2	+ 16	20	− 130·8	− 130·1	+ 23
5	− 130·4	− 130·4	+ 16	21	− 130·9	− 129·4	+ 24
6	− 130·0	− 129·8	+ 16	22	− 130·5	—	—
7	− 130·5	—	+ 18	23	− 130·8	—	+ 24
8	− 130·5	− 131·8	—	24	− 130·7	− 131·1	+ 24
9	—	− 130·3	+ 18	25	− 130·4	− 130·4	+ 25
10	—	− 130·3	+ 20	26	− 130·5	− 129·9	+ 25
11	− 130·3	− 130·2	+ 19	27	− 130·0	− 128·8	+ 26
12	− 130·6	− 130·7	+ 21	28	− 129·2	− 128·8	+ 26
13	− 130·6	− 129·9	+ 19	29	− 129·6	− 129·2	—
14	− 130·2	− 130·1	+ 21	30	− 128·8	—	+ 26
15	− 130·5	− 130·1	+ 21	31	− 129·3	− 128·6	+ 27
16	− 129·7	− 130·6	+ 22				

Nominal frequency corresponds to a value of 9 192 631 770 c/s for the caesium F,m (4,0)–F,m (3,0) transition at zero magnetic field.

Notes: (1) the MSF/GBR frequency offset will be adjusted to -150×10^{-10} at 0000 U.T. on the 1st January 1964.
(2) the GBR transmission was not controlled by MSF oscillator for the period 1500–1850 U.T. on 24th December 1963.

Trans-Pacific Telecommunications

At the inaugural ceremony in London on 2nd December of the opening of COMPAC, the Pacific Ocean section of the Commonwealth telephone cable, the recorded voice of Her Majesty the Queen travelled 16000 miles to guests at ceremonies in Ottawa, Sydney, Wellington and Suva (Fiji).

COMPAC is the first telephone cable to be laid across the Pacific Ocean and cost £26 million. It links Canada with Australia and New Zealand by way of Hawaii and Fiji, a distance of about 9400 statute miles, and forms the second link in the Commonwealth telephone cable system agreed at the Commonwealth Telecommunications Conference in 1958. CANTAT, which was laid between Britain and Canada in 1961, and trans-Canadian microwave and land-line circuits, form the link with COMPAC to provide for the first time high quality telephone circuits to the other side of the world which are not dependent on high-frequency radio.

The COMPAC project has been completed in four stages. The first section between Australia and New Zealand was completed in June 1962 and the section between New Zealand and Fiji in November 1962. The cable laying operations for the sections Fiji to Hawaii and Hawaii to Canada were completed during October this year. Immediately following completion of the laying operations of each section there was a period of intensive testing during which the section operating levels were set to give the design margins against overload and noise.

Differences in clock time and hence differences in peak traffic periods between places along the route enable efficient use to be made of the circuit capacity. For example, trans-Atlantic circuits can be used during Britain's morning for traffic to Australia, during Britain's afternoon, when Australia is asleep, for calls to Canada and in the evening for calls to New Zealand for whom the next day has just begun. G.P.O. operators in London will soon be able to dial direct via the cable to any Australian subscriber served by an automatic exchange; operators in Sydney and Auckland can now dial direct into the inland telephone system in Britain.

The maximum number of telephone circuits that can be provided is 80, many of which will carry multi-channel telegraph systems, facsimile and data transmissions; alternatively, a number of channels can be combined to allow music programme and cable-film transmissions. The two directions of transmission are separated on a frequency basis into the bands 60–300 kc/s and 360–608 kc/s, each direction of transmission accommodating five 48-kc/s groups. The transmission performance of the groups is such that either 12-channel (4-kc/s spacing) or 16-channel (3-kc/s spacing) translating equipment may be used.

The system was designed to meet a noise objective of 1 pW per kilometre averaged over all channels in each direction of transmission; this objective was achieved on all the sections of the COMPAC system. The attenuation/frequency response of groups over any one section of the COMPAC system is within 1·0 dB of nominal. The cable will provide groups of circuits between the various points through which it passes, the longest group being one of 12 circuits between London and Sydney. The attenuation/frequency response of this group is within 2·0 dB of nominal.

In the design, engineering and manufacture of such a system the over-riding requirement is reliability and the methods and equipment employed had to meet the highest possible standards. The majority of the cable used on the COMPAC system is a radically new type of deep sea cable, first used on the CANTAT transatlantic cable, and is unarmoured. The strength member, essential to withstand the tensions during laying and recovery operations, is in the form of high-tensile torsionally-balanced steel strands situated at the centre of the cable and within the inner copper conductor. The inner conductor consists of a thin longitudinal copper tape, box-seamed tightly on to the steel strands. Polythene insulation is extruded on to the conductor and a coaxial return conductor of aluminium tapes is applied. The cable is covered with a sheath of polythene which gives adequate protection in depths greater than 500 fathoms where the water is generally quiet and no interference from fishing trawls or anchors is encountered. This type of cable is known as 'lightweight' cable because its weight in water is only one-fifth of that of a comparable armoured cable. Its attenuation at 608 kc/s is 1·8 dB per statute mile.

Submerged repeaters (totalling 318) are laid at approximately 30-statute-mile intervals along the deep-sea section and at approximately 20-statute-mile intervals in the shallow section of the cable. The basic design of every component and circuit is directed towards giving the longest possible life of the repeater. For example, surfaces are gold-plated to minimize the growth of metal 'whiskers' during the long, undisturbed life of the repeaters on the ocean bed. The performance of the cable, repeaters and transmission paths is monitored by means of pilot signals and repeater monitoring circuits. During laying operations by H.M.T.S. *Monarch* and C.S. *Mercury* equalizers were designed and constructed on the cable ships and inserted at intervals along the cable to maintain the system misalignment within design limits. C.S. *Retriever* will be based in the Pacific Ocean and will carry out any necessary maintenance work on the cable.

A Commonwealth partnership of Australia, Canada, New Zealand and Britain has been responsible for COMPAC. This £26 million project is owned by the Overseas Telecommunications Commission (Australia), Canadian Overseas Telecommunication Corporation, New Zealand Post Office and Cable & Wireless Ltd.; a Management Committee of senior representatives of the partners has been in overall control of design, development and construction. The engineering has been largely undertaken by the British Post Office, working closely with a team of experts of the four countries.

A further telephone cable link with Commonwealth countries will be completed in due course by an extension from Sydney to Hong Kong and Singapore (SEACOM).

U.D.C. 621.384.633

Modes of Oscillation of a Grounded Grid Cyclotron Oscillator

By

D. N. BASU MALLIK, M.Sc.(Tech.)†

AND

B. R. NAG, M.Sc.(Tech.), M.S., D.Phil.‡

Summary: A linear analysis of a grounded-grid cyclotron oscillator considering the effect of all the important circuit elements is presented. It is shown that the oscillator may execute oscillations at one of the cyclotron-dee resonant frequencies and/or at one of the coupling transmission line frequencies. The conditions of excitation of these oscillations are derived. Experimental evidence obtained from a model oscillator verifying the above results are also given.

1. Introduction

Grounded grid oscillators have been used widely in the last few years as a source of r.f. power for cyclotrons.[1-10] The principle of operation and design features of grounded grid oscillators have been discussed by Backus[1] and Relf.[2] In such oscillators, usually there are two capacitively-coupled tuned circuits connected to the oscillator tube by two transmission lines. The oscillator has, therefore, four possible frequencies of oscillation of which only one is the desired frequency. In the design procedure described by Backus and others, main emphasis is laid on the calculation of circuit parameters for the optimum operation at the desired frequency. A general discussion of the other oscillator modes and the conditions under which they are excited is not generally available in published literature. This is possibly due to the distributed nature of the oscillator circuit elements which makes the analysis complicated. However, for the cyclotron oscillator under construction at Saha Institute of Nuclear Physics, Calcutta, and some other existing oscillators,[9] the distributed elements have natural resonant frequencies much above the operation frequency. In such cases one may replace the distributed elements by lumped equivalent elements and a general circuit analysis becomes possible. The authors have made an analysis of the oscillator assuming the above condition and examined different modes of the oscillator. The analytical results have been verified by experiments on a model oscillator. This analysis and the experimental results obtained are presented in this paper.

In Section 2, the oscillator circuit arrangement is described and an equivalent circuit is given. In the equivalent circuit all the essential circuit parameters are included but the distributed parameter elements are replaced by lumped elements.

In Section 3, a linear analysis of the circuit is presented and general expressions for the frequencies of the oscillations and the conditions for their stability are obtained.

In Sections 4 and 5, the above general expressions are examined with reference to practical circuit parameter values. It is shown that of the four possible frequencies of oscillation, oscillations at two frequencies can only be maintained. Of these two frequencies one is undesirable and the condition for its elimination is obtained.

In the concluding section experimental results, obtained from a model oscillator are presented which generally confirm the analytical results.

2. The Oscillator and the Equivalent Circuit

The circuit arrangement of the oscillator is shown in Fig. 1. The anode and the cathode of the oscillator tube are connected by transmission lines through mutual inductances to the cyclotron-dee system, the grid of the tube being grounded for r.f. The circuit which is assumed to be equivalent to the oscillator arrangement is shown in Fig. 2. The dee system has been replaced by two parallel tuned circuits with 'lumped' capacitances and inductances C_1, C_2 and L_1, L_2 respectively. The lumped capacitances are those provided by the dee plates, and the inductances are those provided by the dee stems. The resistances R_1, R_2 shown in parallel with tuned circuits represent the losses in the dee system including ion loss and beam loading. The anode and cathode lines together with the coupling loops are replaced by two L-networks, consisting of lossless inductances and capacitances, L_3, L_4 and C_3, C_4 respectively. The mutual inductances of the coupling loops are M_1 and M_2. The effects of these mutual inductances on the line parameters are included in L_3 and L_4. The grid-anode and cathode-grid capacitances of the tube are included in C_3 and C_4 respectively; C_{ak} represents the anode-cathode capacitance of the tube.

† Saha Institute of Nuclear Physics, Calcutta.

‡ Institute of Radio Physics & Electronics, Calcutta.

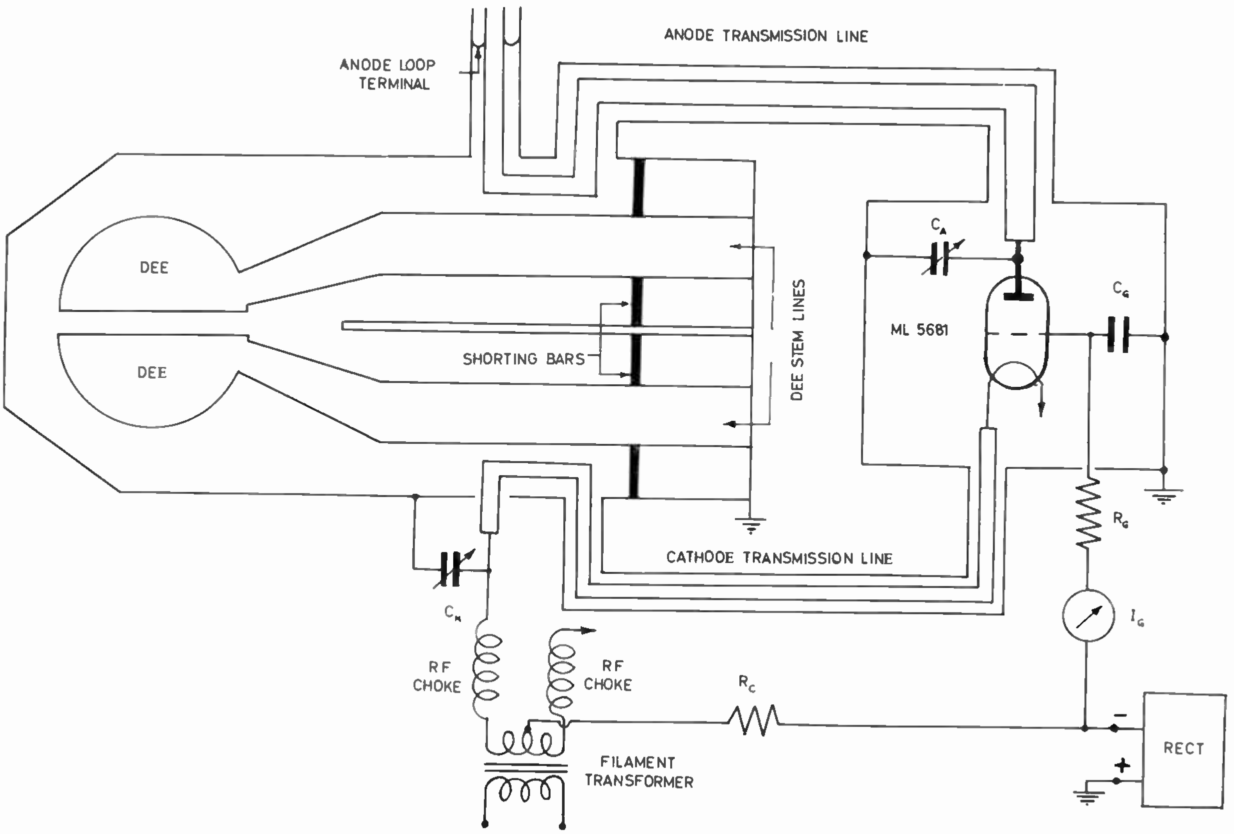


Fig. 1. Circuit arrangement of the oscillator.

3. Differential Equations Describing the Oscillator

Let V_K and V_A represent respectively the voltages at the cathode and anode of the tube with respect to ground and V_{C1}, V_{C2} represent the voltages across the dee capacitances. These voltages are related to V_K and V_A by the equations

$$V_A = \left\{(L_1 - M_1) + \frac{L_1 L_3}{M_1}\right\}\left[C_1 \frac{d^2 V_{C_1}}{dt^2} + \frac{1}{R_1}\frac{dV_{C_1}}{dt} + \right.$$

$$\left. + C_C \frac{d^2 V_{C_1}}{dt^2} - C_C \frac{d^2 V_{C_2}}{dt^2}\right] + V_{C_1}\left(1 + \frac{L_3}{M_1}\right) \quad \ldots(1)$$

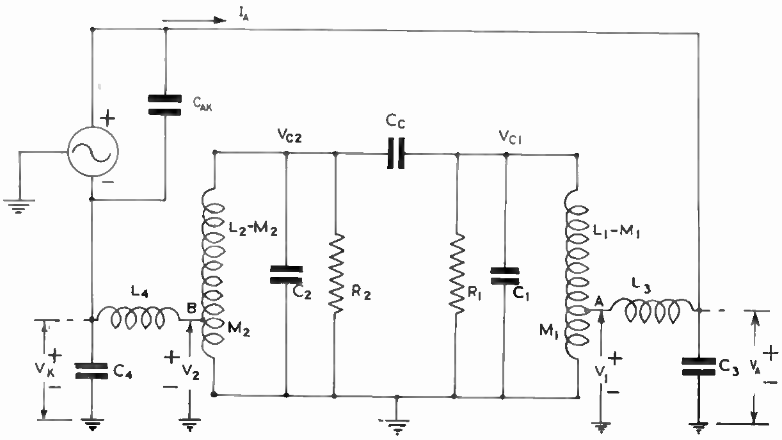


Fig. 2. Equivalent circuit of the oscillator.

$$V_K = \left\{(L_2 - M_2) + \frac{L_2 L_4}{M_2}\right\}\left[C_2 \frac{d^2 V_{C_2}}{dt^2} + \frac{1}{R_2}\frac{dV_{C_2}}{dt} + \right.$$
$$\left. + C_C \frac{d^2 V_{C_2}}{dt^2} - C_C \frac{d^2 V_{C_1}}{dt^2}\right] + V_{C_2}\left(1 + \frac{L_4}{M_2}\right) \quad \ldots(2)$$

The current through the tube is a function of the grid and anode voltages, and may be denoted by

$$I_A = f_A(V_A, V_K) \qquad \ldots\ldots(3)$$

The grid current is represented by

$$I_g = f_g(V_A, V_K) \qquad \ldots\ldots(4)$$

On equating the currents at the anode and cathode nodes one obtains

$$\frac{V_1}{M_1} + \frac{V_1 - V_{C_1}}{L_1 - M_1} + (C_3 + C_{ak})\frac{d^2 V_A}{dt^2} - C_{ak}\frac{d^2 V_K}{dt^2}$$
$$= \left[\frac{d}{dt} f_A(V_A, V_K)\right] \qquad \ldots\ldots(5)$$

$$-\frac{V_2}{M_2} - \frac{V_2 - V_{C_2}}{L_2 - M_2} - (C_4 + C_{ak})\frac{d^2 V_K}{dt^2} + C_{ak}\frac{d^2 V_A}{dt^2}$$
$$= \left[\frac{d}{dt} f_A(V_A, V_K) + \frac{d}{dt} f_g(V_A, V_K)\right] \qquad \ldots\ldots(6)$$

Similarly considering the currents at the points A and B.

$$\frac{(V_A - V_1)}{L_3} = \frac{V_1}{M_1} + \frac{(V_1 - V_{C_1})}{(L_1 - M_1)} \qquad \ldots\ldots(7)$$

$$\frac{(V_K - V_2)}{L_4} = \frac{V_2}{M_2} + \frac{(V_2 - V_{C_2})}{(L_2 - M_2)} \qquad \ldots\ldots(8)$$

Eliminating V_{C_1} and V_{C_2} from eqns. (1) and (2) by using eqns. (5) through (6) and putting

$$a_1 = \frac{1}{R_1(C_1 + C_C)}, \qquad a_2 = \frac{1}{R_2(C_2 + C_C)}$$

$$T_1 = \frac{L_1}{M_1}, \qquad T_2 = \frac{L_2}{M_2}$$

$$K_1 = \frac{C_C}{C_1 + C_C}, \qquad K_2 = \frac{C_C}{C_2 + C_C}$$

$$K_3 = \frac{C_{ak}}{(C_3 + C_{ak})}, \qquad K_4 = \frac{C_{ak}}{(C_4 + C_{ak})}$$

$$\omega_1^2 = \frac{1}{L_1(C_1 + C_C)}, \qquad \omega_2^2 = \frac{1}{L_2(C_2 + C_C)}$$

$$\frac{1}{\omega_3^2} = \left[L_3 + \frac{M_1(L_1 - M_1)}{L_1}\right](C_3 + C_{ak})$$

$$\frac{1}{\omega_4^2} = \left[L_4 + \frac{M_2(L_2 - M_2)}{L_2}\right](C_4 + C_{ak})$$

$$\frac{1}{\omega_5^2} = (L_3 + M_1)(C_3 + C_{ak})$$

$$\frac{1}{\omega_6^2} = (L_4 + M_2)(C_4 + C_{ak})$$

one obtains

$$T_1\left[\frac{d^2}{dt^2} + a_1\frac{d}{dt} + \frac{\omega_1^2 T_1}{(T_1 - 1)}\right]\left[\left(1 + \frac{1}{\omega_3^2}\frac{d^2}{dt^2}\right)V_A - \frac{K_3}{\omega_3^2}\frac{d^2 V_K}{dt^2} - \frac{1}{\omega_3^2(C_3 + C_{ak})}\frac{d}{dt} f_A(V_A, V_K)\right] -$$
$$-\frac{\omega_1^2 T_1}{(T_1 - 1)}\left[V_A + \left\{\frac{T_1}{\omega_3^2} - \frac{(T_1 - 1)}{\omega_5^2}\right\}\left\{\frac{d^2 V_A}{dt^2} - K_3\frac{d^2 V_K}{dt^2} - \frac{1}{(C_3 + C_{ak})}\frac{d}{dt} f_A(V_A, V_K)\right\}\right] -$$
$$-K_1 T_2 \frac{d^2}{dt^2}\left[\left(1 + \frac{1}{\omega_4^2}\frac{d^2}{dt^2}\right)V_K - \frac{K_4}{\omega_4^2}\frac{d^2 V_A}{dt^2} + \frac{1}{\omega_4^2(C_4 + C_{ak})}\cdot\frac{d}{dt}\{f_A(V_A, V_K) + f_g(V_A, V_K)\}\right] = 0 \qquad (9)$$

$$-K_2 T_1 \frac{d^2}{dt^2}\left[\left(1 + \frac{1}{\omega_3^2}\frac{d^2}{dt^2}\right)V_A - \frac{K_3}{\omega_3^2}\frac{d^2 V_K}{dt^2} - \frac{1}{\omega_3^2(C_3 + C_{ak})}\cdot\frac{d}{dt} f_A(V_A, V_K)\right] +$$
$$+T_2\left[\frac{d^2}{dt^2} + a_2\frac{d}{dt} + \frac{\omega_2^2 T_2}{(T_2 - 1)}\right]\left[\left(1 + \frac{1}{\omega_4^2}\frac{d^2}{dt^2}\right)V_K - \frac{K_4}{\omega_4^2}\cdot\frac{d^2 V_A}{dt^2} + \frac{1}{\omega_4^2(C_4 + C_{ak})}\cdot\frac{d}{dt}\{f_A(V_A, V_K) + f_g(V_A, V_K)\}\right] -$$
$$-\frac{\omega_2^2 T_2}{(T_2 - 1)}\left[V_K + \left\{\frac{T_2}{\omega_4^2} - \frac{(T_2 - 1)}{\omega_6^2}\right\}\left\{\frac{d^2 V_K}{dt^2} - K_4\frac{d^2 V_A}{dt^2} + \frac{1}{(C_4 + C_{ak})}\cdot\frac{d}{dt}\{f_A(V_A, V_K) + f_g(V_A, V_K)\}\right\}\right] = 0$$
$$\ldots\ldots(10)$$

Equations (9) and (10) are the describing differential equations of the oscillator based on the equivalent circuit of Fig. 2.

The exact form of the functions f_a and f_g depends on the tube and its operating conditions. However, in the steady state of oscillation, one may write

$$f_a(V_A, V_K) = \left(g_m + \frac{1}{r_a}\right) V_K - \frac{V_A}{r_a} \quad \ldots\ldots(11)$$

and

$$f_g(V_A, V_K) = \frac{V_K}{r_g} \quad \ldots\ldots(12)$$

where g_m, r_a, r_g represent the effective values of the mutual conductance, anode resistance and grid resistance of the tube. These may be evaluated considering only fundamental components of the anode current and grid current in the steady oscillating condition.

Now, on putting

$$b_1 = \left(g_m + \frac{1}{r_a}\right) \frac{1}{\omega_3^2(C_3 + C_{ak})}$$

$$b_2 = \left(g_m + \frac{1}{r_a} + \frac{1}{r_g}\right) \frac{1}{\omega_4^2(C_4 + C_{ak})}$$

$$b_3 = \frac{1}{\omega_3^2(C_3 + C_{ak}) r_a}$$

$$b_4 = \frac{1}{\omega_4^2(C_4 + C_{ak}) r_a} \quad \ldots\ldots(13)$$

one obtains

$$T_1 \left[\left(\frac{1}{\omega_3^2} + \frac{K_1 K_4}{\omega_4^2} \cdot \frac{T_2}{T_1}\right) \frac{d^4}{dt^4} + \left(b_3 + \frac{a_1}{\omega_3^2} + K_1 b_4 \frac{T_2}{T_1}\right) \frac{d^3}{dt^3} + \left(1 + a_1 b_3 + \frac{\omega_1^2}{\omega_5^2}\right) \frac{d^2}{dt^2} + \left(a_1 + b_3 \frac{\omega_1^2 . \omega_3^2}{\omega_5^2}\right) \times \right.$$
$$\left. \times \frac{d}{dt} + \omega_1^2 \right] V_A - T_2 \left[\left(\frac{K_1}{\omega_4^2} + \frac{K_3}{\omega_3^2} \cdot \frac{T_1}{T_2}\right) \frac{d^4}{dt^4} + \left(K_1 b_2 + b_1 \frac{T_1}{T_2} + a_1 \frac{K_3}{\omega_3^2} \cdot \frac{T_1}{T_2}\right) \frac{d^3}{dt^3} + \right.$$
$$\left. + \left(K_1 + a_1 b_1 \frac{T_1}{T_2} + K_3 \frac{\omega_1^2}{\omega_5^2} \cdot \frac{T_1}{T_2}\right) \frac{d^2}{dt^2} + \left(\frac{\omega_1^2 \omega_3^2}{\omega_5^2} \cdot \frac{T_1}{T_2} \cdot b_1\right) \frac{d}{dt}\right] V_K = 0 \quad \ldots\ldots(14)$$

$$-T_1 \left[\left(\frac{K_2}{\omega_3^2} + \frac{K_4}{\omega_4^2} \cdot \frac{T_2}{T_1}\right) \frac{d^4}{dt^4} + \left(K_2 b_3 + b_4 \frac{T_2}{T_1} + a_2 \frac{K_4}{\omega_4^2} \cdot \frac{T_2}{T_1}\right) \frac{d^3}{dt^3} + \left(K_2 + a_2 b_4 \frac{T_2}{T_1} + K_4 \frac{\omega_2^2}{\omega_6^2} \cdot \frac{T_2}{T_1}\right) \frac{d^2}{dt^2} + \right.$$
$$\left. + \left(\frac{\omega_2^2 \omega_4^2}{\omega_6^2} \cdot \frac{T_2}{T_2} \cdot b_4\right) \frac{d}{dt}\right] V_A + T_2 \left[\left(\frac{1}{\omega_4^2} + \frac{K_1 K_3}{\omega_3^2} \cdot \frac{T_1}{T_2}\right) \frac{d^4}{dt^4} + \left(b_2 + \frac{a_2}{\omega_4^2} + K_2 b_1 \frac{T_1}{T_2}\right) \frac{d^3}{dt^3} + \right.$$
$$\left. + \left(1 + a_2 b_2 + \frac{\omega_2^2}{\omega_6^2}\right) \frac{d^2}{dt^2} + \left(a_2 + b_2 . \frac{\omega_2^2 \omega_4^2}{\omega_6^2}\right) \frac{d}{dt} + \omega_2^2\right] V_K = 0 \quad \ldots\ldots(15)$$

On combining eqns. (14) and (15), replacing d/dt by $j\omega$ and equating real and imaginary parts to zero, one obtains

$$[(1 - K_1 K_2)(1 - K_3 K_4)]\omega^8 - \left[(1 - K_1 K_2)(\omega_3^2 + \omega_4^2) + (1 - K_3 K_4)\left(\frac{\omega_1^2 \omega_3^2}{\omega_5^2} + \frac{\omega_2^2 \omega_4^2}{\omega_6^2}\right)\right] \omega^6 +$$
$$+ \left[(1 - K_1 K_2)\omega_3^2 \omega_4^2 + \omega_1^2 \omega_3^2 + \omega_2^2 \omega_4^2 + \omega_3^2 \omega_4^2 \left(\frac{\omega_1^2}{\omega_5^2} + \frac{\omega_2^2}{\omega_6^2}\right) + (1 - K_3 K_4) \frac{\omega_3^2 \omega_4^2}{\omega_5^2 \omega_6^2} \cdot \omega_1^2 \omega_2^2 + \right.$$
$$\left. + \omega_1^2 K_2 K_3 \frac{T_1}{T_2}\left(\omega_4^2 - \frac{\omega_3^2 \omega_4^2}{\omega_5^2}\right) + \omega_2^2 K_1 K_4 \frac{T_2}{T_1}\left(\omega_3^2 - \frac{\omega_3^2 \omega_4^2}{\omega_6^2}\right)\right] \omega^4 - \left[\omega_3^2 \omega_4^2 (\omega_1^2 + \omega_2^2) + \right.$$
$$\left. + \omega_1^2 \omega_2^2 \omega_3^2 \omega_4^2 \left(\frac{1}{\omega_5^2} + \frac{1}{\omega_6^2}\right)\right] \omega^2 + \omega_1^2 \omega_2^2 \omega_3^2 \omega_4^2 = 0 \quad \ldots\ldots(16)$$

$$a_1\left[\frac{(1-K_3K_4)}{\omega_3^2\omega_4^2}\cdot\omega^6-\left(\frac{1}{\omega_3^2}+\frac{1}{\omega_4^2}\right)\omega^4-\frac{(1-K_3K_4)\omega_2^2}{\omega_3^2\omega_6^2}\cdot\omega^4+\omega^2+\omega_2^2\left(\frac{1}{\omega_3^2}+\frac{1}{\omega_6^2}\right)\omega^2-\omega_2^2\right]+$$

$$+a_2\left[\frac{(1-K_3K_4)}{\omega_3^2\omega_4^2}\cdot\omega^6-\left(\frac{1}{\omega_3^2}+\frac{1}{\omega_4^2}\right)\omega^4-\frac{(1-K_3K_4)\omega_1^2}{\omega_4^2\omega_5^2}\cdot\omega^4+\omega^2+\omega_1^2\left(\frac{1}{\omega_4^2}+\frac{1}{\omega_5^2}\right)\omega^2-\omega_1^2\right]+$$

$$+b_2\left[\frac{(1-K_1K_2)}{\omega_3^2}\cdot\omega^6-\left\{(1-K_1K_2)+\frac{\omega_1^2}{\omega_5^2}+\frac{\omega_2^2\omega_4^2}{\omega_3^2\omega_6^2}\right\}\omega^4+\left\{\omega_1^2+\frac{\omega_2^2\omega_4^2}{\omega_6^2}+\frac{\omega_1^2\omega_2^2\omega_4^2}{\omega_5^2\omega_6^2}\right\}\omega^2-\frac{\omega_1^2\omega_2^2\omega_4^2}{\omega_6^2}\right]+$$

$$+b_3\left[\frac{(1-K_1K_2)}{\omega_4^2}\cdot\omega^6-\left\{(1-K_1K_2)+\frac{\omega_2^2}{\omega_6^2}+\frac{\omega_1^2\omega_3^2}{\omega_4^2\omega_5^2}\right\}\omega^4+\left\{\omega_2^2+\frac{\omega_1^2\omega_3^2}{\omega_5^2}+\frac{\omega_1^2\omega_2^2\omega_3^2}{\omega_5^2\omega_6^2}\right\}\omega^2-\frac{\omega_1^2\omega_2^2\omega_3^2}{\omega_5^2}\right]-$$

$$-K_3b_4\left[\frac{(1-K_1K_2)}{\omega_3^2}\cdot\omega^6-\left(\frac{\omega_1^2}{\omega_5^2}+\frac{\omega_2^2\omega_4^2}{\omega_3^2\omega_6^2}\right)\omega^4+\left\{\frac{\omega_1^2\omega_2^2\omega_4^2}{\omega_5^2\omega_6^2}-\frac{K_1}{K_3}\omega_2^2\frac{T_2}{T_1}\left(1-\frac{\omega_4^2}{\omega_6^2}\right)\right\}\omega^2\right]-$$

$$-K_4b_1\left[\frac{(1-K_1K_2)}{\omega_4^2}\cdot\omega^6-\left(\frac{\omega_2^2}{\omega_6^2}+\frac{\omega_1^2\omega_3^2}{\omega_4^2\omega_5^2}\right)\omega^4+\left\{\frac{\omega_1^2\omega_2^2\omega_3^2}{\omega_5^2\omega_6^2}-\frac{K_2}{K_4}\omega_1^2\frac{T_1}{T_2}\left(1-\frac{\omega_3^2}{\omega_5^2}\right)\right\}\omega^2\right]=0 \quad ......(17)$$

Equation (16) gives the frequencies of oscillation in the general case. The condition of steady oscillation at these frequencies are given by eqn. (17).

4. Characteristics of the Oscillator Neglecting the Effects of the Transmission Lines and Tube Capacitances

The expressions given in the above section are too complicated for any general discussion of the characteristics of the oscillator. In an actual oscillator some simplifications may be made. The tubes that are used for the construction of grounded grid oscillators are usually such that the anode-cathode capacitance C_{ak} is negligible compared with other capacitances. Hence, it may be neglected, and one may assume that $K_3 = K_4 = 0$. This approximation simplifies eqns. (16) and (17) and with the simplified equations a general discussion is possible. However, in this section is considered a simpler form of the oscillator, which illustrates the essential modes of the oscillator. It is assumed that the line inductances L_3 and L_4 are negligible compared with L_1 and L_2 and also that ω_3^2, ω_4^2, ω_5^2, ω_6^2 are much larger than ω_1^2 and ω_2^2.

Equations (16) and (17) may then be simplified to

$$(1-K_1K_2)\omega^4-(\omega_1^2+\omega_2^2)\omega^2+\omega_1^2\omega_2^2=0 \quad ...(18)$$

and

$$a_1(\omega^2-\omega_2^2)+a_2(\omega^2-\omega_1^2)-$$
$$-\omega_2^2b_2(\omega^2-\omega_1^2)\left(1-\frac{\omega_4^2}{\omega_6^2}\right)-$$
$$-\omega_1^2b_3(\omega^2-\omega_2^2)\left(1-\frac{\omega_3^2}{\omega_5^2}\right)+$$
$$+K_1b_4\frac{T_2}{T_1}\omega_2^2\omega^2\left(1-\frac{\omega_4^2}{\omega_6^2}\right)+$$
$$+K_2b_1\frac{T_1}{T_2}\omega_1^2\omega^2\left(1-\frac{\omega_3^2}{\omega_5^2}\right)=0 \quad ......(19)$$

Thus, when the effect of the transmission line and the tube capacitances are negligible the oscillator has two possible frequencies of oscillation given by

$$\omega_{10}^2=[\omega_1^2+\omega_2^2-\{(\omega_1^2-\omega_2^2)^2+4K_1K_2\omega_1^2\omega_2^2\}^{\frac{1}{2}}]/2(1-K_1K_2) \quad ...(20)$$

$$\omega_{20}^2=[\omega_1^2+\omega_2^2+\{(\omega_1^2-\omega_2^2)^2+4K_1K_2\omega_1^2\omega_2^2\}^{\frac{1}{2}}]/2(1-K_1K_2) \quad ...(21)$$

On replacing a_1, a_2, etc., by the circuit parameter values, eqn. (19) reduces to

$$\frac{M_1M_2}{L_1}\cdot\frac{K_1\omega_2^2\omega^2}{(\omega^2-\omega_2^2)}\cdot\frac{1}{r_a}+\frac{M_1M_2}{L_2}\cdot\frac{K_2\omega_1^2\omega^2}{(\omega^2-\omega_1^2)}\left(g_m+\frac{1}{r_a}\right)$$
$$=\frac{1}{R_1(C_1+C_C)}+\alpha\cdot\frac{1}{R_2(C_2+C_C)}+$$
$$+\alpha\cdot\frac{M_2^2}{L_2}\omega_2^2\left(g_m+\frac{1}{r_a}+\frac{1}{r_g}\right)+\frac{M_1^2}{L_1}\omega_1^2\cdot\frac{1}{r_a} \quad ......(22)$$

where $\alpha=\dfrac{\omega^2-\omega_1^2}{\omega^2-\omega_2^2}$, $\omega=\omega_{10}$, or ω_{20}.

In eqn. (22) the cyclotron loss terms appear with a positive sign on the right-hand side. Evidently then, in order that oscillations are made possible it is required that the left-hand side representing the regenerative feedback should be positive and equal or larger in magnitude. Considering this requirement one may arrive at the conclusions discussed below.

It is observed that in order to maintain oscillations, $M_1M_2/(\omega^2-\omega_1^2)$ and $M_1M_2/(\omega^2-\omega_2^2)$ must be positive. This means that if M_1M_2 is positive ω must be larger than ω_2 and ω_1. Hence oscillations will occur only at ω_{20}. On the other hand, if M_1M_2 is negative

ω must be less than ω_2 and ω_1, and oscillations at ω_{10} are only possible. It is also evident from eqn. (22) that M_1 is required to be higher than M_2 for exciting oscillations.

It is thus found that the oscillator, under the simplified conditions considered in this section, may oscillate at one of the two possible frequencies only. The frequency at which it will oscillate is determined by the relative signs of M_1 and M_2. It can be easily shown that V_{C_1} and V_{C_2} differ in phase by 180 deg if the oscillation frequency is ω_{10}. Since this is the desired mode of oscillation for a cyclotron, for a cyclotron oscillator the anode and cathode side should be coupled by opposite senses of mutual inductance.

5. Characteristics of the Oscillator including the Effects of the Transmission Lines

In this section it is assumed that ω_3, ω_4, ω_5 and ω_6, though larger than ω_1 or ω_2, the terms containing them are not totally negligible. As mentioned in the previous section for the usual tubes used in such oscillators, the effect of C_{ak} is negligible and one may assume that $K_3 = K_4 = 0$. In addition the effect of C_{ak} on oscillations at ω_{10} or ω_{20} is negligible. It should be further noted that M_1 and M_2 are small fractions of L_1 and L_2. The frequencies ω_3 and ω_4 would, therefore, be nearly equal to ω_5 and ω_6 respectively. If one assumes that $\omega_3 = \omega_5$ and $\omega_4 = \omega_6$, it can be readily shown from eqn. (16) that the oscillator has four possible frequencies of oscillation, namely ω_{10}, ω_{20} (given by eqns. (20) and (21)), ω_3 and ω_4. In the actual case, ω_3 and ω_4 differ from ω_5 and ω_6 by small quantities. Hence the actual frequencies of oscillation will differ from ω_{10}, ω_{20}, ω_3, and ω_4 by small quantities. These deviations may be obtained by substituting $\omega+\Delta\omega$ for ω, and, considering only the first-order terms, ω has any of the above four values and $\Delta\omega$ is the respective deviation.

On putting

$$\left.\begin{aligned}
\frac{1}{\omega_5^2} &= \frac{1}{\omega_3^2} + \Delta\omega_{35}, & \frac{1}{\omega_6^2} &= \frac{1}{\omega_4^2} + \Delta\omega_{46}\\
\beta_1 &= \frac{\omega_3^2-\omega_{10}^2}{\omega_3^2}, & \beta_2 &= \frac{\omega_3^2-\omega_{20}^2}{\omega_3^2}\\
\gamma_1 &= \frac{\omega_4^2-\omega_{10}^2}{\omega_4^2}, & \gamma_2 &= \frac{\omega_4^2-\omega_{20}^2}{\omega_4^2}\\
\alpha_1 &= \frac{\omega_{10}^2-\omega_1^2}{\omega_{10}^2-\omega_2^2}, & \alpha_2 &= \frac{\omega_{20}^2-\omega_1^2}{\omega_{20}^2-\omega_2^2}\\
\delta_1 &= \frac{\omega_2^2}{\omega_{10}^2-\omega_2^2}, & \delta_2 &= \frac{\omega_2^2}{\omega_{20}^2-\omega_2^2}
\end{aligned}\right\} \quad (23)$$

one obtains

$$\Delta\omega_{10}(\omega_1^2+\alpha_1\omega_2^2) = -\omega_{10}^4\left[\frac{\omega_1^2\Delta\omega_{35}}{\beta_1}+\frac{\omega_2^2\alpha_1\Delta\omega_{46}}{\gamma_1}\right] \quad \ldots\ldots(24)$$

$$\Delta\omega_{20}(\omega_1^2+\alpha_2\omega_2^2) = -\omega_{20}^4\left[\frac{\omega_1^2\Delta\omega_{35}}{\beta_2}+\frac{\omega_2^2\alpha_2\Delta\omega_{46}}{\gamma_2}\right] \quad \ldots\ldots(25)$$

$$\Delta\omega_3 = \frac{\omega_1^2(\omega_3^2-\omega_2^2)}{(1-K_1K_2)\beta_1\beta_2}\cdot\Delta\omega_{35} \quad \ldots\ldots(26)$$

$$\Delta\omega_4 = \frac{\omega_2^2(\omega_4^2-\omega_1^2)}{(1-K_1K_2)\gamma_1\gamma_2}\cdot\Delta\omega_{46} \quad \ldots\ldots(27)$$

Equations (24) to (27), together with eqns. (20) and (21), thus give the possible frequencies of oscillation correct up to the first order.

To examine at which of the above four frequencies oscillations may be maintained one has to examine the conditions of equilibrium as given by eqn. (17) in the four cases. These are considered below:

The condition for ω_{10} is

$$\begin{aligned}
b_1\omega_{10}^2\omega_3^2\frac{\omega_1^2}{\omega_2^2}\cdot\frac{T_1}{T_2}\cdot K_2\delta_1\Delta\omega_{35}
&= b_2\beta_1\left(\omega_2^2\omega_4^2\alpha_1\Delta\omega_{46}-\delta_1\frac{(\omega_{10}^2-\omega_{20}^2)}{\omega_2^2}\cdot\Delta\omega_{10}\right)+\\
&\quad +(a_1+\alpha_1 a_2)\left(1+\frac{\Delta\omega_{10}}{\omega_2^2}\cdot\delta_1\right)\beta_1\gamma_1+\\
&\quad +\omega_{10}^2\delta_1\left(a_1\beta_1\Delta\omega_{46}+a_2\gamma_1\frac{\omega_1^2}{\omega_2^2}\cdot\Delta\omega_{35}\right) \quad (28)
\end{aligned}$$

The substitutions are the same as for eqns. (24) to (27). In writing eqn. (28) b_3 and b_4 have been neglected in comparison to b_1 and b_2, as well as other comparatively small terms which have been neglected. The condition for ω_{20} is obtained from eqn. (28) by replacing ω_{10} by ω_{20} and vice versa. This would replace α_1, β_1, γ_1, δ_1 by α_2, β_2, γ_2, δ_2 only.

It has been assumed that ω_3 and ω_4 are larger than ω_1 and ω_2; α_1, β_1, γ_1 in eqn. (28) are hence always positive; $\Delta\omega_{10}$ and $\Delta\omega_{20}$ are negative; δ_1 is negative and δ_2 is positive. As mentioned in the previous section for a cyclotron oscillator T_1/T_2 is adjusted to be negative. Considering the signs of the different quantities one finds that for ω_{20}, the left-hand term of eqn. (28) is negative while the right-hand term is positive. Hence oscillations at ω_{20} cannot lead to a stable equilibrium value and would not be maintained. On the other hand, since for ω_{10}, δ_1 is negative the left-hand side term of eqn. (28) is positive and the terms on the right-hand side are mostly positive. Hence the equality of eqn. (28) may be satisfied with proper adjustments. The oscillator may then execute stable oscillations at ω_{10}. It should also be noted that the maintenance of this oscillation would be most favoured if ω_1 is equal to ω_2. This is evident on considering the magnitudes of the terms on the two sides of eqn. (28). Deviations from this condition would reduce the value of the left-hand side term of eqn. (28) due to a reduction in the value of ω_1^2/ω_2^2 or δ_1, thereby reducing the amplitude of stable oscillation, and an extreme condition would arise when oscillations cannot be maintained.

In the above discussion we have assumed that T_1/T_2 is negative; if, however, T_1/T_2 is positive one can argue similarly and show that oscillations at ω_{10} cannot be maintained while oscillations at ω_{20} will be maintained.

The conditions of equilibrium for ω_3 and ω_4 are:

$$K_2 b_1 \frac{T_1}{T_2} r_1 \Delta\omega_{35} = \frac{\Delta\omega_3}{\omega_1^2\omega_3^2\omega_4^2}(a_1\eta_2 + a_2\eta_1) - $$

$$-a_2\frac{\Delta\omega_{35}}{\omega_4^2} + \frac{b_3}{\omega_1^2}\left[\xi_1 + \xi_1 r_1 \frac{\Delta\omega_3}{\omega_3^2} - \right.$$

$$\left. -\eta_1 r_1 \omega_2^2 \Delta\omega_{46} + \frac{\omega_1^2\omega_3^2}{\omega_4^2}\left(\frac{1}{\beta_1} + \frac{1}{\beta_2} - 1\right)\eta_2\Delta\omega_{35}\right] \quad \ldots\ldots(29)$$

$$-K_2 b_1 \frac{T_1}{T_2} r_1 \Delta\omega_{35} = \frac{\Delta\omega_4}{\omega_1^2\omega_3^2\omega_4^2}(a_1\eta_4 + a_2\eta_3) - $$

$$-a_1\frac{\omega_2^2}{\omega_1^2\omega_3^2}\,.\,\Delta\omega_{46} + \frac{b_2}{\omega_1^2}\left[\xi_2 - \xi_2 r_1 \frac{\Delta\omega_4}{\omega_3^2} + \right.$$

$$\left. +\eta_4 r_1 \frac{\omega_1^2\omega_4^2}{\omega_3^2}\,.\,\Delta\omega_{35} + \frac{\omega_2^2\omega_4^2}{\omega_3^2}\left(\frac{1}{\gamma_1} + \frac{1}{\gamma_2} - 1\right)\eta_3\Delta\omega_{46}\right] \quad \ldots\ldots(30)$$

where

$$\eta_1 = \frac{(\omega_3^2 - \omega_1^2)}{\omega_3^2}, \qquad \eta_2 = \frac{(\omega_3^2 - \omega_2^2)}{\omega_3^2}$$

$$\eta_3 = \frac{(\omega_4^2 - \omega_1^2)}{\omega_4^2}, \qquad \eta_4 = \frac{(\omega_4^2 - \omega_2^2)}{\omega_4^2}$$

$$\xi_1 = \frac{(\omega_3^2 - \omega_{10}^2)(\omega_3^2 - \omega_{20}^2)}{\omega_3^2\omega_4^2}, \quad \xi_2 = \frac{(\omega_4^2 - \omega_{10}^2)(\omega_4^2 - \omega_{20}^2)}{\omega_3^2\omega_4^2}$$

$$r_1 = \frac{\omega_3^2}{(\omega_3^2 - \omega_4^2)}$$

It may be noted that in the above equations all the η's and ξ's are positive when ω_3, ω_4 are larger than ω_1, ω_2 as assumed earlier. The sign of r_1 is, however, positive when $\omega_3 > \omega_4$ and negative when $\omega_3 < \omega_4$. If one considers eqns. (29) and (30) taking into account the above-mentioned signs of the different quantities, the following conclusions may be drawn.

(*a*) Oscillations at ω_3 may be excited, if T_1/T_2 is positive and r_1 is positive or if T_1/T_2 is negative and r_1 is negative.

(*b*) Oscillations at ω_4 may be excited if T_1/T_2 is positive and r_1 is negative or if T_1/T_2 is negative and r_1 is positive.

The conditions are mutually exclusive; hence, of the two possible frequencies due to the transmission line, only one may be excited in the oscillator. Further, b_2 is larger than b_3; oscillations at ω_3 are hence more easily excitable than that at ω_4. One may, thus, conclude that the favourable conditions for elimination of the oscillations due to transmission lines would be to make r_1 positive since T_1/T_2 is negative for cyclotron oscillators. This requires that the anode line frequency should be higher than that of the cathode line.

In the above discussion the effect of C_{ak}, the anode-cathode capacitance has been neglected. Though, C_{ak} has negligible effect on the conditions of oscillations for the frequencies ω_{10} and ω_{20}, it may have considerable effect on the conditions for ω_3 or ω_4. The effect of C_{ak} on ω_3 or ω_4 may be discussed neglecting the coupling between the two lines through the dee circuit. Under the above conditions eqn. (17) may be simplified to

$$a_3\frac{(\omega_3')^2}{\omega^2}\left\{1 - \frac{(\omega_4')^2}{\omega^2}\right\} + a_4\frac{(\omega_4')^2}{\omega^2}\left\{1 - \frac{(\omega_3')^2}{\omega^2}\right\} + $$

$$+\omega_3^2 b_3\left\{1 - \frac{(\omega_4')^2}{\omega^2}\right\} + \omega_4^2 b_2\left\{1 - \frac{(\omega_3')^2}{\omega^2}\right\} - $$

$$-K_3\omega_4^2 b_4 - K_4\omega_3^2 b_1 = 0 \quad \ldots\ldots(31)$$

where

$$a_3 = \frac{R_3}{L_3}, \qquad a_4 = \frac{R_4}{L_4}$$

$$(\omega_3')^2 = \frac{1}{L_3(C_3 + C_{ak})} \simeq \omega_3^2$$

$$(\omega_4')^2 = \frac{1}{L_4(C_4 + C_{ak})} \simeq \omega_4^2$$

The expression for frequency is given by

$$(1 - K_3K_4)\omega^4 - \{(\omega_3')^2 + (\omega_4')^2\}\omega^2 + (\omega_3')^2(\omega_4')^2 = 0 \quad \ldots\ldots(32)$$

The above equation may also be derived directly from the reduced equivalent circuit of the oscillator shown in Fig. 3. In this equivalent circuit, the effect of the dee circuit is represented by including only the losses presented by them.

On considering eqn. (31) one observes that the oscillator may execute oscillations only at a frequency higher than both ω_3 and ω_4 due to feedback through C_{ak}. Further, this oscillation is favoured if ω_3 is larger than ω_4, since then the effect of cathode damping is minimized.

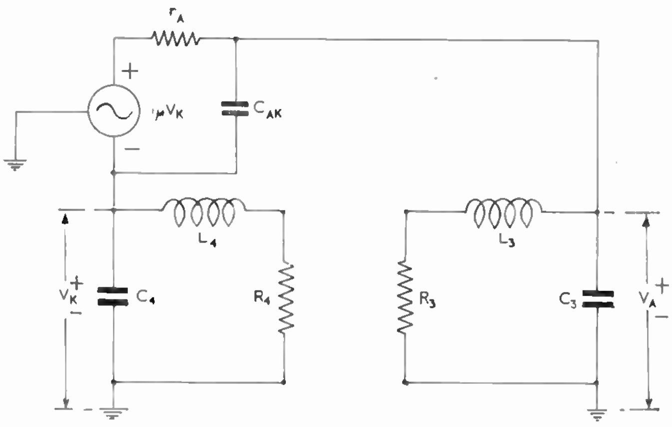


Fig. 3. Simplified equivalent circuit of the oscillator for analysing the effect of C_{ak} on transmission-line-frequency oscillations.

If the above result is considered together with the results presented earlier one may draw the following conclusions regarding the elimination of oscillations at the transmission line frequencies.

(*a*) If T_1/T_2 is positive, oscillations at the line frequencies may be eliminated by making ω_3 lower than ω_4.

(*b*) If T_1/T_2 is negative, oscillations at the line frequencies would be eliminated if ω_3 is less than ω_4 and the feedback through C_{ak} dominates. However, if the feedback through the dee circuits dominates, one has to make ω_3 larger than ω_4.

It should be mentioned in this connection that the comments made above are based on the equivalent circuit in which the tube parameters have been replaced by the steady state values. This linear analysis only predicts if there is any possibility of oscillations at a particular frequency. In an actual oscillator whether a frequency, at which there may be a stable equilibrium amplitude, will be excited or not can be decided by considering the tube non-linearities properly. A non-linear analysis of the present oscillator is rather complicated due to high order of the differential equations describing the oscillator. It was, therefore, decided to study this question by constructing a model oscillator.

6. Experimental Results

The circuit arrangement of the oscillator constructed to verify the theoretical analysis is shown in Fig. 4. The parameter values of the tuned circuits and the coupling inductances were chosen to make the impedance of the tuned circuits at resonance nearly equal to that of the cyclotron existing at the Saha Institute of Nuclear Physics. This required scaling of the inductances by a factor of 10, as the Q of the coils could be made only 1/10th of the cyclotron circuit Q which is approximately equal to 2500. All the capacitances were kept identical to those of the cyclotron system. The resultant resonant frequencies of the tuned circuits of the experimental oscillator was 3·8 Mc/s.

The vacuum tube often used for the construction of a grounded grid oscillator for cyclotron purposes is the type ML-5681 triode. The equivalent characteristics of the tube were realized by using five ECC-84 tubes in parallel. The anode to cathode capacitance, amplification factor and mutual conductance of the tubes when operated in parallel are approximately equal to those of ML-5681. The anode to grid and grid to cathode capacitances were made equal to that of ML-5681 by adding extra capacitances.

The characteristics of the oscillator as obtained experimentally were as follows:

(*a*) When the two tuned circuits were approximately tuned to the same frequency, oscillations at only ω_{20}, shown in Fig. 5(*a*), were obtained for the same signs of M_1 and M_2. For opposite signs of M_1 and M_2 under the above conditions oscillations at ω_{10}, shown in Fig. 5(*b*), only occurred. However, it was observed that ω_{20} could be excited for smaller

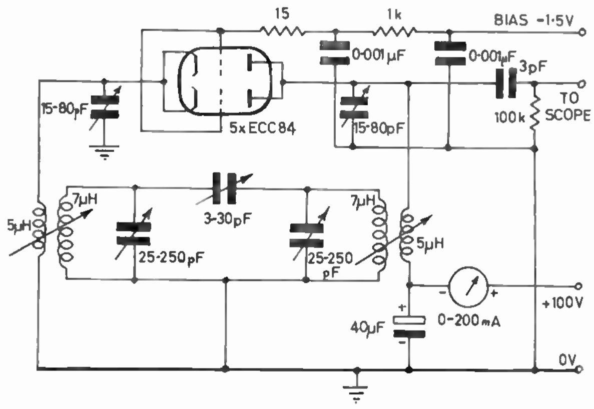
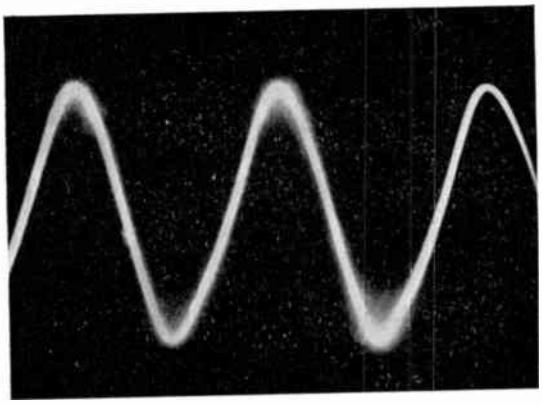
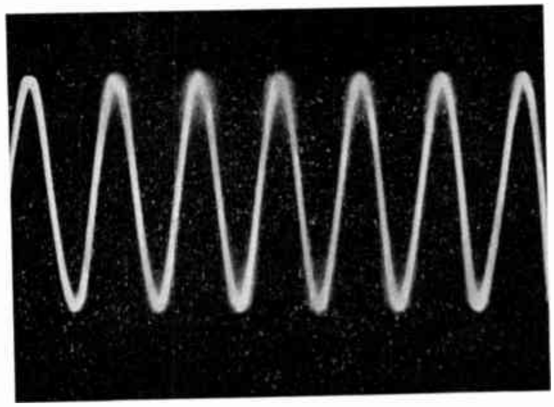


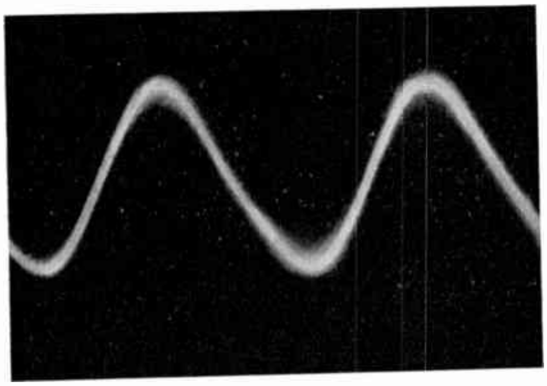
Fig. 4. Circuit arrangement of the experimental oscillator.



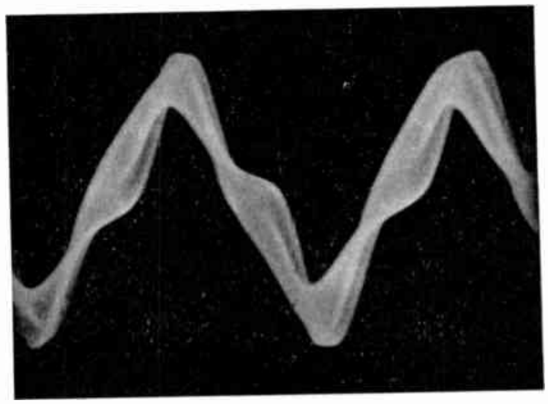
(*a*) Oscillations at ω_{20} ($= 2\pi \times 4{\cdot}1 \times 10^6$ rad/s).



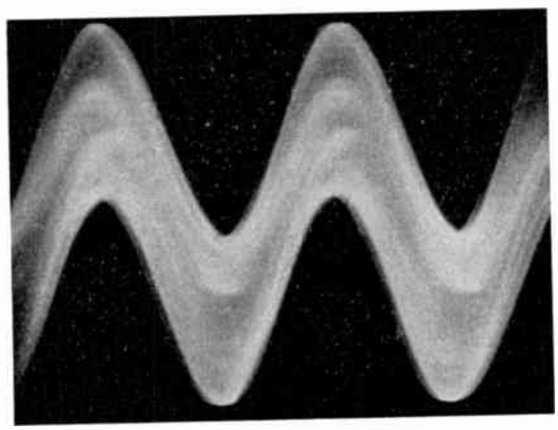
(*c*) Oscillations at ω_3 ($= 2\pi \times 10{\cdot}4 \times 10^6$ rad/s), the anode line frequency.



(*b*) Oscillations at ω_{10} ($= 2\pi \times 3{\cdot}1 \times 10^6$ rad/s).



(*d*) Simultaneous oscillations at ω_{10} and ω_3.



(*e*) Simultaneous oscillations at ω_{20} and ω_3.

Fig. 5. Oscillograms showing the various modes of oscillation. Time scale—1 cm = 0·08 μs.

values of coupling capacitance than were required for exciting ω_{10}. Further, oscillations at ω_{20} were 'soft' excited, whereas oscillations at ω_{10} were 'hard' excited. Once the oscillations at ω_{10} were stopped by circuit adjustments, they did not start when the circuits were put back in the original condition. It could be re-excited only by reducing the anode voltage and slowly increasing it.

(*b*) For a fixed coupling, and fixed frequency of one of the tuned circuits, oscillations at ω_{20} or ω_{10} existed within a range of tuning of the second tuned circuit.

(*c*) Oscillations at ω_3, the anode line frequency shown in Fig. 5(*c*), were excited when the coupling between the two circuits was reduced and ω_3 was larger than ω_4 for both positive and negative signs of T_1/T_2.

(*d*) When the anode line was tuned to a frequency lower than that of the cathode no oscillation at ω_3 or ω_4 were excited.

(*e*) It was also observed that for the tuning conditions as mentioned in (*c*) for some adjustments of the cathode line frequency and coupling capacitor C_C, simultaneous oscillations at both ω_{20} or ω_{10} and ω_3 were obtained as shown in Fig. 5(*d*) and 5(*e*). It should be mentioned that ω_4 and ω_{20} were not harmonically related and the photographs represent the wave shape when synchronized to ω_3.

The observations noted above under (*a*) and (*b*) agree generally with the analytical results. However, the difference in the excitation conditions of ω_{10}, and ω_{20}, namely that ω_{20} is self-excited, whereas ω_{10} is hard-excited, is difficult to explain from theory, since in the analysis the tube parameters were linearized. This peculiarity may be attributed to the non-linear characteristics of the tubes. It is interesting to note that the actual cyclotron oscillator also behaves similarly, i.e. oscillations at ω_{10} are hard-excited.

It is also evident that conditions of excitation of oscillations at the line frequencies also agree with theory. That oscillations at ω_3 are excited for opposite signs of M_1 and M_2 when ω_3 is greater than ω_4 only indicates that feedback through C_{ak} dominates over that through dee-tuned circuits in the experimental oscillator. In an actual cyclotron oscillator, this condition may be reversed since feedback through C_{ak} may be further reduced by proper shielding between anode and cathode.

7. Conclusions

A linear analysis of the lumped equivalent circuit of a grounded grid oscillator for use in cyclotrons has shown that such oscillators have four possible frequencies of oscillation, at only two of which are steady oscillations possible. The conditions determining the excitation of these frequencies have been obtained and verified from studies on an experimental oscillator. The most favoured oscillation frequency is that determined by the dee circuit when the tuning conditions and coupling between the dee circuits are correctly adjusted. However, oscillations at transmission line frequency only or simultaneously with the dee circuit frequency are also possible, for low coupling between the dee circuits. The proper transmission line tuning conditions for the elimination of these undesired oscillations have also been obtained and verified experimentally.

8. Acknowledgments

The authors are indebted to Professor J. N. Bhar for his kind interest and constant encouragement during the progress of the work. One of the authors (D.N.B.) is indebted to Professors B. D. Nag and D. N. Kundu for their interest in the problem. The authors are also grateful to Dr. A. P. Patro and Dr. B. Basu for making available the circuit parameter values of the cyclotron dee system and for general experimental help.

9. References

1. J. Backus, "The design of cyclotron oscillators"; *Rev. Sci. Instrum.*, **22**, pp. 84–92, 1951.
2. K. E. Relf, "The Radio Frequency System of the Cyclotron at Berkeley", Report No. BP42 (3rd Ed.), 1950, Croker Laboratory, University of California, Berkeley, California.
3. F. H. Schmidt, G. W. Farwell, J. E. Henderson, T. J. Morgan and J. F. Streib, "The University of Washington sixty-inch cyclotron", *Rev. Sci. Instrum.*, **25**, pp. 499–510, 1954.
4. K. R. Mackenzie and V. B. Waitham, "R-f system for frequency modulated cyclotron", *Rev. Sci. Instrum.*, **18**, pp. 900–7, 1947.
5. H. Atterling and G. Lindstrom, "The 225-cm cyclotron at the Nobel Institute of Physics, Stockholm", *Ark. for Fys.*, **15**, No. 36, pp. 483–502, 1959.
6. R. S. Livingstone, "The Oak Ridge 86-inch cyclotron", *Nature*, **170**, pp. 221–3, 1952.
7. "The University of Birmingham cyclotron", *Nature*, **169**, p. 476, 1952.
8. F. T. Howard, "Cyclotrons and High Energy Accelerators", Oak Ridge National Laboratory Report No. ORNL-2644 (1958), pp. 15–170.
9. S. Kikuchi *et al.*, "A 160-cm synchro and variable energy ordinary cyclotron", *J. Phys. Soc., Japan*, **15**, pp. 41–59, 1960.
10. F. H. Schmidt and M. J. Jakobson, "Cyclotron oscillators and the shifting inter-dee ground surface", *Rev. Sci. Instrum.*, **25**, pp. 136–9, 1954.

Manuscript received by the Institution on 3rd May 1963. (Paper No. 875.)

U.D.C. 621.375

Design of Transistor I.F. Amplifiers—Stability and Power Gain

By

K. R. WILSHER†

AND

A. DABROWA, B.Sc.‡

Summary: The two most important stability criteria are presented and expressions for power gain are given in terms of them. Expressions for the input admittance of a stage are given, and the beneficial effects of loading the input port are shown. The interaction of stages is discussed with reference to overall stability. The advantages and disadvantages of making all stages of a multistage amplifier identical are given, and compared with a method of designing each stage separately.

List of Symbols

A	g_{in}/g_{11}
b_{in}	Transistor input susceptance (ωC_{in})
b_L	Transistor load susceptance (ωC_L)
B_{22}	Total susceptance at output port
b_{11}, b_{22}	(ωC_{11}, ωC_{22}) susceptance terms in Y_{11}, Y_{22}
$C_{in}, C_L, C_{11}, C_{22}$	See above
d	General two port immittance parameter
F	A function of K and θ
$g_{in}, g_L, g_{11}, g_{22}$	Conductance terms in Y_{in}, Y_L, Y_{11}, Y_{22}
G_{11}, G_{22}	Total conductance at input and output ports
g_e	Shunt conductance added at input port
G_p	Power gain as a ratio
G_{pn}	Normalized power gain
K	Stern's stability factor
K_1, K_p, K_m	Inherent, partial and minimum stability factors
M	$\lvert Y_{21} Y_{12} \rvert$
n	Coupling coil ratio
P_{in}, P_{out}	Input and output powers
Q	Quality factor of the load
W	Bandwidth
$Y_{11}, Y_{22}, Y_{21}, Y_{12}$	Two-port Y parameters
Y_e	Shunt admittance added at input port
Y_{in}, Y_L	Input and load admittances
α'	Common base and common emitter current gains
ϕ	Lathi's stability figure
ϕ_1, ϕ_p, ϕ_m	Inherent, partial and minimum stability figures
ϕ_1'	ϕ_1 modified by g_e
θ	arg Y_{21}, Y_{12}
ω	Angular frequency
ω_m	Angular frequency at maximum output
Re	Real part of . . .
Im	Imaginary part of . . .

1. Introduction

It is usually required to design an amplifier for a particular gain and bandwidth. Present methods of transistor design deal only with the question of gain, bandwidth calculation being too complicated mathematically to be capable of a simple solution. The problem here is the change of transistor parameters over the frequency range of interest, so that only for narrow bandwidths, or more correctly, small fractional bandwidths, can simplifications be made, because in this case bandwidth is determined mainly by circuit components. The common base configuration is to some extent an exception to this rule for when corrected in an h environment the overall parameters can be made sufficiently independent of frequency to allow a simplified calculation of bandwidth even at large fractional bandwidths.

In this paper no attempt is made to calculate the bandwidth of a common emitter amplifier and to allow for the effect of bandwidth on power gain, although these topics are discussed. Results are given so as to provide examples of what can be achieved with available r.f. transistors, and also to verify the theory.

No comparison is made of the three basic configurations with regard to bandwidth, but gain and sensitivity are compared.

2. Design Methods

Obviously all amplifier design methods are governed by the problem of instability. In general, transistors are potentially unstable over a major part of their useful frequency range due to internal feedback. The best way to deal with this problem is to treat the transistor as a black box whose terminal

† E.M.I. Electronics Ltd., Hayes, Middlesex.
‡ Formerly with E.M.I. Electronics Ltd.

parameters only can be measured. A number of workers have used this method to define a criterion of stability, for a transistor fed from a fixed source conductance, connected to a fixed load conductance and not by neutralizing. There are many different design methods based on these stability criteria but in the authors' opinion the most comprehensive and simplest work is "Optimal design of multistage tuned transistor amplifiers" by B. P. Lathi.[1] Unfortunately the report does not deal with bandwidth and care must be exercised by the user if the results are to be satisfactory. In this paper Lathi's method is simplified and the effects of wide bandwidth on design are discussed.

3. The Active Four Pole Network

3.1. *Stability Factor of an Active Four Pole Network*

The input admittance of a four pole network is given by

$$Y_{\text{in}} = Y_{11} - \frac{Y_{12}\,Y_{21}}{Y_{22}+Y_L} \qquad \ldots\ldots(1)$$

If a value of Y_L can be found which makes the real part of Y_{in} negative, instability is possible. Putting a tuned circuit at the input whose losses are smaller than the negative input resistance will produce oscillation. Both the source and the load admittances are important when considering stability. When tuning an amplifier, both the source and load susceptances of any one stage can vary over a wide range and even change sign. As it is impossible to know what the limits of source and load susceptance will be before the amplifier is designed, it must be assumed that both these susceptances can take any value from $+j\infty$ to $-j\infty$. A useful stability criterion must show whether any particular circuit is stable for the worst possible combination of source and load susceptance. (See Fig. 1.)

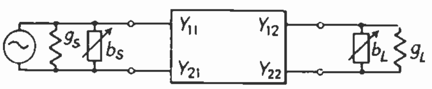


Fig. 1. Active four-pole network with variable source and load admittances.

It has been shown by Linvill[2] and others that if

$$|Y_{21}\,Y_{12}| + R_e(Y_{21}\,Y_{12}) \geqslant 2g_{11}\,g_{22} \qquad \ldots\ldots(2)$$

the transistor may be made to oscillate by suitable tuning at input and output. This immediately gives the stability factor first proposed by Stern[3]:

$$K = \frac{2g_{11}\,g_{22}}{|Y_{21}\,Y_{12}| + R_e(Y_{21}\,Y_{12})} \qquad \ldots\ldots(3a)$$

This factor applies to a transistor whose input and output ports are terminated in pure susceptances. Because the source and load conductances appear in parallel with g_{11} and g_{22} respectively, the stability factor in an actual circuit is modified.

$$K = \frac{2(g_{11}+g_s)(g_{22}+g_L)}{|Y_{21}\,Y_{12}| + R_e(Y_{21}\,Y_{12})} \qquad \ldots\ldots(3b)$$

using Lathi's notation

$$G_{11} = g_s + g_{11} \qquad M = |Y_{21}\,Y_{12}|$$

$$G_{22} = g_L + g_{22} \qquad \theta = \tan^{-1}(Y_{21}\,Y_{12})$$

$$K = \frac{2G_{11}\,G_{22}}{M(1+\cos\theta)} \qquad \ldots\ldots(3c)$$

The circuit is stable if $K > 1$. A number of workers have developed similar criteria, for instance Bahrs' criterion.[4]

Three stability factors may be defined.

(1) *Inherent stability factor*

$$K_i = \frac{2g_{11}\,g_{22}}{M(1+\cos\theta)} \qquad \ldots\ldots(3d)$$

This factor, being dependent only on the transistor parameters, indicates whether the device can become unstable at the frequency of parameter measurement. If $K_1 > 1$ the transistor is said to be unconditionally or inherently stable, as no combination of passive source or load admittances could produce oscillation. Usually $K_1 < 1$ and this is the case given most consideration in this report.

(2) *Partial stability factor*

$$K_p = \frac{2g_{11}\,G_{22}}{M(1+\cos\theta)} \qquad \ldots\ldots(3e)$$

In this case only the effects of g_L are taken into account.

(3) *Actual stability factor*

$$K = \frac{2G_{11}\,G_{22}}{M(1+\cos\theta)} \qquad \ldots\ldots(3c)$$

This gives the actual stability of a stage loaded at each port with passive terminations.

3.2. *Stability Figure of a Four Pole Network*

One disadvantage of the stability factor is that it takes different values for different sets of parameters (Y, h, Z, g). Also, the expression for power gain using this factor K is quite complex.

Lathi has developed a stability *figure* ϕ which is invariant with respect to the parameter set used, and it has the advantage of giving a simple expression for power gain.

It should be noted that K is only invariant at the limit condition $K = 1$.

It is easily shown that the power gain of a four terminal network in terms of Y parameters is:

$$G_p = \frac{P_{out}}{P_{in}} = \frac{|Y_{21}|^2 g_L}{\mathrm{Re}\left(Y_{11} - \dfrac{Y_{12}Y_{21}}{Y_{22}+Y_L}\right)|Y_{22}+Y_L|^2} \qquad \text{......(4)}$$

The power gain is normalized by putting

$$G_p = \left|\frac{Y_{21}}{Y_{12}}\right| G_{pn} \qquad \text{......(5)}$$

G_{pn} is called the normalized power gain. As G_p would have the same value whatever parameter set was used to calculate it, or in other words G_p is matrix invariant, and also

$$\left|\frac{Y_{21}}{Y_{12}}\right| = \left|\frac{Z_{21}}{Z_{12}}\right| = \left|\frac{h_{21}}{h_{12}}\right| = \left|\frac{g_{21}}{g_{12}}\right| \qquad \text{......(6)}$$

then G_{pn} is also matrix invariant.

The imaginary part of Y_L, b_L must be chosen to give the maximum value of G_p (tuning the collector load). This value is

$$b_L = \frac{M \sin\theta}{2g_{11}} - b_{22} \qquad \text{......(7)}$$

Then putting $$\phi_1 = \frac{2g_{11}g_{22}}{M} - \cos\theta \qquad \text{......(8a)}$$

(Inherent stability figure)

and $$\phi_p = \frac{2g_{11}G_{22}}{M} - \cos\theta \qquad \text{......(8b)}$$

(Partial stability figure)

$$G_{pn} = \frac{2(\phi_p - \phi_1)}{\phi_p^2 - 1} \qquad \text{......(9)}$$

Since G_{pn} is matrix invariant so are ϕ_1 and ϕ_p. Thus

Actual stability figure

$$\phi = \frac{2G_{11}G_{22}}{M} - \cos\theta \qquad \text{......(8c)}$$

The condition for stability is

$$\phi > 1 \qquad \text{......(10)}$$

The expressions for ϕ and K are very similar and there is a simple relationship between them.

$$\phi = K(1 + \cos\theta) - \cos\theta \qquad \text{......(11)}$$

This is plotted in Fig. 2.

Normalized power gain in terms of K is given below.

$$G_{pn} = \frac{2(K_p - K_1)}{K_p^2(1+\cos\theta)^2 - 2K_p\cos\theta(1+\cos\theta) - \sin^2\theta} \qquad \text{......(12)}$$

The stability figure ϕ will be used wherever possible.

The normalized power gain G_{pn} has a maximum value for inherently stable devices ($\phi > 1$), at some optimum value of ϕ_p. When the device is inherently unstable G_{pn} increases continuously as ϕ_p is reduced.

Venkateswaran[5] has also derived an invariant stability factor, S, which is similar, though not identical to G_{pn}.

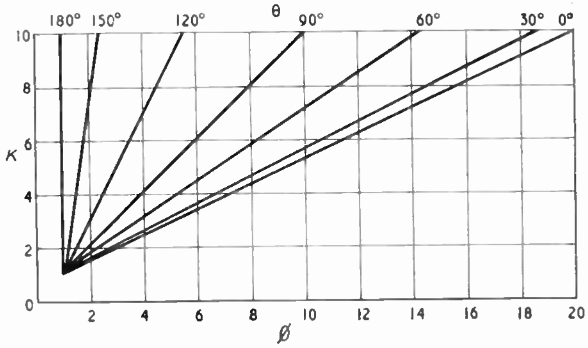


Fig. 2. $\phi = K(1 + \cos\theta) - \cos\theta$ for different values of θ. (After Lathi.[1])

3.3. *Power Gain of a Four Pole Active Network*

From four-pole theory

$$G_p = \frac{P_{out}}{P_{in}} = \frac{|Y_{21}|^2 g}{\mathrm{Re}\left(Y_{11} - \dfrac{Y_{12}Y_{21}}{Y_{22}+Y_L}\right)|Y_{22}+Y_L|^2} \qquad \text{......(4)}$$

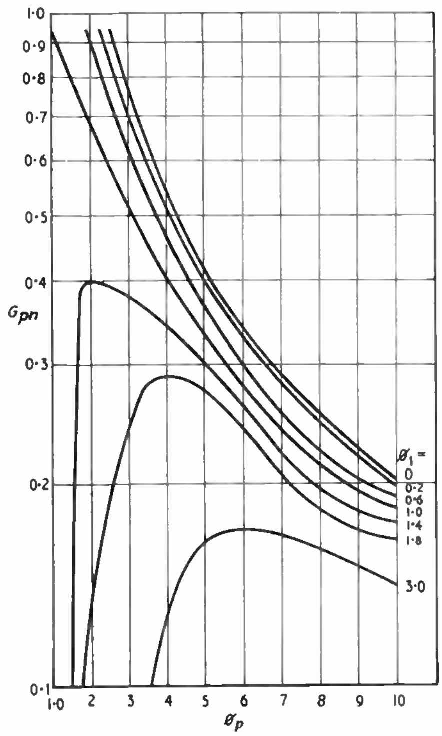


Fig. 3. $G_{pn} = \dfrac{2(\phi_p - \phi_1)}{\phi_p^2 - 1}$ for different values of ϕ_1. (After Lathi.[1])

As described previously this can be represented by

$$G_p = \left|\frac{Y_{21}}{Y_{12}}\right| G_{pn} \qquad \ldots\ldots(5)$$

where $$G_{pn} = \frac{2(\phi_p - \phi_1)}{\phi_p^2 - 1}$$

Figure 3 shows G_{pn} plotted against ϕ_p for various values of ϕ_1. From this graph it can be seen that when $\phi_1 > 1$, there is a maximum value for G_{pn} which is obtained by putting $\phi_p = \phi_{p(\mathrm{opt})}$

where $$\phi_{p(\mathrm{opt})} = \phi_1 + \sqrt{\phi_1^2 - 1} \qquad \ldots\ldots(13)$$

If $\phi_p > 1$, G_{pn} is defined completely by ϕ_p and ϕ_1. If $\phi_1 \ll 1$, ϕ_1 can be neglected in the equation for G_{pn}. The rather weak dependence of G_{pn} on ϕ_1 when $\phi_1 \ll 1$ allows the value of ϕ to be altered without much effect on the power gain. This can be done by shunting the input port with conductance.

3.4. *Conductance Loading of the Input Port*

In the grounded emitter configuration the input conductance of any transistor type is not particularly well defined, and although there is some improvement at high frequencies, there may be a three to one variation between units. Also, this parameter varies with frequency because of its dependence on α' and this may be important when the bandwidth of an amplifier is an appreciable fraction of its centre frequency. As the load conductance of an iterative stage is the input conductance of the next stage it is desirable to add conductance at the input port to reduce variations in power gain, bandwidth, centre frequency and stability factor. (See Fig. 4.)

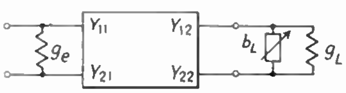


Fig. 4. Active four-pole network with conductance added to input port.

If a conductance g_e is added at the input the real part of the input admittance becomes $g_{11}+g_e$; all other terms in the matrix are not affected.

ϕ_1 now becomes ϕ_1'

where $$\phi_1' = \frac{2(g_e+g_{11})g_{22}}{M} - \cos\theta \qquad \ldots\ldots(8d)$$

and $$G_{pn} = \frac{2(\phi_p - \phi_1')}{\phi_p^2 - 1} \qquad \ldots\ldots(9a)$$

For g_e to have negligible effect on G_{pn} the term in ϕ_1' dependent on g_e, $\left(\frac{2(g_e+g_{11})g_{22}}{M}\right)$ must be small compared with ϕ_p. This condition is of greater necessity than $\phi_1 \ll \phi_p$ because the term $(-\cos\phi)$ can have a maximum value of unity making $\phi_1 \simeq 1$ and so, perhaps, not $\ll \phi_p$.

So the stage gain of a potentially unstable device can be stabilized against g_{11} changes with very small loss in power gain. Conductance loading will also give flexibility in the design of the interstage transformer.

In order to keep the value of ϕ_p constant as g_e is increased, g_L has to be reduced and this increases the sensitivity of G_{pn} to variations in g_{22}. This may be troublesome, but in most designs even with $g_e \gg g_{11}$, $g_L \gg g_{22}$.

4. Multistage Amplifiers

4.1. *The Interdependence of Stages in Multistage Amplifiers*

Assume that all stages are identical—this is true of most amplifiers. The power gain of a stage terminated by a passive load is determined by its inherent stability figure, ϕ_1 and partial stability figure, ϕ_p. In the case of cascaded amplifiers where source and load admittances are active networks which may present negative values of conductance under certain conditions of tuning, the actual stability figure ϕ, may be less than the partial stability factor ϕ_p. The value of ϕ_p must be made large enough so that ϕ is always greater than 1. It is necessary to know what the minimum value of ϕ_p can be.

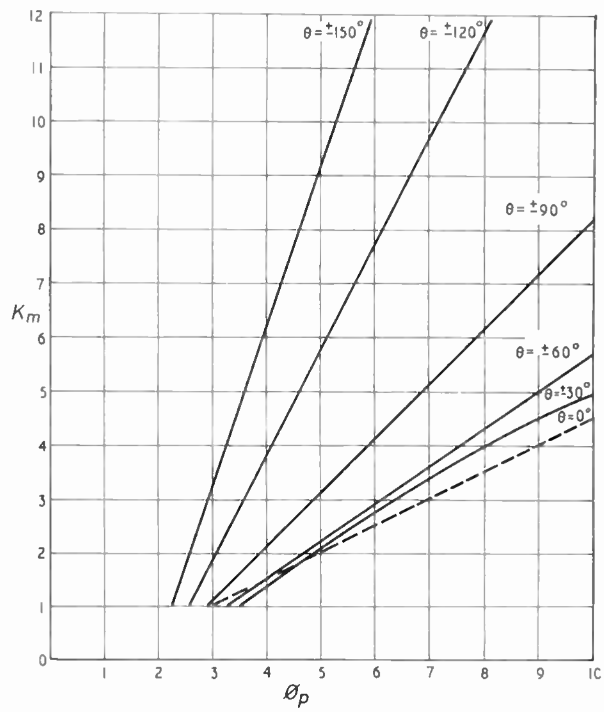


Fig. 5. K_m against θ and ϕ_p. (After Lathi.[1])

During alignment the source and load conductances will vary for every stage. It is possible to find a condition where, for one particular stage, the source and load conductances are simultaneously at a minimum. This stage will then have the minimum stability factor possible, expressed as K_m. The derivation of K_m is given by B. P. Lathi.[1] Some approximations were made by him to arrive at the equation:

$$\frac{(K_m+1)^2}{K_m} = \frac{\phi_p + \cos\theta}{1+\cos\theta}\,\frac{\phi_p^2+2\phi_p\cos\theta + 1}{\phi_p^2-1} \qquad \ldots\ldots(14)$$

K_m is plotted against ϕ_p and θ in Fig. 5.

It may be seen that, for a particular transistor, fixing K_m fixes ϕ_p and so also the power gain.

The approximations made in deriving eqn. (14) were such that it always gives a value slightly lower than the actual value of K_m. A safe method is to use a value of ϕ_p obtained from the graph.

For $\phi_p > 7$, K_m is always greater than three for any value of ϕ; a value of K_m less than two should not be used because errors in parameter measurements and stray feedback could reduce K_m to near or below unity.

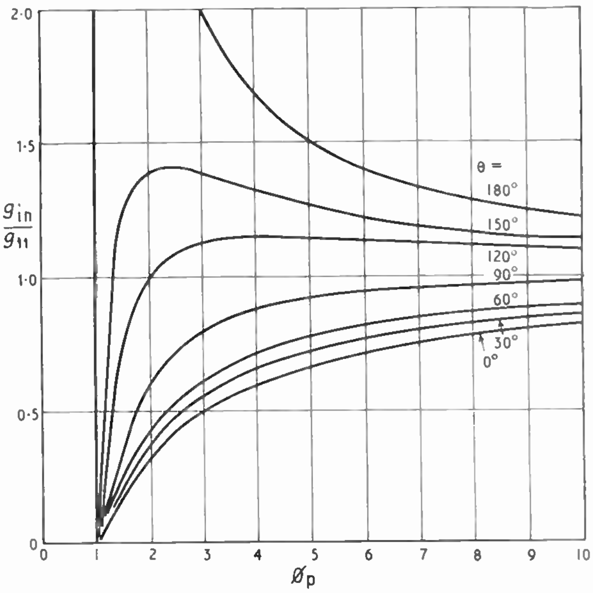


Fig. 7. g_{in}/g_{11} against ϕ_p and θ for identical stages. (After Lathi.[1])

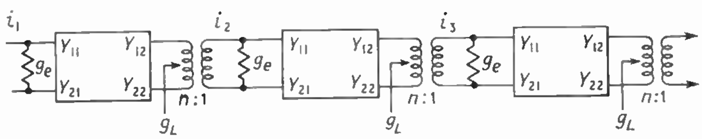


Fig. 6. Chain of identical stages.

4.2. *The Coupling of Stages* (Fig. 6)

Re-arranging eqn. (8b)

$$g_L = \frac{M}{2g_{11}}(\phi_p + \cos\theta) - g_{22} \qquad \ldots\ldots(15)$$

Also when the stage is tuned for maximum power gain

$$b_L = \frac{M\sin\theta}{2g_{11}} - b_{22} \qquad \ldots\ldots(7)$$

and

$$\frac{g_{in}}{g_{11}} = \frac{\phi_p^2-1}{\phi_p^2+2\phi_p\cos\theta+1} \qquad \ldots\ldots(16)$$

$$\frac{b_{in}-b_{11}}{g_{11}} = \frac{-2\phi_p\sin\theta}{\phi_p^2+2\phi_p\cos\theta+1} \qquad \ldots\ldots(17)$$

$\frac{g_{in}}{g_{11}}$ and $\frac{b_{in}-b_{11}}{g_{11}}$ are plotted in Figs. 7 and 8.

The coupling network must present a conductance g_L and susceptance b_L at its input when loaded with g_{in} and b_{in} at its input port. Thus, for an ideal transformer the turns ratio is given by

$$n = \sqrt{\frac{g_{in}}{g_L}} \qquad \ldots\ldots(18)$$

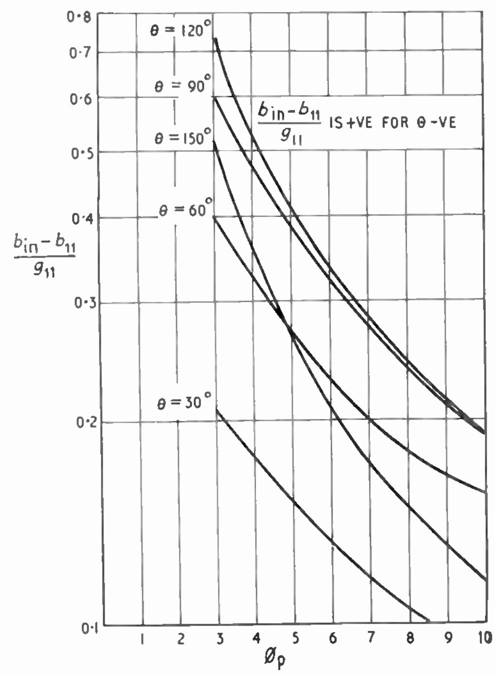


Fig. 8. $\frac{b_{in}-b_{11}}{g_{11}}$ against ϕ_p for various θ. (After Lathi.[1])

and the primary inductance by

$$\left(b_L - \frac{b_{\text{in}}}{n^2}\right) \qquad \text{......(18a)}$$

In all these equations g_{11} includes any input shunt g_e.

4.3. *Design Procedures for Multistage Amplifiers using Inherently Unstable Devices*

From the specification of a required multistage amplifier it may be deduced that one or several of the variables listed below must be fixed at the start of the design.

(1) Stage gain

(2) Stability figure

(3) Load impedance

(4) Coil ratio

In general, only two of these can be chosen, e.g. stability figure and coil ratio, or stage gain and coil ratio, the others then being fixed.

From the equations and graphs already given it is possible to derive design procedures for perhaps almost any combination.

It is sometimes required to obtain maximum power gain from a stage but, as has been shown, the maximum gain is only limited by constraint upon the stability figure, if the transistor is inherently unstable. Stability is not the only limiting factor. In wide-band amplifiers gross distortion of the response can occur even at reasonably high stability figures. (These effects are dealt with elsewhere.) The value of ϕ_p is therefore important in high-gain narrow-band amplifiers and wide-band low-gain stages, and is therefore the most reasonable starting point for design. Sometimes the stage gain must be fixed, for instance, in a logarithmic amplifier. Design procedures starting with ϕ_p and n, G_p and n will now be derived.

4.4. *Derivation of Procedure given K_m, or ϕ_p and n*

Using the relationships discussed previously, a reasonably simple design procedure can be evolved if the starting point to the design is choice of K_m or ϕ_p.

Power gain for a stage is given by

$$\text{power gain} = \left|\frac{Y_{21}}{Y_{12}}\right| G_{pn} \qquad \text{......(5)}$$

where

$$G_{pn} = \frac{2(\phi_p - \phi_1)}{(\phi_p^2 - 1)} \qquad \text{......(9)}$$

The inherent stability factor ϕ_1 can be changed by shunting the input with a conductance g_e, and ϕ_p can be altered by altering g_L. The power gain can be varied over a wide range by choice of these conductances. To couple the stages together it is sufficient to use a transformer of such a ratio that a secondary load g_{in} gives a primary conductance of g_L. The design problem is then to find a value of g_e and coil ratio n to give the required stability figure.

n is chosen to be some convenient value (from 2 to 5).

$$g_L = \frac{M(\phi_p + \cos\theta)}{2(g_{11} + g_e)} - g_{22} \qquad \text{......(15)}$$

For identical stages:

From eqn. (18)

$$g_{\text{in}} = n^2 g_L = \frac{g_{\text{in}}}{g_{11} + g_e}(g_{11} + g_e) \qquad \text{......(19)}$$

also

$$\frac{g_{\text{in}}}{g_{11} + g_e} = \frac{\phi_p^2 - 1}{\phi_p^2 + 2\phi_p \cos\theta + 1} = C, \text{ say} \qquad \text{......(16)}$$

when input and output susceptances are adjusted for maximum power gain. This function C is plotted in Fig. 7 for ϕ_p and θ.

From eqn. (18) and eqn. (16):

$$g_{11} + g_e = \frac{n^2 g_L}{C} \qquad \text{......(20)}$$

From eqn. (15) and eqn. (20):

$$g_L = \frac{CM(\phi_p + \cos\theta)}{2^2_n g_L} - g_{22} \qquad \text{......(21a)}$$

$$0 = g_L^2 + g_L g_{22} - \frac{CM(\phi_p + \cos\theta)}{2^2_n} \qquad \text{......(21b)}$$

$$g_L = -\tfrac{1}{2}g_{22} + \tfrac{1}{2}\sqrt{g_{22}^2 + 2CM(\phi_p + \cos\theta)/n^2} \qquad \text{......(21c)}$$

If ϕ_p and n have been fixed, then all the terms on the right-hand side are known. Having found g_L, g_e can be calculated by re-arranging eqn. (20):

$$g_e = \frac{n^2 g_L}{C} = g_{11} \qquad \text{......(22)}$$

To find the power gain it is necessary to calculate ϕ_1'.

$$\phi_1' = \frac{2(g_{11} + g_e)(g_{22})}{M} = \cos\theta \qquad \text{......(8d)}$$

$$\text{Power gain } G_p = \left|\frac{Y_{21}}{Y_{12}}\right| \frac{2(\phi_p - \phi_1)}{\phi_p^2 - 1} \qquad \text{......(12)}$$

(from eqns. (5) and (9)).

The total circuit capacitance referred to the collector is composed of collector capacitance b_{22} and reflected input capacitance C_{in}/n^2, and stray capacitance.

The input capacitance can be calculated from eqn. (17).

$$\frac{b_{\text{in}} - b_{11}}{(g_{11} + g_e)} = \frac{-2\phi_p \sin\theta}{\phi_p^2 + 2\phi_p \cos\theta + 1} \qquad \text{......(17)}$$

$\dfrac{b_{\text{in}} - b_{11}}{(g_{11} + g_e)}$ is plotted against ϕ_p for various θ's in Fig. 8.

Maximum gain will not occur when the total load susceptance is zero (resonant frequency of the load). Maximum gain occurs when

$$B_{22} = \frac{M \sin \theta}{2(g_{11}+g_e)} \qquad \ldots\ldots(15a)$$

where B_{22} is the total load susceptance. This effect is small.

4.5. *Summary of Design Procedure for Unstable Devices*

(*a*) From the transistor Y parameters find M, θ, $\sin \theta$ and $\cos \theta$. Calculate also $\left|\frac{Y_{21}}{Y_{12}}\right|$

(*b*) Choose K_m and from Fig. 5 find ϕ_p, or choose ϕ_p directly.

(*c*) Choose a suitable value for n (from 2 to 5).

(*d*) From Fig. 7 find C.

(*e*) From eqn. (21) calculate g_L.

(*f*) From eqn. (22) calculate g_e.

(*g*) From eqn. (8d) calculate ϕ_1' and find G_p from eqn. (9a).

(*h*) Find the input capacitance from Fig. 8 and then find the total load capacitance.

If step (*f*) gives a negative value for g_e this means that the values of ϕ_p and n are not compatible. Either ϕ_p or n must be increased. As increasing ϕ_p would mean reducing the gain it is usually better to start again using a higher value of n.

5. Designing for a Particular Gain/Stage

For some applications it is necessary to fix the gain/stage of a multistage amplifier, i.e. G_{pn} must be fixed at the start of the design. When this is done and the coil ratio also fixed there does not seem to be a simple method of calculating g_e. The method described previously could be used by choosing values of ϕ_p and calculating values of G_{pn} until the desired G_{pn} was found, but this is tedious. A method can be found which also relies on successive approximation but is much faster.

The problem is to find the value of g_e.

From eqn. (16)

$$\phi_p^2+2\phi \cos \theta + 1 = \frac{\phi_p^2-1}{g_{\text{in}}}(g_{11}+g_e) \qquad \ldots\ldots(16a)$$

From eqn. (9)

$$\phi_p^2-1 = \frac{2(\phi_p-\phi_1')}{G_{pn}} \qquad \ldots\ldots(23)$$

Therefore

$$\phi_p^2+2\phi_p \cos \theta + 1 = \frac{2(\phi_p-\phi_1')(g_{11}+g_e)}{G_{pn}g_{\text{in}}} \qquad \ldots\ldots(24)$$

For identical cascaded stages

$$g_{\text{in}} = n^2 g_L \qquad \ldots\ldots(18)$$

Also from eqn. (8b) and eqn. (12)

$$2(\phi_p-\phi_1') = 2\left[\frac{2(g_{11}+g_e)(g_L+g_{22})}{M} - \frac{2(g_{11}+g_e)g_{22}}{M}\right] \qquad \ldots\ldots(25)$$

This reduces to

$$2(\phi_p-\phi_1') = \frac{4(g_{11}+g_e)g_L}{M} \qquad \ldots\ldots(25a)$$

Substituting eqn. (18) and eqn. (25a) in eqn. (24)

$$\phi_p^2+2\phi_p \cos \theta + 1 = \frac{4(g_{11}+g_e)^2}{M_n^2 G_{pn}} \qquad \ldots\ldots(26)$$

Now from eqn. (8b)

$$\phi_p + \cos \theta = \frac{2(g_{11}+g_e)(g_L+g_{22})}{M} \qquad \ldots\ldots(27)$$

and

$$\phi_p^2-1 = \frac{2(\phi-\phi_1')}{G_{pn}} \qquad \ldots\ldots(23)$$

Substituting eqn. (25a) in eqn. (23)

$$\phi_p^2-1 = \frac{4(g_{11}+g_e)g_L}{MG_{pn}} \qquad \ldots\ldots(23a)$$

Dividing eqn. (23a) into eqn. (27)

$$\frac{\phi_p + \cos \theta}{\phi_p^2-1} = \frac{G_{pn}}{2}\left(1+\frac{g_{22}}{g_L}\right) \qquad \ldots\ldots(28)$$

Consider the equations

$$\phi_p^2+2\phi_p \cos \theta + 1 = \frac{4(g_{11}+g_e)^2}{M_n^2 G_{pn}} \qquad \ldots\ldots(26)$$

and

$$\frac{\phi_p + \cos \theta}{\phi_p^2-1} = \frac{G_{pn}}{2}\left(1+\frac{g_{22}}{g_L}\right) \qquad \ldots\ldots(28)$$

M and θ are known from the Y parameters, and n is chosen to be some convenient value. G_{pn} is found from the Y parameters and the required gain/stage.

Inserting any value of ϕ_p into eqn. (26) will allow $(g_{11}+g_e)$ to be calculated. The value of C can be found from Fig. 7 and then g_L calculated from eqn. (20). Inserting this value of g_L into eqn. (28) will now yield another value of ϕ_p. If this new value does not agree with the original value, the new value can be used to repeat the whole process. Convergence will be more rapid if a value even more remote from the original is chosen.

The whole process is much simplified by plotting graphs of eqn. (26) and eqn. (28). These are Figs. 9 and 10.

Once ϕ_p is fixed g_e can be calculated easily from

$$\frac{4(g_{11}+g_e)^2}{M_n^2 G_{pn}} \qquad \ldots\ldots(26)$$

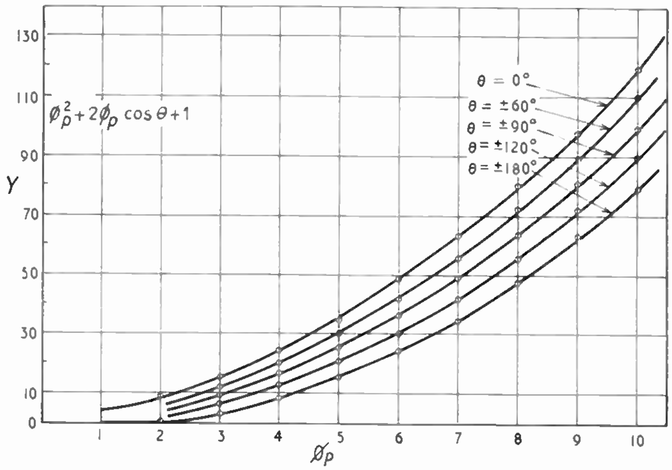


Fig. 9. Graph of equation (26).

The procedure may be summarized as follows:

(*a*) Find M, θ and G_{pn} and choose n.

(*b*) Choose ϕ_p.

(*c*) Find C from Fig. 7.

(*d*) Find $\dfrac{4(g_{11}+g_e)^2}{M_n^2 G_{pn}}$ from Fig. 9.

(*e*) Calculate $(g_{11}+g_e)$ and find g_L from

$$g_{11}+g_e = \frac{n^2 g_L}{C}$$

(*f*) Calculate the value of $\dfrac{G_{pn}}{2}\left(1+\dfrac{g_{22}}{g_L}\right)$

(*g*) Find ϕ_p from Fig. 10.

(*h*) Compare the two values ϕ_p. If the second value is greater than the first, start again with a value even larger. If the second value is smaller than the first start again with a value even smaller.

The whole process of calculating for a particular gain/stage is simple if $g_L \gg g_{22}$ as then g_L has no effect in eqn. (28).

Then
$$\frac{\phi_p+\cos\theta}{\phi_p^2-1} = \frac{G_{pn}}{2} \qquad \ldots\ldots(29)$$

This gives the value of ϕ_p directly. From Fig. 9 the value of $\dfrac{4(g_{11}+g_e)^2}{M_n^2 G_{pn}}$ can be found and from this the value of g_e is easily calculated. g_L should then be calculated to check that $g_L \gg g_{22}$.

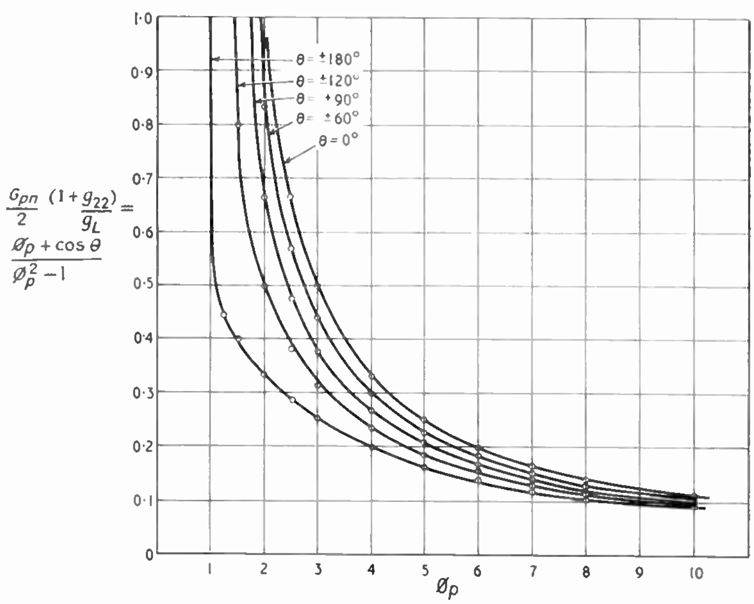


Fig. 10. Graph of equation (28).

6. Design of a Multistage Amplifier

It is now convenient to point out the difference between these methods and those of Lathi.

A multistage amplifier can be constructed by cascading identical stages designed by either of the methods shown. If the source impedance is a well-defined positive conductance the first stage of such an amplifier will have a higher value of K_m than the others. If it is required to obtain the maximum possible gain for the value of K_m chosen this would mean inefficient use of the first stage. In fact this is not a requirement often encountered, but it is the basis of Lathi's method.

6.1. *Design of the First Stage*

A chain of stages designed by the methods just given could be used as a multistage amplifier without modification by matching the source conductance to g_{in} (see Fig. 11). The power gain will then be the sum of the stage power gains in terms of decibels.

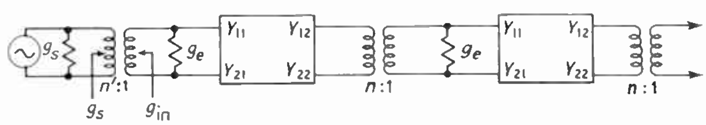
This system has two disadvantages.

(*a*) The presence of g_e will increase the noise figure.

(*b*) Maximum gain is not obtained for the particular value of K_m used.

The best way to design the first stage is by employing the idea of transducer gain. The transducer gain of an amplifier is defined as the increase in power delivered to a load when an amplifier is inserted into the line feeding the load, if the source was first matched to the load. (See Fig. 12.)

Fig. 11.
Connection of source to an identical stage amplifier.



To obtain the maximum gain for a given value of K_m, K_m must be the same for every stage, and the source and load conductances of the first stage must be optimized. To do this the first stage must be designed first and a value for the first load conductance found. Using this value the second stage is designed and the second load conductance found, and so on. For a multistage amplifier this must be time consuming. The resulting amplifier will have a different value of g_e at every stage.

When the identical stage amplifier and the optimized constant K_m amplifier are compared it can be seen that a simple modification to the first stage of the former can reduce the gain difference to one or two decibels. In view of this it seems hardly worthwhile to design each stage separately.

The design of the first stage will now be dealt with in detail.

$$\text{transducer gain} = \frac{P_{out}}{\text{power available from source}}$$

Many workers have analysed the transducer gain stage, or single stage amplifier (e.g. Bahrs[4]). The main problem is that the analysis involves cubic equations when the device is potentially unstable. Bahrs solved these using a digital computer and produced useful design graphs. Lathi extended this approach and has set out a simple method of design:

$$G_T = \text{transducer gain}$$

$$= \frac{4|Y_{21}|^2 g_s g_L}{|(G_{11}+jB_{11})(G_{22}+jB_{22}) - M(\cos\theta\, g \sin\theta)|^2} \qquad \ldots\ldots(30)$$

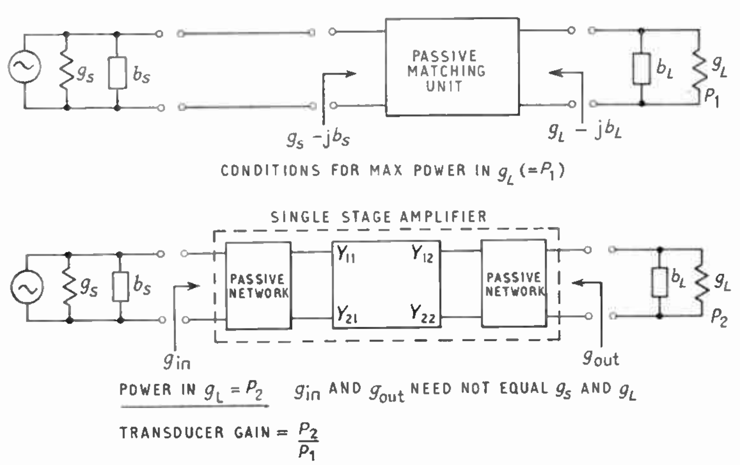


Fig. 12. Illustrating the definition of transducer gain.

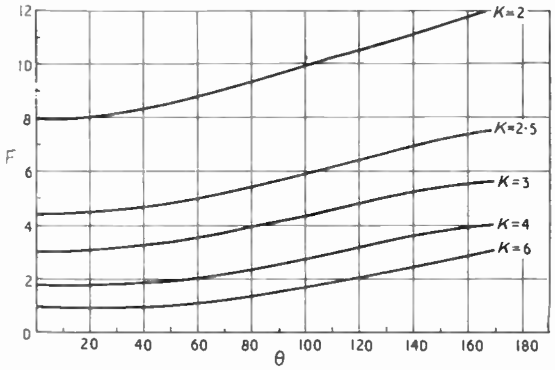


Fig. 13. F against θ for various K. (After Lathi.[1])

Putting
$$G_T = G_{TN}\left|\frac{Y_{21}}{Y_{12}}\right| \qquad \ldots\ldots(31)$$

It was shown by Lathi that when the input and output susceptances were adjusted for maximum output.

$$G_{TN} = \frac{2g_s g_L F}{MK(1+\cos\theta)} \qquad \ldots\ldots(32)$$

where F is a very complicated function of K and θ. He plotted a graph of F against K for various θ values. This is reproduced in Fig. 13. If the transducer gain stage is to be the first stage of a multistage amplifier it is desirable to make its load admittance equal to the load admittance of all the other stages. Since

$$K = \frac{2(g_s+g_{11})(g_L+g_{22})}{M(1+\cos\theta)} \qquad \ldots\ldots(3c)$$

fixing K and g_L also fixes g_{TN}. G can then be calculated easily from eqn. (32). The K_m of the first stage is very nearly equal to K, because its source conductance is fixed, and the variation of load conductance will probably be small. (See Fig. 14.)

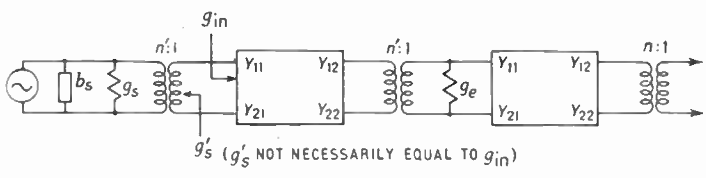


Fig. 14. First stage modification.

For any particular K and θ there are optimum values for g_s and g_L which maximize G_{TN}. These values are usually impractical for the first stage of a multistage amplifier, though perhaps acceptable for single stage amplifiers, the load conductance being low and the source conductance high. Lathi then applies a correction factor to the source and load conductances so that the gain is kept as near optimum as possible for non-optimum source and loads. However, most design requirements can be met by following one of the simpler design procedures described below.

6.2. *Factors Affecting the Design of the First Stage*

Any one or all of the following requirements may be important:

(1) The amplifier must have the lowest possible noise figure.

(2) The input impedance must be well defined.

(3) The amplifier must be stable over a wide range of source impedances.

(4) The amplifier must have the maximum gain possible for the particular K_m used.

It can be seen immediately that some of these requirements conflict. Consider (1). To obtain the lowest noise figure the first stage must be run at its optimum direct current and voltage from the optimum source conductance. Stability can only be obtained by choosing g_L, no g_e being allowed as this shunts the input. If the load conductance is made equal to the load conductance used in the later stages and the stability factor K calculated, in most cases K will be high enough. If it is not then the only alternative is to increase g_L.

Consider (2). The input conductance variations at high stability factors will be due only to the variation of g_{11} between units. If this variation is too great some conductance g_e must be added at the input. When the stability factor is low the input conductance will depend on the other Y parameters (eqn. (1)) and variations in these will affect the value of g_{in}.

Consider (3). The widest variations of source conductance possible are obtained by shorting or by making the input an open circuit.† When the input is short circuited K will be infinite, but when the input is an open circuit K could be less than 1, indicating the possibility of oscillation. This applies particularly to case (1). If this danger is to be avoided K_p or ϕ_p must be greater than 1 for the first stage.

Consider (4). This is Lathi's requirement. In most cases making g_L equal to the other g_L's in the amplifier, and calculating g_s to give the same value of K_m used in the design of the iterative stage will give a first-stage gain only one or two decibels below the optimum. If the absolute maximum gain is needed, the design of each stage must be carried out separately, as described by Lathi.

† If the source is another amplifier the source impedance could be negative.

If an input shunt conductance is used the transducer gain must of course be calculated using $g_{11}+g_e$ and not merely g_{11} in the expression for K, etc.

6.3. *Frequency Response Distortion*

In the two design methods given, ϕ_p and then G_{pn} were chosen at the start. It is important to know what range of values will avoid (*a*) instability, (*b*) distortion of the frequency response. The first point has been dealt with comprehensively. The second point cannot be treated in such a precise manner. Consider the causes of frequency response distortion.

The reasons for the response of a multi-stage amplifier utilizing single tuned circuit coupling differing from the response of a single tuned circuit can be grouped under two headings:

(*a*) Feedback effects due to Y_{12}.

(*b*) Variation of transistor parameters over the passband.

(*a*) To study the feedback effect consider a hypothetical multistage amplifier using devices whose parameters (g_{11}, g_{22}, C_{11}, C_{22}, Y_{21}, Y_{12}) are frequency invariant, tuned only with the input and output capacitance of each device (no capacitance added to the coupling circuit). If the phase of the feedback is such as to make $b_{in} > b_{11}$, as it is in most cases, and the stability figure is low, a response similar to that shown

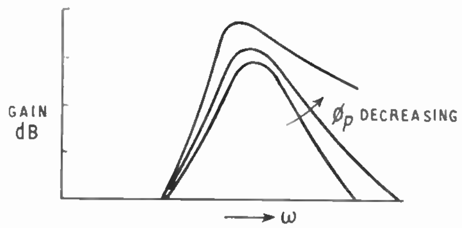


Fig. 15. Frequency response of feedback amplifier stage.

in Fig. 15 will be obtained. An input signal of frequency lower than the amplifier centre frequency will give a small output. As the frequency is increased the load admittance of every stage decreases, so increasing the feedback and increasing C_{in}. As C_{in} is used to tune the load, the load admittance is further decreased. This means the amplifier response will be very steep on the low-frequency side. After the frequency of maximum gain has been passed, the amount of feedback decreases, so decreasing C_{in} and this tends to keep the coupling circuits near resonance. As the frequency is increased the gain falls slowly.

If no shunt conductance g_e is present, the load conductance (g_{in}/n^2) will change over the passband, but this effect on the response will be approximately symmetrical.

This distortion can be reduced by

(i) putting extra capacitance in the coupling circuits, so that the contribution of C_{in} to the total tuning capacitance is small;

(ii) increasing ϕ_p so as to reduce the difference between b_{in} and b_{11}.

If feedback effects can be neglected distortion can still be caused by variation of g_{11}, g_{22}, Y_{21}, and also C_{11} and C_{22} over the passband.

(b) When used in a high frequency common emitter amplifier, most transistors will be operating above their α' cut-off frequency, $f_{\alpha'}$. It is well known that above this frequency the current gain is inversely proportional to frequency. Because of this g_{11} increases with frequency, and C_{11} decreases. The other main effect is a decrease of Y_{21} with frequency when the emitter current is several milliamperes (in the case of a small signal r.f. transistor), whereas Y_{21} is almost independent of frequency at low emitter currents. Variation of g_{22} with frequency is large but as $g_L \gg g_{22}$ usually this does not matter. G_{22} variations are relatively small.

It is not too difficult to calculate the effect of α' fall-off in amplifier response, but becomes impossible when the effects of C_{11} and Y_{21} are included, unless a digital computer is used.

The effect of α' variation can be exactly simulated for the case of a single stage fed from a current source, and terminated by a low load impedance by feeding a parallel tuned circuit from a generator whose output current is inversely proportional to frequency. (See Fig. 16.)

The admittance of a parallel tuned circuit is given by

$$(Y)^2 = G^2 + \left(\omega C - \frac{1}{\omega L}\right)^2 \qquad \ldots\ldots(33)$$

$$Q_0 = \frac{\omega_0 C}{G} = \frac{1}{\omega_0 LG} \qquad \ldots\ldots(34)$$

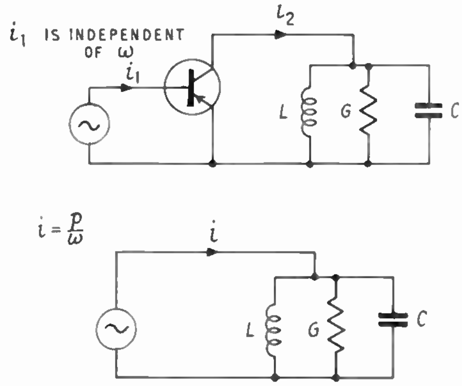


Fig. 16. Circuit for simulating variation in α'.

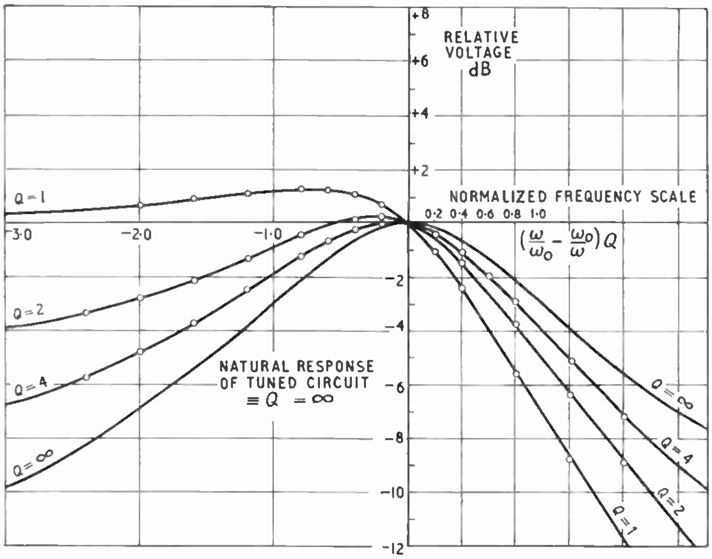


Fig. 17. Response of voltage developed across tuned circuit when fed with a current inversely proportional to frequency.

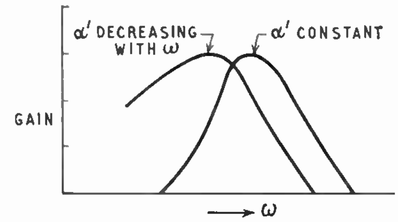


Fig. 18. Frequency response of wide band amplifier.

substituting

$$Y^2 = G^2\left[1+Q_0^2\left(\frac{\omega}{\omega_0}-\frac{\omega_0}{\omega}\right)^2\right] \quad \ldots\ldots(35)$$

When this circuit is fed with a current inversely proportional to frequency (i.e. $i = P/\omega$), the output voltage V_{out} is given by

$$\left(\frac{P}{V_{out}}\right)^2 = (\omega G)^2\left[1+Q_0^2\left(\frac{\omega}{\omega_0}-\frac{\omega_0}{\omega}\right)^2\right] \quad \ldots\ldots(36)$$

$$= G^2\left[Q_0^2\frac{\omega^4}{\omega_0^2}+(1-2Q_0^2)\omega^2+Q_0^2\omega_0^2\right] \quad \ldots\ldots(37)$$

$$\frac{d}{d\omega}\left(\frac{P}{V_{out}}\right)^2 = \left[\frac{4Q_0^2\omega^3}{\omega_0^2}+2(1-2Q_0^2)\omega\right]G^2 \quad \ldots\ldots(38)$$

When V_{out} is a maximum

$$\frac{d}{d\omega}\left(\frac{P}{V_{out}}\right)^2 = 0 \quad \text{at } \omega = \omega_m$$

$$0 = \frac{4Q_0^2\omega_m^3}{\omega_0^2}+2(1-2Q_0^2)\omega_m \quad \ldots\ldots(39)$$

$$\omega_m = \omega_0\sqrt{1-\frac{1}{2Q_0^2}} \quad \ldots\ldots(40)$$

The bandwidth of such a system was calculated as

3 dB bandwidth

$$= \omega_0\sqrt{2-\frac{1}{Q_0^2}-2\sqrt{1-\frac{2}{Q_0^2}-\frac{1}{Q_0^4}}} \quad \ldots\ldots(41)$$

Substituting $Q_0 = 1$ in this equation does not give an answer for the bandwidth because the term

$$\left(1-\frac{2}{Q_0^2}-\frac{1}{Q_0^4}\right)$$

is negative.

The reason is shown by plotting the response of the circuit for various Q_0's. The response becomes so distorted at low Q_0 values that the lower 3-dB point ceases to exist (see Fig. 17).

The responses of a wide band amplifier with α' constant and with α' decreasing at 6 dB/octave are shown in Fig. 18. There are two effects; the centre frequency of the amplifier is lowered and the response is distorted.

The effective current gain of a stage may be considerably modified in a typical multistage amplifier. Calculating the current gain of an iterative stage, such as shown in Fig. 6:

$$\frac{i_1}{i_2} = \frac{Y_{22}}{Y_{21}}+\frac{g_e}{nY_{21}}+\frac{Y_{11}}{nY_{21}} \quad \ldots\ldots(42)$$

If $nY_{22} \ll Y_{11}$ and $g_e \ll Y_{11}$

$$\frac{i_1}{i_2} = \frac{Y_{11}}{nY_{21}} = \frac{1}{n\alpha'} \quad \ldots\ldots(43)$$

If $g_e \gg Y_{11}$ and $g_e \gg nY_{22}$

$$\frac{i_1}{i_2} = \frac{g_e}{nY_{21}} \quad \ldots\ldots(44)$$

If the 6 dB/octave reduction of α' with frequency is reflected in Y parameter values as an increase in Y_{11} only, then using a large shunt conductance at the input will make the current gain independent of frequency. If Y_{21} alters with frequency the current gain of the stage will alter with frequency whatever loading is applied to the input port.

As ω_m, the frequency of maximum output, is less than the resonant frequency of the interstage coupling networks, the gain of each stage is reduced by a small amount. This does not usually exceed 1·5 dB.

To summarize therefore, there are two processes which distort the frequency response of a multistage amplifier; one due to the interdependence of stages, and the other due to the variation of device parameters over the passband. The interdependence of stages can be reduced by increasing the stability figure, or to some extent by adding capacitance to the coupling network. Parameter variations over the passband can be reduced by choosing the d.c. operating conditions carefully and also by shunting the input port with a large g_e.

6.4. *Choosing the Value of ϕ_p or G_{pn}*

For high gain narrow band amplifiers. To avoid distortion of the frequency response ϕ_p should be greater than 4 or 5, or $G_{pn} > 0{\cdot}2$. As the factor Y_{21}/Y_{12} can be very large high gains can still be obtained, for instance, an AFZ12 will have a gain > 20 dB at 50 Mc/s under these conditions.

For wide band amplifiers. It is difficult to find a general rule as the two types of distortion mentioned in section 1 tend to cancel one another. Amplifiers have been constructed with bandwidths of 18 Mc/s at 45 Mc/s using $\phi_p = 6$ with very good response shape. On the other hand a ϕ_p of 10 might not be sufficient. The shape of the response will indicate what kind of distortion is dominant, so that once an amplifier has been constructed logical steps can be taken to improve the response (see Sect. 6.3), such as increasing ϕ_p, increasing g_e or choosing another d.c. operating point to reduce the dependence of Y_{21} on frequency.

6.5. *Sensitivity of Power Gain to Parameter Changes*

Assume that the variations of different parameters are not correlated for any particular transistor type.

If the feedback parameter is neglected, the power gain equation is simplified.

From Fig. 19

$$V_{in} = \frac{I_1}{Y_{11}} \qquad \text{......(45)}$$

$$P_{in} = V_{in}^2 g_{11} \qquad \text{......(46)}$$

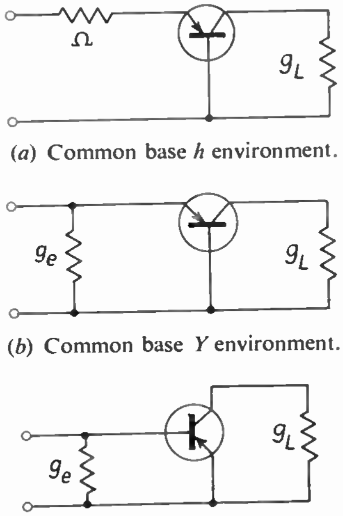
$$I_2 = Y_{21} V_{in} \frac{Y_L}{Y_{22} + Y_L} \qquad \text{......(47)}$$

Also

$$I_2' = \frac{g_L}{Y_L} I_2 \qquad \text{......(48)}$$

$$P_{out} = I_2'^2 \qquad \text{......(49)}$$

$$P_{out} = \frac{V_{in}^2 |Y_{21}|^2 g_L}{|Y_{22} + Y_L|^2} \qquad \text{......(50)}$$



(*a*) Common base *h* environment.

(*b*) Common base *Y* environment.

(*c*) Common emitter *Y* environment.

Fig. 19.

$$\frac{P_{out}}{P_{in}} = \frac{|Y_{21}|^2 g_L}{|Y_{22} + Y_L| g_{11}} \qquad \text{......(51)}$$

This expression is maximized when $b_{22} + b_L = 0$, i.e. maximum gain is obtained when the load susceptance cancels the output susceptance:

$$\frac{P_{out}}{P_{in}} = \frac{|Y_{21}|^2 g_L}{(g_{22} + g_L) g_{11}} \qquad \text{......(52)}$$

Similarly it can be shown that the power gain equation for *h*, *g* and *Z* parameters takes an identical form. If *d* stands for *Y*, *Z*, *h*, or *g*.

$$\frac{P_{out}}{P_{in}} = \frac{|d_{21}|^2 \mathrm{Re}\, d_L}{(\mathrm{Re}\, d_{22} + \mathrm{Re}\, d_L) \mathrm{Re}\, d_{11}} \qquad \text{......(53)}$$

From this it can be seen that the dependence of power gain on d_{22} can be reduced by making $d_L \gg d_{22}$. Similarly the effect of d_{11} is reduced by loading the input with an immittance† d_e, so that d_{11} can be replaced by $(d_e + d_{11})$ in eqn. (52):

$$\frac{P_{out}}{P_{in}} = \frac{|d_{21}|^2 \mathrm{Re}\, d_L}{(\mathrm{Re}\, d_{22} + \mathrm{Re}\, d_L)(\mathrm{Re}\, d_{11} + \mathrm{Re}\, d_e)} \qquad \text{......(54)}$$

If $\mathrm{Re}\, d_L \gg \mathrm{Re}\, d_{22}$ and $\mathrm{Re}\, d_e \gg \mathrm{Re}\, d_{11}$

$$\frac{P_{out}}{P_{in}} = \frac{|d_{21}|^2 \mathrm{Re}\, d_L}{\mathrm{Re}\, d_L\, \mathrm{Re}\, d_e} = \frac{|d_{21}|^2}{\mathrm{Re}\, d_e} \qquad \text{......(55)}$$

The only parameter now affecting the gain is the forward transfer parameter. Loading at input and output ports cannot change the dependence of gain on this parameter. It is interesting, therefore, to compare the variations in parameter d_{21} between different units when the transistor is operated in different con-

† *Immittance:* Term first used by Linvill to denote impedance or admittance.

figurations (common emitter, common base, common collector) and in different matrix environments (h, Y, Z, g). It is obvious that the best defined parameter is α (h_{21}, in grounded base), as it is very close to unity over nearly the whole useful operating frequency range. Y_{21} in grounded emitter and in grounded base is quite well defined at low emitter currents, its value being approximately 40 I_e millimhos. Both Z_{21} and g_{21} are less well defined than this in any configuration.

From the above it would appear that three types of circuit would give a well defined gain. These are shown in Fig. 19 and are: (*a*) common base h environment, (*b*) common base Y environment, (*c*) common emitter Y environment. Type (*b*) must be rejected because, in general, the input admittance is so large that it is difficult to shunt with y_e. Also, the coupling transformer will not be perfect, and this low secondary load impedance will emphasize the effects of leakage inductance.

The gain of type (*a*) stage is better defined than that of type (*c*), but type (*c*) can give a higher power gain. It is difficult to compare their gain-bandwidth products theoretically. Judging from a small number of amplifiers constructed, type (*c*) would seem to have the highest gain-bandwidth product, although this cannot be taken as a rule.

It should be noted that the assumption made at the beginning of this section, that there is no correlation between the variations of the parameters, may not be true. The lower limit to gain variations may not be set by variations in d_{21} in this case, because port terminations could be used which allowed d_{11} or d_{22} variations to affect the gain, and so tend to offset the effects of d_{21} variations. This correlation would seem to have been assumed by Lathi, who approaches sensitivity via the equivalent circuit. The manufacturing process would seem to be the vital point here.

6.6. *Gain-Bandwidth Product*

Giacolletto[6] has proposed that the gain-bandwidth product be taken as a figure of merit for a transistor. However, it has been shown by Venkateswaran that the gain-bandwidth product is not a constant, but depends on bandwidth even when the parameters are constant over the band.

Consider a transistor whose parameters are constant over the frequency range of interest and whose reverse transfer parameter is zero (lossless unilateralization). The maximum (conjugately matched) gain and bandwidth may be easily calculated. Also, it is obvious that there need be no reduction of gain for bandwidths up to a certain critical bandwidth (obtained when no extra capacitance is added across the load). To increase the bandwidth further the load conductance must be increased, hence reducing the gain.

If $Y_0 = g_L/g_{22}$, G_{pc} is the conjugately matched gain, W_c is the bandwidth at G_{pc}, G_p is the gain at any y_0, and W is the bandwidth at any y_0.

Then it can be shown that when the Q of the output circuit is much greater than that of the input circuit ($Q_2 \gg Q_1$)

$$G_p = G_{pc}\left[\frac{4y_0}{1+y_0}\right]^2 \qquad \ldots\ldots(56)$$

$$W = W_c\left(\frac{1+y_0}{2}\right) \qquad \ldots\ldots(57)$$

If $Q_1 = Q_2$

$$G_p = G_{pc}\left[\frac{4y_0}{(1+y_0)^2}\right]^2 \qquad \ldots\ldots(58)$$

$$W = W_c\left(\frac{1+y_0}{2}\right) \qquad \ldots\ldots(59)$$

Then for $Q_2 \gg Q_1$

$$G_p W = G_{pc} W_c \frac{2y_0}{1+y_0} \qquad \ldots\ldots(60)$$

If $y_0 \gg 1$

$$G_p W = 2G_{pc} W_c \qquad \ldots\ldots(61)$$

For $Q_2 = Q_1$

$$G_p W = G_{pc} W_c \frac{8y_0^2}{(1+y_0)^3} \qquad \ldots\ldots(62)$$

so that if $y_0 \gg 1$

$$G_p W = \frac{8G_{pc} W}{y_0} \qquad \ldots\ldots(63)$$

but

$$G_p W^2 = 4G_{pc} W_c^2 \qquad \ldots\ldots(64)$$

Usually $Q_2 \simeq Q_1$ so that eqn. (57) is more important. It shows that the product (gain) × (bandwidth)^2 is a constant, so that halving the gain will only increase the bandwidth by a factor of 1·4 when the transistor is heavily mismatched.

7. Tuning a Multistage Common Emitter Amplifier

Narrow-band amplifiers can be tuned quite normally with a signal generator and r.f. detector for maximum output at the required centre frequency. With $\phi_p > 5$ the tuning of one stage will only slightly affect the tuning of its immediate neighbours.

If the fractional bandwidth is large then the frequency dependence of the transistor parameters will affect the response shape (see Section 6.3). Figure 20 shows the locus of power gain as a multistage narrow-band amplifier is tuned over a wide range of frequencies. If the bandwidth is now increased and the amplifier tuned for maximum output with an input frequency f_0, the power gain at f_0 will be the same as for a narrow-band amplifier tuned to f_0. However, plotting the response will show that maximum power gain occurs at a lower frequency so that the amplifier is not centred on f_0. To centre the response on f_0

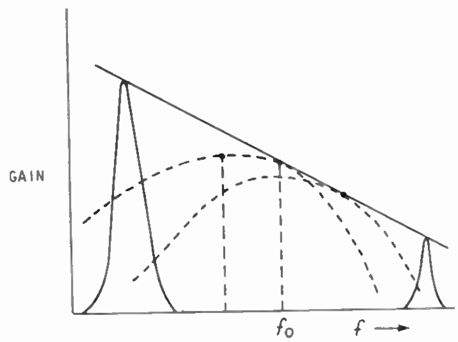


Fig. 20. Locus of power gain against frequency for multistage narrow band amplifier.

the interstage coupling networks must be tuned to a higher frequency. Gain at f_0 will then be slightly less than for the narrow-band case.

There are many possible tuning methods.[7] The three given below all use a swept frequency source and panoramic display.

(*a*) Observe the response at the first collector or second base and tune to centre at the required frequency. Repeat with succeeding stages, then check again.

(*b*) Tune for maximum gain at some higher frequency and then observe the overall response. If not centred correctly tune for maximum gain at a slightly different frequency and so on.

(*c*) Observing the output tune all stages together to obtain the correct centre frequency. For slug tuning this would mean moving all slugs to the same position in each coil.

The last method is obviously the quickest. Due to variations of stray and collector capacitance the stages will not be synchronously tuned by this method, but it has been found quite successful in practice.

8. Practical Interstage Coupling

Any four terminal network which presents a conductance g_L at its input port when loaded with g_{in} at its output port can be used to couple the stages together. The power losses in the network should be kept small because these will reduce the overall gain of the amplifier. In Section 4.2 the coupling network was assumed to be a perfect transformer of ratio n. At high frequencies a practical transformer may have a coupling coefficient less than 0·7 although its power losses may be quite small. Due to this poor coupling coefficient the impedance transformation will not equal the square of the turns ratio. By reducing the turns ratio the impedance transformation can be brought to the correct value. This can be done quite easily with the aid of a r.f. bridge, by measuring the primary conductance with the secondary loaded with g_{in} and adjusting the turns ratio until the required value of g_L is obtained. (The input conductance g_{in} is taken here to include any g_e.)

The shunt conductance g_e can be transferred to the primary side of the transformer, but in this case the partial stability factor of the stage will be reduced because the admittance looking into the secondary will consist of g_e in series with any leakage reactance present. In practice this is only a minor disadvantage.

Hetterscheid[7] has examined bandpass coupling of transistor i.f. amplifiers.

9. Two Examples of the Design Method

Parameters of several AFZ12 transistors were measured at 45 Mc/s in common emitter. Typical parameters are given below for $I_e = 3$ mA, $V_{ce} = 10$ V.

$$Y_{11} = 8{\cdot}5+\text{j}12 \text{ millimhos}$$
$$Y_{12} = 0-\text{j}0{\cdot}12 \text{ millimhos}$$
$$Y_{21} = 65-\text{j}50 \text{ millimhos}$$
$$Y_{22} = 0{\cdot}04+\text{j}0{\cdot}56 \text{ millimhos.}$$

From these $M = 0$, $\cos\theta = -0{\cdot}5$

and $$\frac{Y_{21}}{Y_{12}} = 690 \text{ millimhos}$$

9.1. *First Design: Three Stage Amplifier, 60 dB Nominal Gain* (Fig. 21)

A stage gain of 20 dB would require

$$G_{pn} = \frac{100}{690} = 0{\cdot}145$$

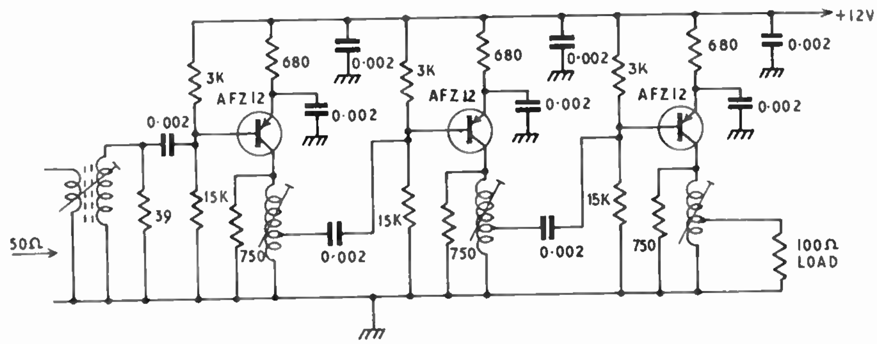


Fig. 21.
45 Mc/s AFZ12 amplifier first design.

As this is greater than 0·2 the response distortion due to feedback will not be serious.

Now $$\frac{\phi_p + \cos\theta}{\phi_p^2 - 1} = G_{pn}\left(1 + \frac{g_{22}}{g_L}\right) \qquad \ldots\ldots(28)$$

If $g_{22} \ll g_L$

$$\frac{\phi_p + \cos\theta}{\phi_p^2 - 1} \simeq \frac{G_{pn}}{2} = \frac{0{\cdot}145}{2} \qquad \ldots\ldots(29)$$

Therefore $$\phi_p = 13{\cdot}5$$

The coil ratio was chosen to be 4·4 : 1. M, G_{pn}, ϕ_p, $\cos\theta$, n^2 are now known.

From $$\phi_p^2 + 2\phi_p\cos\theta + 1 = \frac{4(g_{11}+g_e)^2}{M_n^2 G_{pn}} \qquad \ldots\ldots(26)$$

$$g_{11} + g_e = 34{\cdot}4 \text{ millimhos}$$

and $$g_e = 34{\cdot}4 - 8{\cdot}5 = 26{\cdot}9 \text{ millimhos}$$

From eqn. (16), the input conductance at each stage is

$$g_{\text{in}} = 34{\cdot}4 \times \frac{\phi_p^2 - 1}{\phi_p^2 + 2\phi_p\cos\theta + 1}$$

$$g_{\text{in}} = 36{\cdot}8 \text{ millimhos}$$

The load conductance is $g_{\text{in}}/n^2 = 1{\cdot}9$ millimhos.

9.1.1. Circuit details

The conductance g_e was moved into the primary circuit of the coupling transformer.

$g'_e = g_e/n^2 = 1{\cdot}4$ millimhos (resistance 720 Ω)

The secondary load was then $g_{\text{in}} - g_e$, or about 10 millimhos. The source impedance was matched to g_{in} with an input transformer.

9.1.2. Design to check the effects of bandwidth

No account was taken of parameter variations over the passband. The Q of the coupling circuits was calculated from $Q = \omega_0 CR$, C being the total circuit capacitance.

From $$\frac{b_{\text{in}} - b_{11}}{g_{11} + g_e} = \frac{-2\phi_p \sin\theta}{\phi_p^2 + 2\phi_p\cos\theta + 1} \qquad \ldots\ldots(17)$$

$$= 0{\cdot}068$$

$$b_{\text{in}} = 2{\cdot}4 + 12 = 14{\cdot}4 \text{ millimhos}$$

$$C_{\text{in}} = 51 \text{ pF}$$

This appears at the collector as $51/n^2 = 2{\cdot}6$ pF. If C_c = collector capacitance, C_s = stray capacitance.

Total tuning capacitance $= C_c + C_{\text{in}}/n^2 + C_s$

$$= 2 + 2{\cdot}6 + 2 = 6{\cdot}6 \text{ pF}$$

The load resistance was $\dfrac{1000}{1{\cdot}9} = 530\ \Omega$

Therefore with no extra capacitance across the load

$$Q = \omega_0 CR = 1$$

This will give a stage bandwidth of 45 Mc/s. The reduction factor for three cascaded stages is

$$\sqrt{2^{1/3} - 1} \quad \text{or } 0{\cdot}51$$

Therefore the overall calculated bandwidth is 23 Mc/s. With 5·6 pF added across the load the overall calculated bandwidth is 12·5 Mc/s.

9.1.3. Results

(1) For the first test the amplifier was tuned for maximum power gain with a 45 Mc/s input signal (Centre frequency ≠ 45 Mc/s). The voltage gain of each stage was measured (base to base)

Stage	1	2	3
Voltage gain	19 dB	19 dB	19·5 dB

Measured total gain: 57 dB.

(2) Tuning the amplifier to centre on 45 Mc/s; all dust cores in the same position in the coils.

Stage	1	2	3
Voltage gain	17·5 dB	17·5 dB	18 dB

Measured total gain: 52 dB.

The symmetry of the response can be expressed as a symmetry factor $\dfrac{\sqrt{f_1 f_2}}{f_0}$ where f_1 and f_2 are frequencies below and above the centre frequency, f_0, where the gain of the amplifier is the same. For example, these could be the 3 dB points. The factor is unity for a simple LCR circuit of any Q factor.

Upper 3 dB point	Lower 3 dB point	Band-width	$\dfrac{\sqrt{f_1 f_2}}{f_0}$
56 Mc/s	39 Mc/s	17 Mc/s	1·04

With 5·6 pF added across the tuned circuit and returned to centre on 45 Mc/s

Measured total gain = 54 dB.

Upper 3 dB point	Lower 3 dB point	Band-width	$\dfrac{\sqrt{f_1 f_2}}{f_0}$
53·5 Mc/s	40 Mc/s	13·5 Mc/s	1·028

These results show clearly the effects of bandwidth on gain mentioned in Sections 6.3 and 6.5 and also response distortion, predominantly due to internal feedback. (From Section 6.3 it is seen that for feedback distortion the symmetry factor will be greater than unity, whilst for distortion caused by parameter changes in the amplifier passband it will be less than unity.)

As predicted in Section 6.3, adding capacitance across the tuned circuit improved the symmetry of the response. When tuned with stray capacitance the bandwidth was much narrower than that calculated. When the tuning capacitance was increased reasonable agreement was obtained between calculated and measured values.

9.1.4. Alternative design with partial stability figure $\phi_p = 33$ (also three stages)

The coil ratio is the same as before (4·4 : 1).

From eqn. (21) $g_L = 2{\cdot}94$ millimhos.

From eqn. (22) $g_e = 47$ millimhos.

The value of G_{pn} is then computed to be 0·06, $g_{\text{in}} = 56{\cdot}5$ millimhos and $b_{\text{in}} = 13{\cdot}5$ millimhos. The gain/stage is therefore 16·2 dB.

9.1.5. Bandwidth calculation

$$\text{Total tuning capacitance} = C_c + C_{\text{in}}/n^2 + C_s = 2 + 2{\cdot}5 + 2 = 6{\cdot}5 \text{ pF}$$

With no extra capacitance across the load bandwidth = 36 Mc/s overall.

With 8·2 pF added across the load bandwidth = 16 Mc/s overall.

9.1.6. Results

When tuned for maximum power gain with 45 Mc/s input frequency (centre frequency ≠ 45 Mc/s) the measured power gain over 3 stages is 46·5 dB, or 15·5 dB/stage.

Tuning with stray capacitance to centre on 45 Mc/s. All dust cores adjusted to the same position in the coils.

Gain	Lower 3 dB point	Upper 3 dB point	Band-width	$\dfrac{\sqrt{f_1 f_2}}{f_0}$
42 dB	36 Mc/s	57 Mc/s	21 Mc/s	1·01

Tuning with 8·2 pF across each load, as before.

Gain	Lower 3 dB point	Upper 3 dB point	Band-width	$\dfrac{\sqrt{f_1 f_2}}{f_0}$
44 dB	39 Mc/s	54 Mc/s	15 Mc/s	1·02

9.1.7. Conclusions

An improvement in response shape was obtained by increasing ϕ_p from 13·5 to 33. Agreement between calculated and measured values of bandwidth was poor unless a small padding capacitor was used. The gain-bandwidth product was not the same at the two values of ϕ_p, and for this amplifier there was an even poorer exchange between gain and bandwidth than indicated in Section 6.6.

9.2. *Design of a Wide Band Silicon Transistor Amplifier for a Particular Gain/Stage*

Centre frequency: 45 Mc/s

Transistor: 2S131

D.C. conditions: 10 mA I_e 3 V V_{ce}

These unusual operating conditions were determined by other requirements.

Parameters: (Average of 10 in common emitter)

$Y_{11} = 4 + j4{\cdot}2$ millimhos

$Y_{12} = -0{\cdot}05 - j0{\cdot}5$ millimhos

$Y_{21} = 45 - j50$ millimhos

$Y_{22} = 2{\cdot}5 + j1{\cdot}6$ millimhos

These results were obtained using a Wayne Kerr admittance bridge

required gain/stage: 11 dB.

9.2.1. Design

(*a*) From the above parameters

$$M = 34 \qquad \theta = 143^\circ$$

$$\left|\frac{Y_{21}}{Y_{12}}\right| = 136 \text{ millimhos}$$

Allowance must be made for the difference in power gain between narrow band and wide band amplifiers centred on the same frequency. (See Sections 6.3 and 6.5.)

The amplifier will therefore be designed for 12·5 dB gain/stage

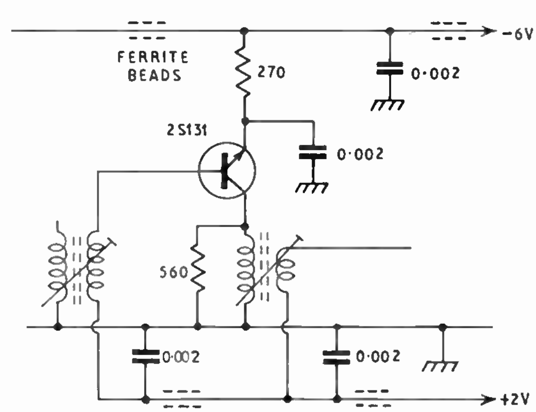


Fig. 22. Typical stage of the 11 dB/stage 2S131 amplifier.

$$G_p = 17{\cdot}8$$

$$G_{pn} = \left|\frac{Y_{12}}{Y_{21}}\right| G_p = 0{\cdot}131$$

As $G_{pn} < 0{\cdot}2$ the response distortion due to feedback will be small. (See Section 6.3.)

The coil ratio was chosen to be 3 : 1.

(*b*) To start the calculation choose $\phi_p = 6$.

(*c*) From Fig. 7, $\phi_p = 6$, $\theta = 143°$, $C = 1{\cdot}2$.

(*d*) From Fig. 9, $\dfrac{4(g_{11}+g_e)^2}{Mn^2G_{pn}} = 29$

for $\phi_p = 6$ and $\theta = -143°$.

(*e*) From (*d*) $g_{11}+g_e = 17$ millimhos

$$g_L = \frac{C(g_{11}+g_e)}{n^2}$$

$$g_L = 2{\cdot}27 \text{ millimhos.}$$

(*f*) $\dfrac{G_{pn}}{2}\left(1+\dfrac{g_{22}}{g_L}\right) = \dfrac{0{\cdot}131}{2}\left(1+\dfrac{2{\cdot}5}{2{\cdot}27}\right) = 0{\cdot}137.$

(*g*) From Fig. 10, $\dfrac{G_{pn}}{2}\left(1+\dfrac{g_{22}}{g_L}\right) = 0{\cdot}137$

and $\theta = -143°$.

(*h*) The second value of ϕ_p is slightly greater than the first. Starting again with $\phi_p = 6{\cdot}5$.

(*e*) $C = 1{\cdot}18$.

(*d*) $\dfrac{4(g_{11}+g_e)^2}{Mn^2G_{pn}} = 35.$

(*e*) $g_{11}+g_e = 18{\cdot}7$ millimhos

$g_L = 2{\cdot}45$ millimhos.

(*f*) $\dfrac{G_{pn}}{2}\left(1+\dfrac{g_{22}}{g_L}\right) = 0{\cdot}132.$

(*g*) $\phi_p = 6{\cdot}5$.

The stage input conductance will be $C(g_{11}+g_e)$ or $18{\cdot}7 \times 1{\cdot}18 = 22$ millimhos.

The shunt conductance g_e will be

$$18{\cdot}7 - 4 = 14{\cdot}7 \text{ millimhos.}$$

The input capacitance b_{in} was found from Fig. 8 for $\phi_p = 6{\cdot}5$ and $\theta = -143°$.

$$\frac{b_{in} - b_{11}}{g_{11}} = 0{\cdot}23$$

$$b_{in} = 0{\cdot}23 g_{11} + b_{11}$$

$$= (0{\cdot}23 \times 18{\cdot}7) + 4{\cdot}2$$

$$b_{in} = 8{\cdot}5 \text{ millimhos or } 30\,\text{pF}$$

This appears at the collector as 30/9 pF or 3·3 pF. The collector capacitance is about 5·5 pF, giving a total capacitance of 8·8 pF without stray capacitance. If a stray capacitance of 1 pF is assumed the total tuning capacitance is approximately 10 pF. The total conductance appearing across the tuned circuit primary is g_L+g_{22} if feedback is neglected.

The Q factor of each interstage coupling network is then

$$Q = \frac{\omega C}{g_L+g_{22}} = \frac{3{\cdot}5}{5} = 0{\cdot}7$$

The stage bandwidth is therefore 45/0·7 = 64 Mc/s.

For six stages the bandwidth reduction factor is $\sqrt{2^{1/6}-1}$ or 0·35.

The overall bandwidth is thus calculated to be 22·5 Mc/s.

9.2.2. Results

The measured gain of a six stage amplifier was 71 dB (11·7 dB/stage) and the overall bandwidth 18 Mc/s. The response was not distorted.

10. Acknowledgments

The authors thank the Directors of E.M.I. Electronics Ltd. for permission to publish the paper.

11. References

1. B. P. Lathi, "Optimal Design of Multi-stage Tuned Transistor Amplifiers considering Gain, Stability and Sensitivity". Technical Report No. 755–3, July 1960. Stanford Electronics Laboratories, Stanford University, California.
2. A. Linvill and L. G. Schimpf, "Design of tetrode transistor amplifiers", *Bell Syst. Tech. J.*, **35**, p. 813, 1956.
3. A. P. Stern, "Stability and power gain of tuned transistor amplifiers", *Proc. Inst. Radio Engrs*, **45**, p. 335, March 1957.
4. G. S. Bahrs, "Stable amplifiers employing potentially unstable transistors", *I.R.E. Wescon Convention Record*, Part 2 (Circuit Theory), p. 185, 1957.
5. S. Venkateswaran, "Stability, Power Gain and Bandwidth of Linear Active Four-Pole Networks with particular reference to Transistor Amplifiers at High Frequencies". Imperial College, University of London, Ph.D. Thesis, June 1961.
6. L. J. Giacoletto, "Study of pnp alloy junction transistors from d.c. through medium frequencies", *R.C.A. Rev.*, **15**, p. 506, December 1954.
7. W. Hetterscheid, "Transistor i.f. amplifiers. Design procedure for obtaining optimum performance", *Electronic Applications*, **21**, No. 4, p. 141, 1960–61, and **22**, No. 1, p. 1, and No. 2/3, p. 41, 1961–62.

Manuscript received by the Institution on 29th July 1963 (Paper No. 876)

U.D.C.620.178.3: 621.375.4

A High-efficiency Low-frequency Power Source for Vibration Excitation

By

B. H. VENNING,
B.Sc.(Eng.), A.C.G.I.
(*Associate Member*)†

Presented at the Convention on "Electronics and Productivity" in Southampton on 19th April 1963.

Summary: The production of continuous sinusoidal vibrations of low frequency or transients of a comparably long period is beyond the scope of existing electro-dynamic force generators in which displacement is relatively limited. Moreover, the transformer-coupled amplifiers employed to drive the units have a limited low-frequency bandwidth unable to cater for the pulse shapes required.

This paper describes the power source of a system which has been devised to overcome these disadvantages and to demonstrate a means of high-efficiency amplification which can also be employed with the existing types of vibration generators. The power amplifier employs a pulse-width modulated signal and output transistors operating as saturated switches, with a frequency bandwidth from d.c. to about 350 c/s. The efficiency of the output stage is high (85–90%), and considerable economy of design is possible at high powers.

1. Introduction

Vibration waveforms which come under the general classification of 'long-period and large displacement' arise in the operation of a number of different systems. Among these are:

(*a*) A transient shock arising in environmental testing for packaging and fragility assessment, which is essentially a single transient with a waveform varying from rectangular to a half-sinusoid, of up to 6 in or more displacement over a period 5–50 ms.

(*b*) A compression force pulse during the measurement of the dynamic stiffness characteristics of visco-elastic materials, such as foam rubber, foam plastic and other bulk cushioning materials, where samples have to be compressed to perhaps 80% of a normal thickness of 6–12 in.

(*c*) The motion of a platform used in the calibration of accelerometers by the application of a single acceleration pulse of known characteristics.

(*d*) Continuous sinusoidal oscillations in the low-frequency range 0·1–100 c/s.

Existing techniques for producing single transient shocks employ either drop-testing machines of various patterns with the load platform falling on to some form of arresting device, or pendulum-type machines where the load swings in an arc until it impacts against a suitable cushion. Alternatively, a force pulse can be generated by a pneumatic system or by hydraulic means, and the latter is certainly a very efficient system when dealing with forces of the order of tons. However, such a system becomes uneconomic with thrusts in the region of 10–500 lb, which is a range which is quite adequate for many of the applications quoted above, and it was with this in mind that a system using an electro-dynamic exciter was designed.

Conventional electro-dynamic vibrator systems are largely intended for continuous sinusoidal testing in the range 10 c/s–10 kc/s. They employ an electro-dynamic exciter in which the moving coil, located by one or more flexible 'spiders', moves in the annular gap of a magnet, travel being limited to $\pm$ 0·5 in by the degree of deflection that can be obtained from the coil mounting. The frequency response of the vibrator extends to d.c. at the lower end, with a low-frequency resonance depending upon the total moving mass and the stiffness of the suspension. The electrical drive to the exciter employs a sine-wave oscillator feeding a suitable power amplifier, usually operating as a transformer-coupled Class-B circuit, and at powers above 500 W, the output transformer becomes a massive (and expensive) item. Moreover, it imposes an effective low-frequency cut-off which is seldom below 10 c/s and this, coupled with the low-frequency resonance of the suspension, makes it unsuitable for use for sinusoidal testing below about 5 c/s, or for single pulses longer than 5 ms duration.

An alternative system has been developed which employs rather different principles. The power amplifier uses a pulse-width modulated signal which controls an output stage consisting of power transistors operating as saturating switches. This stage has an efficiency in the region 85–90%, so that the transistor dissipation is only a small fraction of the volt-amperes

† Department of Electronics, University of Southampton.

switched into the load. The overall frequency response is from d.c. to 350 c/s, so that there is no limitation on low-frequency or long-pulse operation. The upper-frequency limit is largely a matter of economics and development, and it should be able to be extended to the low audio frequency region. This type of amplifier, with its high overall efficiency and compact size for a given power output, may well find applications outside the vibration field. The electro-dynamic force generators use sliding coils, the travel being 2–6 in depending upon the employment. Finally, associated with the operation are high input-resistance transistor pre-amplifiers for use with piezo-electric transducers, and variable-capacitance force and displacement transducers.

This paper is concerned with the design of the power source for the exciter.

2. Pulse-width Modulated Power Amplifiers

A system using pulse-width modulation conveys its information by the variation of the duration of individual pulses of constant amplitude and having a fixed p.r.f. which is higher than the maximum signal frequency (Fig. 1). The signal is detected by a low-pass filter, which removes all the harmonics of the p.r.f. and provides an output proportional to the original signal input. When this system is used in an amplifier, the input signal is first converted to a pulse-width modulated signal at a low power level, and the pulse waveform is then amplified to a level suitable to drive the output stage. As the requirement is that of constant pulse amplitude, the output transistors can operate as saturated switches, and the resulting collector current waveform is then filtered to remove harmonics of the switching frequency, so that the final load current is linearly related to the input signal over the working bandwidth.

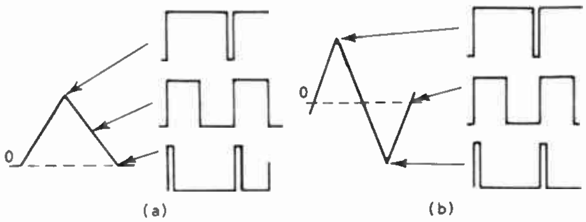


Fig. 1. Pulse-width modulated signals.

(*a*) Pulse trains for unidirectional signal. Zero signal = 0·0 duty-cycle.

(*b*) Pulse train for bidirectional signal. Zero signal = 0·5 duty-cycle.

Amplifiers based on this principle have received some attention as d.c. power amplifiers[1,2] and as a.c. coupled audio amplifiers,[3,4] but the development to

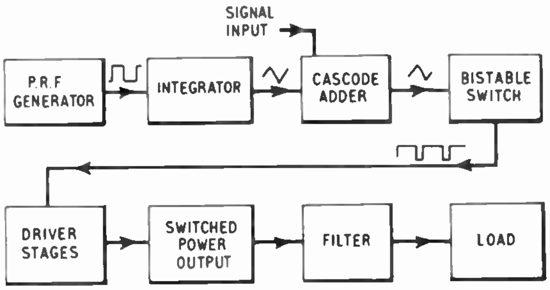


Fig. 2. Block schematic for complete amplifier.

give a bidirectional output catering for low-frequency a.c. signals seems to have been ignored. This is possibly because the low switching speeds obtainable with germanium power transistors severely limit the signal bandwidth, but the introduction of high-speed switching transistors has now extended the limit well into the audio range. The complete circuit of the amplifier (Fig. 2) falls into two basic sections:

(i) the pulse-width modulator, which operates at low power and produces a pulse train with the duty-cycle proportional to input signal amplitude;

(ii) the switched power output stages, which include a driver stage to transform impedance levels between the modulator and the heavy-current switched output transistors, and finally the filter circuit to remove switching frequency components from the load.

3. Frequency Response and Efficiency

The primary advantage of the pulse-width modulated (or 'switched') amplifier lies in the high efficiency which can be obtained. When conventional amplifiers are used with a d.c. coupling to the load, the maximum efficiency of the stage is low (25% for Class A operation) and the fact that the power to the load can be no greater than one-half of the transistor dissipation immediately necessitates a large total dissipation for the output stage. Class-B transformer-coupled circuits realize much higher values of efficiency (ideally 78·5%), but the frequency response is limited to that of the output transformer. Design considerations normally impose a lower limit of 5–10 c/s on this, and even if a reduction of high-frequency response is accepted to allow a higher value of shunt inductance to be used, the result is an exceedingly large and expensive transformer. There is also still a large power wastage in the output stage, which has to be shared between a number of output transistors in parallel, but this approach has been used to produce transistor audio amplifiers with outputs up to kilowatts.[5]

When an absolute d.c. response is not essential, the low-frequency cut-off must be determined by operating considerations. The reduction in signal for continuous a.c. testing is obvious from the lower -3 dB limit, but when working with single transients, some criterion of acceptable pulse distortion must be used. For example, if the peak value of a 20-ms triangular or half-sinusoidal input pulse is not to be more than 5% below its true value because of the low-frequency cut-off, then this must be below 1·5 c/s. In such cases, it is often preferable to extend the response down to d.c. rather than attempt to obtain suitable values from transformer coupling.

The transistor makes an exceedingly efficient switch, and a typical power transistor ($V_{CE(max)} = 60$ V, $I_{C(max)} = 6$ A) will have a collector saturation voltage of about 1 V. If used at its maximum rating, it will therefore dissipate 6 W at the collector when 360 W are being dissipated in the collector load resistor, giving an efficiency, on these figures, of 98%. Other losses arise however, as the waveforms of voltage and current are not perfectly rectangular, and the total transistor power loss arises from:

(*a*) the collector saturation loss, proportional to the saturation voltage;

(*b*) collector leakage loss, during the part of the cycle that the transistor is nominally cut-off;

(*c*) base loss, from the product of base-emitter voltage and base current;

(*d*) transient switching losses during the rise and fall times of the collector current.

Calculations of these can be made if simplified waveshapes are assumed,[6] but in fact the base voltage and current waveforms are far from rectangular, and with the inductive load presented by the filter, the transient losses are complicated by the phase shift between collector voltage and current waveforms. If the switching times are short compared to the pulse length, the predominant loss is that due to collector saturation, but as the switching p.r.f. is increased to obtain a higher overall signal frequency response, then the contribution from the transient losses increases.

A filter circuit of some description must be incorporated between the output transistors and the load in order to reduce the magnitude of the harmonics of the switching frequency in the load circuit. There are some applications in which these frequency components are unimportant, but in most cases it is necessary to consider the maximum tolerable ripple current and to filter accordingly. The circuits described later in this paper have been developed using a single L–C filter section, similar to a choke-input filter, but more complex filter networks can obviously be used. Component values can be calculated from a consideration of the upper signal bandwidth and the magnitude of ripple at the switching frequency. Work is still proceeding to relate the optimum values of switching frequency, maximum signal frequency, and load current ripple, as it is obviously important to keep the switching frequency as low as possible.

This aspect is also connected with the actual collector efficiency obtained, as the figure of 95–98% quoted above would only be obtainable with a low-frequency (< 150 c/s) unidirectional output. In the bi-directional circuit (Section 5.4), the peak load voltage must not exceed half the $V_{CE(max)}$ rating of the transistors, which reduces the output VA to half the unidirectional case, whilst maintaining the same collector saturation loss. Silicon power transistors ($f_\alpha = 2$ Mc/s, $I_{C(max)} = 6$ A) can be obtained with a superior switching performance to the lower-priced germanium power transistors ($f_\alpha = 200$ kc/s, $I_{C(max)} = 8$ A), and these enable a higher switching speed to be used. However, $V_{CE(sat)}$ for non-selected transistors of this type may be as much as 4–5 V at 6 A, giving a collector saturation loss of 25–30 W when switching 150 W into the load, and a collector efficiency of only 85%. Further power losses occur in the driving stage and in the diodes and the choke of the output stage, but as these are all distributed over different components they do not increase the complexity of the output stage. The chief disadvantage of the lower efficiency of the conventional amplifier (Class A or B) arises because the loss is concentrated at the collectors of the output transistors, limiting the ratio of transistor dissipation to output power, and thus necessitating a large number of parallel output units.

4. The Pulse Width Modulator

The basic requirement is to vary the duty-cycle of a rectangular pulse-train of constant p.r.f. in accordance with the amplitude of the input signal. If the amplifier is to cater only for unidirectional outputs, i.e. either true d.c. or d.c. plus a superposed a.c., the resultant never reversing sign, then zero output corresponds to zero duty-cycle, and maximum output to something as near as possible to unity duty-cycle. Thus for linear amplification for signals down to 1% of maximum, the variation in duty-cycle must exceed 100 : 1, with a minimum pulse-length less than 1% of the maximum. (A modification which reduces this stringent requirement is given in Section 5.3.) The more useful case of an amplifier with a true bidirectional output can be constructed with an output circuit which gives zero load current at 0·5 duty cycle, the negative output increasing as the cycle decreases from 0·5 to zero, and the positive output increasing from 0·5 to unity. In this case, the ability to cater for very low duty-cycles is not so essential, as it only reduces the peak output signal slightly, without affecting linearity at small signals. If the load is suitable for dividing into two halves, such as a centre-tapped driving coil, then it is also possible to obtain a bidirectional output from two identical unidirectional amplifiers, one driven by positive half-cycles and the other driven by negative half-cycles.

The most satisfactory method of obtaining a large linear variation of duty-cycle employs a triangular wave of a constant amplitude which exceeds the peak signal, and a p.r.f. which is also greater than signal frequency (Fig. 3). This is added to the signal, and the combined waveform used to trigger a bistable voltage-sensitive circuit of the Schmitt trigger type.

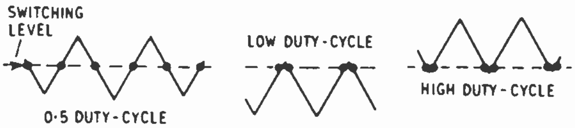


Fig. 3. Waveforms at input to bistable switch.

The output from this is therefore a rectangular pulse of fixed amplitude but a duration proportional to input signal amplitude. If the triangular waveform has a linear rise and fall, not necessarily of equal duration, or a linear rise and infinitely fast fall, then the linearity between output pulse and input amplitude is true even when the trigger circuit has finite 'backlash', i.e. the voltage to trigger 'on' is different to that to trigger 'off'. The waveform with linear rise and fall can be generated by integration of a rectangular wave, but if an attempt is made to obtain a very rapid flyback, this usually entails an exponential recharging of a capacitor. It therefore becomes essential to make the flyback duration very short compared to the linear rise, and with sweep periods of 20–100 μs, it becomes difficult to keep the flyback to the order of 1%.

The pulse-width modulator therefore consists of:

(*a*) *Square-wave generator.* This produces a square wave of about 0·5 duty cycle at the desired switching frequency, which may be 2–50 kc/s depending upon the ultimate switching speed of the output transistors. An emitter-timed circuit employing fast switching transistors offers advantages of frequency stability, fast rise and fall times and output amplitude stability.[7]

(*b*) *Integrator.* The square wave is a.c. coupled to a single transistor amplifier using capacitance feedback, as in a Miller integrator circuit. With the integrating time-constant of the same order as the pulse period, sufficient linearity of output is obtained.

(*c*) *Cascode adder.* It is again advantageous to a.c.-couple the triangular waveform to this circuit to avoid the d.c. drift of the integrator output. The cascode circuit employed for this has a d.c. input from the signal source to one transistor and the a.c. input from the integrator to the other, so that there is no interaction between the two inputs, which can occur if the two inputs are added directly to the base of a single transistor, or via a virtual earth. The output is the direct combination of the two inputs.

(*d*) *Amplitude switch.* This is an emitter-coupled bistable circuit, with adjustments of trigger level and backlash. As this circuit triggers as the input base voltage passes the critical level, it is essential to supply this base from a low-impedance source, so the output from the cascode-adder is d.c.-coupled to it via an emitter-follower. The bistable circuit triggers between its two states as the input base signal moves either side of the critical switching level, with a small backlash voltage between the two. This is normally about 100 mV, but it can be reduced to as little as 20 mV at some sacrifice of switching speed. The trigger voltage level remains constant (apart from a small adjustment) and a bias control on the d.c. channel of the adder enables a choice to be made between the two requirements of zero signal being equivalent to zero duty-cycle for unidirectional outputs, and zero signal corresponding to 0·5 duty-cycle for bidirectional outputs.

The pulse-width modulator is designed to work up to a maximum p.r.f. of 50 kc/s, and the output pulse has an amplitude of 6 V across a 1 kΩ collector load, with rise and fall times of approximately 0·2 μs.

5. Power Output Stage

There are a number of different possibilities for this stage, depending upon the load requirements.

5.1. *Unidirectional Unfiltered Output*

This is the simplest arrangement (Fig. 4), consisting only of a common-emitter transistor, switched between cut-off and saturation, giving substantially rectangular pulses of current through the load with a mean value proportional to the duty cycle of the base

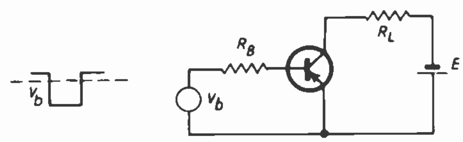
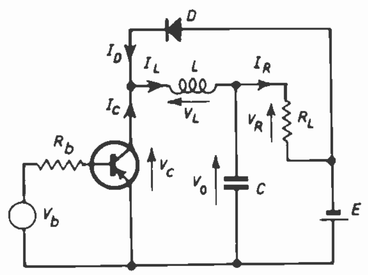


Fig. 4. Unidirectional unfiltered output.

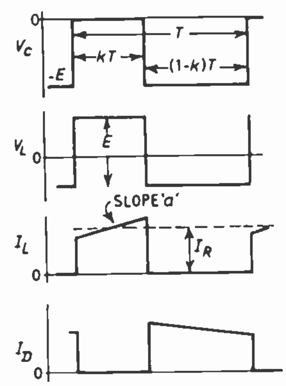
driving waveform. The collector current waveform exhibits switching transients with a turn-on delay, finite rise and fall times and a desaturation time, and these contribute additional collector power dissipation as well as limiting the values of minimum and maximum duty-cycles. For maximum switching speeds, the source impedance for the base drive circuit must be low, with resistor R_B chosen to give the required degree of overdrive into saturation. These transients contribute a non-linearity between the nominal duty-cycle and the average load current at low values of input. The circuit is obviously only applicable when the harmonics of the switching frequency have no effect upon the load, and when the load itself has no effect upon the switching performance of the transistors.

5.2. *Unidirectional Filtered Output*

If the presence of higher frequencies in the load is undesirable, they can be considerably attenuated by the inclusion of a suitable filter, which must have a very low d.c. resistance (typically less than 0·1 Ω) to keep the power dissipation to a reasonable value. The



(*a*) Circuit.

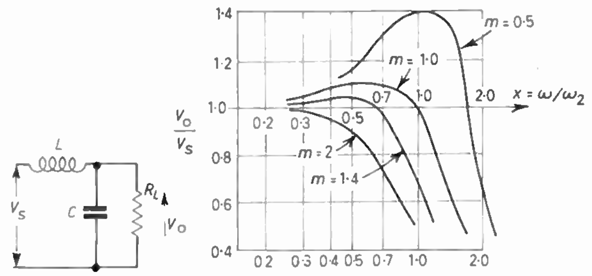


(*b*) Idealized waveforms.

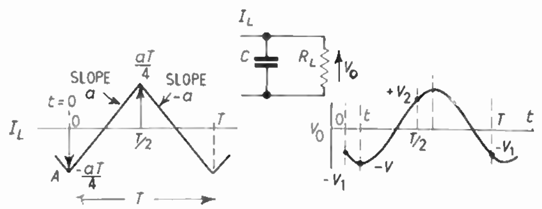
Fig. 5. Unidirectional filtered output.

filter must not have an adverse effect upon the collector switching waveform (which would occur if its first element was a shunt capacitance), and it must also act as a buffer to isolate the transistor from the effects of actual or motional reactance in the load. The simplest filter to fulfil these requirements is a single inductor-input section (Fig. 5(*a*)), and with a switching frequency of 5 kc/s and an upper signal frequency limit of about 350 c/s, suitable values for L and C are in the region 1 mH and 100 μF, depending upon the load resistance. The time-constant CR_L must be long compared to the switching period, or of the same order, so that V_o remains virtually constant over this time, only changing at the relatively slower signal frequency. As the transistor switches off, the current I_L decreases and the induced voltage V_L across L drives V_C negative, tending towards a maximum of $-2E$. The waveforms in Fig. 5 (*b*) are idealized, the actual waveform being influenced by the necessity to remove the saturation base charge from the transistor before any change in V_C can be observed, and the subsequent motion being part of a damped oscillation between the inductance and transistor collector capacitance. As V_C becomes more negative than $-E$, the diode D conducts, ensuring that V_C swings between $V_{CE(sat)}$ (ideally 0V) and $-E$, so that the waveform contains a d.c. level V_o equal to $-(1-k)E$, and an a.c. rectangular wave with peak values $E(1-k)$ and $-kE$. The filter design must be such that the d.c. component of V_C appears across C, with $V_R = kE$ and $I_R = kE/R_L$, with negligible ripple in V_o. As k varies with input amplitude, then V_o varies, and the filter will eventually set an upper-frequency limit to the rate of variation, i.e. it will determine the signal bandwidth limit. Meanwhile, the a.c. rectangular wave at the switching frequency appears across L, so that the current through L consists of the d.c. output I_R (for k constant) and a triangular waveform with a rate of rise equal to $E(1-k)/L$ and rate of fall equal to Ek/L. The part of the cycle with rising amplitude is supplied via the transistor whilst it is conducting during period kT, so that I_C consists of a rectangular pulse plus a rising sawtooth component. The other portion of the inductor current waveform, a rectangular pulse with decreasing sawtooth top, is supplied via the diode during the period $(1-k)T$, the energy for this being drawn from the reducing magnetic field of the inductor. As the triangular component of the inductor current passes through capacitor C, the ripple voltage across this is exponential in form, and this in turn causes a ripple current of similar waveform through R_L, this being kept small by suitable design.

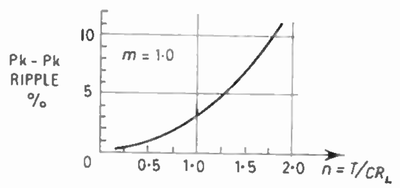
A complete analysis of the output circuit is complicated by the conflicting requirements of low ripple and adequate signal frequency response, coupled with a pulse-train with a duty-cycle and harmonic content varying at signal frequency. A simplified analysis, considering signal voltage and ripple current separately, is summarized in the Appendix and this enables satisfactory values of the filter components to be calculated. The parameters that are normally fixed are the required upper signal bandwidth ($\omega_2 = 2\pi f_2$), the load resistance R_L, and the period of the switching waveform (T, from the choice of output transistors). The parameter m, which is related to the Q of the filter circuit, has an optimum value of about unity (Fig. 6(*b*)), in which case the upper signal bandwidth is given by $\omega_2 = 1/CR_L$ (actually, -3 dB at $1·5/CR_L$). Hence C is defined to satisfy ω_2 and R_L at $m = 1$, and L is then given by $L = R_L/\omega_2$.



(*a*) Simplified filter circuit. (*b*) Amplitude—frequency response.



(*c*) Ripple current and voltage waveforms.



(*d*) Percentage ripple voltage against filter time-constant.

Fig. 6. Filter circuit characteristics.

The ripple voltage can be calculated by considering the response of the parallel CR_L combination to the sawtooth current at a duty-cycle of 0·5, which gives the maximum amplitude of the fundamental of the switching frequency, and hence the worst ripple condition. Figure 6 (*d*) shows the variation of the percentage peak-peak ripple voltage for different values of parameter n $(= T/CR_L)$, at $m = 1$, from which it can be seen that the percentage does not exceed 5% until $n = 1{\cdot}25$. Lower values of m give a proportionately larger ripple.

One other calculation needs to be made to ensure that the values of E, L, T and the current slope a do not mean that the peak current at the end of the sawtooth exceeds the transistor rating. At duty-cycle $k = 0{\cdot}5$, $E/2 = aL$, so that $a = E/2L$ and the peak current, including d.c. output and ripple, is

$$E/2R_L + ET/4L.$$

The minimum pulse length depends upon the turn-on, desaturation and turn-off times of the transistor, the total values of which are of the order of 30 µs with p-n-p germanium power transistors (f_α = 350 kc/s, I_c = 6 A), about 4 µs with n-p-n high-current silicon transistors (f_α = 1 Mc/s, I_C = 6 A), or less than 1 µs with intermediate-current n-p-n silicon transistors (f_α = 50–100 Mc/s, I_c = 0·5 A).

If the diode connection in Fig. 5 (*a*) is modified by adding a Zener diode of breakdown voltage bE, in series with D, then zero output voltage occurs at a duty cycle $k = b/(1+b)$. If b is therefore made about 0·1, non-linearity of output at small signals due to the minimum pulse length can be avoided, at the expense of losing 10% of the possible total VA.

5.3. *Bidirectional Unfiltered Output*

A bidirectional current through a single load resistor can be obtained using a single supply and a bridge of four switched transistors as in Fig. 7 (*a*), or twin supplies and two transistors, Fig. 7 (*b*). Considering the latter arrangement, transistor VT1 is switched on for period kT and VT2 is subsequently switched on for $(1-k)T$. The mean current through the load is $E(2k-1)/R_L$, which is obviously zero for $k = 0{\cdot}5$, with maximum values of $\pm E/R_L$. The transistors may be both p-n-p, both n-p-n or a complementary pair of p-n-p, n-p-n, the latter form allowing a simpler base driving circuit as both transistors switch in common-emitter mode, requiring only a small base voltage waveform. The complementary pair also has the advantage of a lower switching-off time, as the saturation base charge is extracted via the other collector, instead of only through its own base. The arrangement of Fig. 7 (*a*) is more economical in power supplies, at the expense of additional switching transistors, which must be switched in diagonal pairs. The mean current is the same, and the same combination of transistor types can be used.

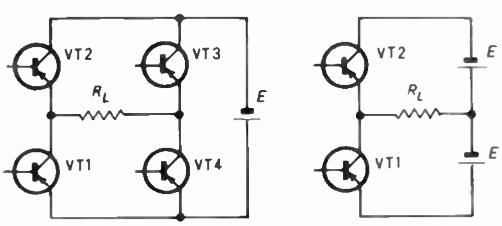
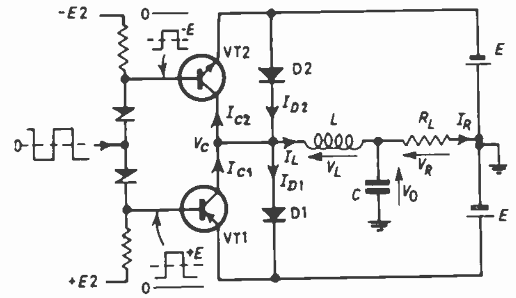


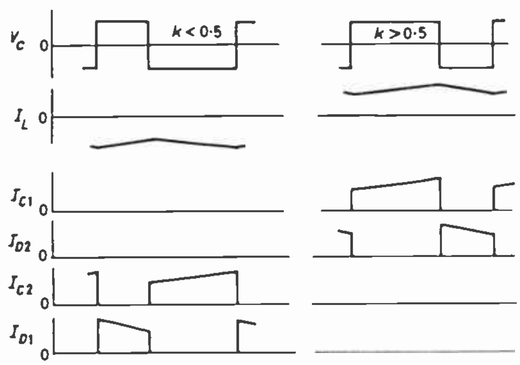
Fig. 7. Bidirectional unfiltered outputs.

5.4. *Bidirectional Filtered Output*

An inductor input filter is now added to the circuit of Fig. 7 (*b*), giving the arrangement shown in Fig. 8 (*a*). In this case a complementary pair of p-n-p and n-p-n transistors are used, requiring the base driving waveforms as shown, which can be produced from a single drive input via a Zener diode coupling arrangement to give the correct d.c. shift. The operation is similar to that of the unidirectional filtered output (Section 5.2), except that the switched transistors VT1, VT2 and diodes D1, D2 now ensure that V_C



(a) Circuit.



(b) Idealized waveforms.

Fig. 8. Bidirectional filtered output.

switches between $\pm E$, with a mean d.c. level $E(2k-1)$, which allows the output current to reverse at $k = 0{\cdot}5$ (Fig. 8(*b*)). For a positive output ($0{\cdot}5 < k < 1{\cdot}0$), VT1 conducts during period kT to provide a collector current pulse I_{C1} with a rising amplitude. During the remaining period $(1-k)T$, the reducing magnetic field of the inductor provides an induced e.m.f. to maintain the inductor current I_L in the same direction, around the loop D2, L, R_L and the upper power supply. The current passes into this supply against the source voltage, charging the internal smoothing capacitor. With a positive output normal conduction only takes place through VT1, VT2 making no contribution, so that it can in fact be omitted if D2 is in circuit. Similarly, VT1 is not essential for negative outputs. For $k = 0{\cdot}5$, with no d.c. output, the only power input balances the losses in the saturating transistors, the diodes and in the inductor. The smoothing capacitor C provides a low-impedance path for the saw-tooth current components and reduces load voltage ripple, but this is again not essential for the operation. The switching time of the transistors, governing the minimum pulse length and hence the minimum and maximum usable values of duty-cycle k now affects the regions of maximum positive and negative outputs, reducing the peak output but not affecting linearity at low signals, as in the case of the unidirectional output. Maximum load ripple voltage also occurs at zero signal output.

The transistors VT1 and VT2 need not be a complementary pair, but can both be p-n-p, in which case VT2 must switch as a common-collector stage. This necessitates a base voltage drive in anti-phase to that of VT1 and of a magnitude greater than $2E$. Thus when switching speed is not of prime importance the cost of the output transistors can be greatly reduced, particularly when a number are being used in parallel, but the base drive circuit consumes more power and may limit the choice of the final supply voltage.

In many cases the load R_L is not a pure resistance but contains a reactive component, either from a circuit element or via the motional impedance of an electrodynamic driving coil. Any reactance here does not affect the switching times, as it is bypassed by the smoothing capacitor at these frequencies. The load simply behaves as a normal impedance at the signal frequencies, with reduction in signal current as with a conventional power amplifier.

5.5. *Driver Stage*

The transformation from the output impedance at the collector of the bistable switch to the low source impedance and high base current requirements of the final output stage requires several stages of current amplification. This is best carried out by cascading emitter-followers, until the final stage can supply a sufficiently large base current. If two p-n-p transistors are employed in the circuit of Fig. 8, then the base voltage drive to VT2 exceeds $2E$, whereas that to VT1 need only be a few volts. This requires two separate driver amplifiers, one being a cascade of emitter-followers, whilst the other includes a grounded-emitter stage for voltage amplification, followed by an emitter-follower to obtain the low driving impedance.

6. Experimental Circuits

Several experimental circuits have been constructed along the lines detailed above, and the main performance features are as follows.

6.1. *Unidirectional Filtered Output*

This employs a switching frequency of 5 kc/s, and a switched output stage of four Ferranti ZT1488 silicon n-p-n power transistors in parallel, giving a maximum current capacity of 24 A. With a 24 V supply, an output power of 400 W is delivered into a load of 1·0 Ω, with a total collector power loss of 60 W and an inductor loss of 20 W. The collector efficiency is therefore 87%, but this would be increased to 94% if the supply voltage was increased to 50 V and the output power to 920 W, which is still well within the transistor ratings. The upper frequency bandwidth is 350 c/s, and the maximum ripple voltage is 3% of the peak output.

6.2. *Bidirectional Filtered Output*

The output stage consists of a pair of ZT1488 n-p-n transistors, one switching as a common-emitter circuit and the other as a common-collector. Output power supplies are ± 24 V, with a load resistance (including filter) of 6 Ω. The peak output power is 70 W, with a collector loss of 14 W, giving a collector efficiency of only 82%. As each transistor has to stand the peak-peak supply voltage, there is little margin left for increasing the working voltage, and substitution of transistors of low saturation resistance is the only means of increasing the efficiency. The upper frequency cut-off is 350 c/s, with a total harmonic distortion of less than 5%.

7. Conclusions

The principle of using pulse-width modulation of switched output transistors gives a d.c. coupled amplifier with an output stage efficiency of 85–95%. The actual values obtainable are very dependent upon the transistor saturation characteristics, and it is unfortunate that the use of high-speed silicon switching transistors leads to about three times the power loss of the low-speed germanium types. However, the power loss appropriate to maximum current rating can be dissipated by a single transistor.

The upper-frequency response of the amplifier is determined by the choice of switching frequency and the filter circuit, and an approximate analysis is given as an Appendix to enable a simple filter to be designed. With the switching frequency largely determined by the choice of output transistor types, a compromise has to be made between upper bandwidth and the ripple component in the load. Reasonable performance has been obtained to 350 c/s, with 3% ripple, at powers up to 400 W for a unidirectional output and 70 W for a bidirectional output. Higher power amplifiers and more complex filter circuits are at present being developed, and the bandwidth will almost certainly be increased considerably.

8. Acknowledgments

The author wishes to acknowledge the interest of Professor E. E. Zepler, of Southampton University, in whose Department most of the work was carried out and the assistance of Flt. Lt. G. R. Toule and Flt. Lt. E. A. Mansfield, who worked on the experimental circuits.

9. References

1. G. M. Ettinger and B. J. Cooper, "The design of high power transistor amplifiers", *Proc. Instn Elect. Engrs*, **106**, Pt. B, Supplements 15–18, p. 1285, 1959.
2. J. P. Eadon-Smith, "A direct current switching amplifier", *Electronic Engng*, **34**, p. 472, July 1962.
3. D. R. Birt, "Modulated pulse a.f. amplifiers", *Wireless World*, **69**, p. 76, February 1963.
4. D. A. Smith, "Transistors as switches", *Wireless World*, **69**, p. 195, April 1963.
5. R. F. C. Bennett, "A high power audio frequency push-pull transistor amplifier", *The Radio & Electronic Engineer* (*J. Brit.I.R.E.*), **26**, No. 3, p. 191, September 1963.
6. G. M. Ettinger, "Measurement of power loss in switching transistors", *Proc. Nat. Electronics Conf.*, **4**, p. 464, 1958.
7. B. Gilbert, "An emitter-timed monostable circuit", *Mullard Technical Communications*, No. 49, p. 345, July 1961.

10. Appendix

Output Filter Design, Unidirectional Output

The action of the switched output transistor, the inductor L, diode D and supply E (Fig. 5), is such that the collector voltage waveform V_C becomes rectangular, between ideal limits of $-E$ and 0, with a duty cycle k proportional to signal voltage. When k is varying at signal frequency ($f = \omega/2\pi$), a voltage V_s at this frequency can be considered to be applied to the input of the filter (Fig. 6(*a*)), together with the numerous harmonics of the switching frequency. The filter design must then be such that the total ripple voltage due to switching harmonics is suitably small, but that the component due to the signal voltage is not attenuated within the signal bandwidth of the amplifier.

Considering the response of the filter to a sinusoidal signal voltage V_s

$$V_o/V_s = 1/(1-\omega^2 LC + \mathrm{j}\omega L/R_L)$$

If $\omega_2 = 1/CR_L$, $x = \omega/\omega_2$ and a parameter m is introduced with $m = \omega_2 L/R_L$ then

$$V_o/V_s = [(1-mx^2)^2 + m^2x^2]^{-\frac{1}{2}}$$

This ratio is plotted on Fig. 6(*b*) to a base of x for several values of m, from which it can be seen that an optimum value of m is about unity, giving no attenuation of signal below ω_2 and a 3-dB fall at $1{\cdot}5\,\omega_2$. Values of m below 1·0 give some extension of bandwidth but add an appreciable resonant peak which would give an undesirable transient response.

The effect of the filter on the ripple current can be considered separately, as the requirement of a rectangular voltage across L means that the ripple current in L must always show a linear rise and fall. At a duty-cycle of 0·5, the current into the CR_L combination has the waveform shown in Fig. 6(*c*). Starting from point A, the input current is $-aT/4+at$ for $t > 0$, with a second contribution of $-2a(t-T/2)$ for $t > T/2$, and the voltage levels V_1 and V_2 must repeat over each complete period T. Considering the response to each input in turn, V_2 can be found to be:

$$V_2 = -V_1\exp(-T/2CR_L) - aTR_L \times \times[1-\exp(-T/2CR_L)]/4 + aR_LT/2 - aCR_L^2 \times \times[1-\exp(-T/2CR_L)]$$

Introducing parameter $n = T/CR_L$, V_1 and V_2 can be reduced to

$$V_1 = aCR_L^2[n/4-1-(1+n/4)\exp(-n)+ +2\exp(-n/2)]/[1-\exp(-n)]$$

$$V_2 = -V_1\exp(-n/2)+aCR_L^2\times \times[(1+n/4)\exp(-n/2)+n/4-1]$$

At a duty cycle of 0·5, the voltage across the inductor has the limits $\pm E/2$, so the rate of change of current a is determined by $E/2 = aL$, $aCR_L^2 = ER/2L\omega_0 = E/2m$. Therefore

$$V_1/E = [n/4-1-(1+n/4)\exp(-n)+2\exp(-n/2)]/ 2m[1-\exp(-n)]$$

and

$$V_2/E = -V_1\exp(-n/2)/E+ +[(1+n/4)\exp(-n/2)+n/4-1]/2m$$

The actual peak values of the voltage waveform do not occur at $-V_1$ and $+V_2$ except when there is negligible smoothing, (n is large), in which case the waveform remains triangular. When CR_L is large and n is small, only the fundamental component remains and the peak of the sinusoidal waveform is shifted by $T/4$. Normal requirements will lead to an intermediate case when n is near unity, in which case the time at which the peak voltage is reached has to be found by differentiating the voltage expression. This leads to:

$$t_1 = T\log_e[k + n/4 - 2mV_1/E]/n,$$

which in turn gives the value of the negative peak, $-V_3$, as $V_3/E = -V_1\exp(-t_1/CR_L)/E-(1+n/4)$ $[1-\exp(-t_1/CR_L)]/2m + t_1/2mCR_L$

As the waveform must be symmetrical, the peak-peak ripple is twice this value.

Figure 6 (*d*) shows the variation in peak-peak ripple voltage as a percentage of peak output voltage E for different values of $n (= T/CR_L)$ at one value of m.

Ripple can therefore be reduced by increasing m, at the expense of reducing bandwidth, or reducing n, which means switching faster if ω_2 is to remain unchanged. In the case of a bidirectional output, $k = 0{\cdot}5$ at zero signal and $E = aL$, so that the peak-peak ripple figures are double those above.

Manuscript first received by the Institution on 4th February 1963 and in final form on 28th October 1963. *(Paper No. 877/EA12)*

INSTITUTION NOTICES

New Year Honours

The Council of the Institution has congratulated the following members whose names appeared in the New Year Honours List.

John Norman Toothill, C.B.E. (Companion), is to receive the Honour of Knighthood for services to Scottish industry. Mr. Toothill is general manager of Ferranti Ltd. in Edinburgh; he was a member of the Institution's Council from 1962–1963.

Lieutenant-Colonel (now Colonel) Dennis Dibsdall (Associate Member) is to be appointed an Ordinary Officer of the Military Division of the Most Excellent Order of the British Empire. Col. Dibsdall was formerly with the Royal Corps of Signals and is now senior military advisor to the Ministry of Aviation at the Signals Research and Development Establishment, Christchurch, Hampshire. He served on the Examinations Committee of the Institution from May 1962 to October 1963.

Convention on "Advances in Automatic Control" April 1965

The Institution has accepted an invitation from the Institution of Mechanical Engineers to nominate a representative to join an Organizing Committee for a 2½-day convention which is to be held under the aegis of the United Kingdom Automation Council and has appointed Mr. J. A. Sargrove (Member). Other member bodies of the U.K.A.C. serving on the Organizing Committee are the Royal Aeronautical Society, the Institution of Chemical Engineers, the Institution of Electrical Engineers, the Society of Instrument Technology and the Institution of Production Engineers. At its first meeting the Organizing Committee agreed on a provisional programme for the Convention and now invites offers of papers.

The object of the Convention is to provide a forum for the presentation and discussion of the results of progress in research and development in automatic control and related subjects, including applications of computers in automatic control and the effects of advances in components on control techniques. After an invited address on the first evening, the second day will be devoted to the presentation of papers in three parallel groups of sessions. It is hoped that approximately eight papers will be presented in each group. The third day will be for the presentation of invited survey papers and general discussion sessions.

Offers of papers should be accompanied by brief synopses and sent not later than 27th March 1964 to the Secretary of the Institution of Mechanical Engineers, 1 Birdcage Walk, London, S.W.1.

The Secretary to Visit India

The Secretary of the Institution, Mr. G. D. Clifford, is to visit India again during January–February 1964. Mr. Clifford's last visit to India in April–May 1963 was concerned with the setting up of an Institution office in Bangalore, the inauguration of the Indian Divisional Council and the establishment of the Indian *Proceedings*. On this occasion he will be attending meetings of the Indian Divisional Council and will meet members in Delhi, Calcutta and Bangalore.

Mr. Clifford will leave London on 20th January and will be away for about three weeks.

International Conference on Magnetic Recording

An International Conference on Magnetic Recording will take place during the week beginning 6th July 1964, and will be held in the headquarters of the Institution of Electrical Engineers, Savoy Place, London. The Conference is being sponsored by the British Institution of Radio Engineers, the Institution of Electrical Engineers, and the United Kingdom and Eire Section of the European Region of the Institute of Electrical and Electronics Engineers.

The scope of the Conference will cover all methods of magnetic recording on moving media and include sessions on audio, video, digital and analogue recording techniques, recording media and general aspects.

Papers on these and associated topics, of a length not greater than 2,000 words, or equivalent including figures, are invited. Prospective authors should submit synopses of approximately 200 words by 29th February and full texts (3 copies) will be required by 30th April 1964.

A programme of social events and technical visits will be associated with the Conference and there will be an opportunity for items to be displayed in support of papers included in the programme.

Further details concerning the Conference and registration forms can be obtained from The International Conference on Magnetic Recording Secretariat, c/o The Institution of Electrical Engineers, Savoy Place, London, W.C.2, to which address synopses of papers and manuscripts should also be sent.

Binding of Volume 26

An Index to Volume 26 of *The Radio and Electronic Engineer* is sent out with this issue.

Members may have their half-yearly Volume (six issues) bound at 17s. per volume (postage and packing extra: 4s. Great Britain; 5s. Overseas). The issues and the appropriate indexes should be sent to the Publications Department, at 8–9 Bedford Square, London, W.C.1, with remittance.

U.D.C. 621.395.625.3: 681.32

Design Principles of the Magnetic Tape System for the Atlas Computer

By

D. B. G. EDWARDS, Ph.D., M.Sc.†

D. ASPINALL, Ph.D., M.Sc.†

AND

Professor T. KILBURN, D.Sc., M.A.†

Presented at the International Conference on "Signal Recording on Moving Magnetic Media" in Budapest on 15th–18th October 1962.

Summary: *Atlas* is a high speed computer typically performing a floating point addition in 1·4 μs. It has been designed to accommodate information transfers from eight tape channels simultaneously, information either passing into or out of the computer. Under these circumstances the computer has to deal with a transfer rate of $7{\cdot}2 \times 10^5$ 6-bit characters per second. A switching system interconnecting tape decks and channels permits more than eight tape decks to be provided on an *Atlas* computer.

After a brief survey of the *Atlas* computer the paper discusses tolerances on the packing density of the tape system, the layout of information on tape and the virtues of a fixed block length on tape. The program and automatic circuit techniques for dealing with the tape store are described briefly, and the system of checking tape transfers is outlined.

1. Introduction

1.1. *Magnetic Tape Requirements of the Atlas Computer*

The *Atlas* digital computing machine is capable of a computation rate of 5×10^5 operations per second, a high rate which demands extensive storage capacity. The main store of the *Atlas* at Manchester University is provided by 16 384 words of 48 digits on magnetic cores, backed by a magnetic drum containing 98 304 words. These two stores are presented to the programmer as one level of storage of 98 304 words.[1] The storage can be extended on other versions of the *Atlas* and augmented by large file drums or disc stores. There is still a need for a substantial bulk store which is provided for by a library of magnetic tapes. These tapes serve different purposes in the machine and in order to utilize them simultaneously there is a need for several parallel channels all of which can operate at the same time. Three channels are used to accommodate the operating system of the machine,[2] one to accept all the input data from various paper tape and card equipment, one to hold output information which is being fed to tape punches, parallel printers, etc., and the third to act as a 'dump' for programs in abeyance. Different problems require different amounts of computation and input/output time and on any one problem these times cannot be overlapped efficiently. The significance of the above operating procedure is to provide a delay in the system so that these times overlap over several problems.

Certain problems within the machine will have input data in the library of magnetic tapes which must be updated, an operation requiring an extra three channels; one for the old record, one for the updating information and one for the new tape. A further two channels are provided to transfer data between the library of tapes and the other peripheral equipments such as line printers or card readers. This makes a total of eight channels in all.

When the transfer from one deck is completed and its reel is being rewound, the channel can be freed to be used by another transferring tape deck. The switching between a channel and several decks is accomplished in a special switching unit, the cross points being transistor switches. The most usual arrangement is for two channels to be shared among eight tape decks allowing an installation to include up to 32 tape decks, as shown in Fig. 1.‡ This illustration

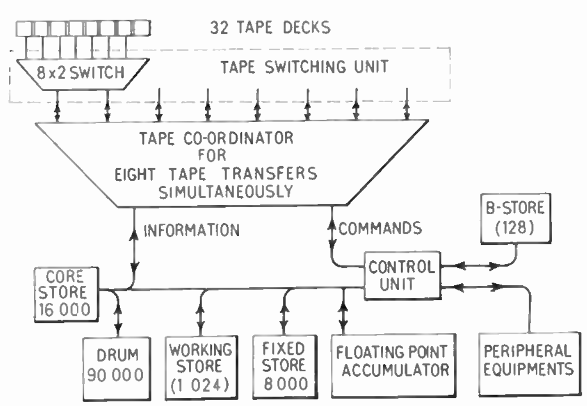


Fig. 1. Block schematic of the *Atlas* computer and the tape system.

‡ For economic reasons the Manchester University *Atlas* does not include this unit, and each of its eight channels has one tape deck.

† Electrical Engineering Laboratories, University of Manchester.

also shows the single information path between the main core store and the tape co-ordinator.

The co-ordinator organizes the transfer between the appropriate decks and this single path, and it also routes commands between the tape decks and the main control unit. This unit is shared between the tape co-ordinator and the working store of 1024 words, a fixed store of 8192 words, a main core store, a floating point accumulator and the peripheral equipments, all of which can be in operation at the same time.

1.2. *The Tape Deck*

The tape deck is based upon the 1-in version of the Ampex TM.2 tape transport (Fig. 2). The tape speed is 120 in per second and there are separate read and write heads for each of 16 tracks; a packing density of 375 bits per inch is achieved. The stop time is 1·5 milliseconds and the start time to within 10% of nominal speed is 2 milliseconds. A servo restriction forces a delay of 90 milliseconds to be included between a stop and a direction change command. The deck has been designed to contain the minimum amount of equipment, as much as possible being shared within the channel.

1.3. *The Tape Co-ordinator*

The nominal transfer rate between a tape deck and the co-ordinator is 90 000 characters per second and the total transfer rate between co-ordinator and main core store is 90 000 words per second. The packing or unpacking of characters is carried out in two words of buffer storage for each channel, which are contained in a core store of 16 words.

Words are transferred from the main core store as 48 information bits plus two parity bits. The parity is checked as each 12-bit stripe is written to tape (Fig. 3). During the write-to-tape transfer a check-sum is formed which is written to tape after the block of information. On reading from tape the check-sum of the block is formed and compared with that on tape; as each stripe is packed into the buffer store a parity circuit also operates to give each half word its own parity bit. These are checked as the 50-bit word is transferred to the main core store.

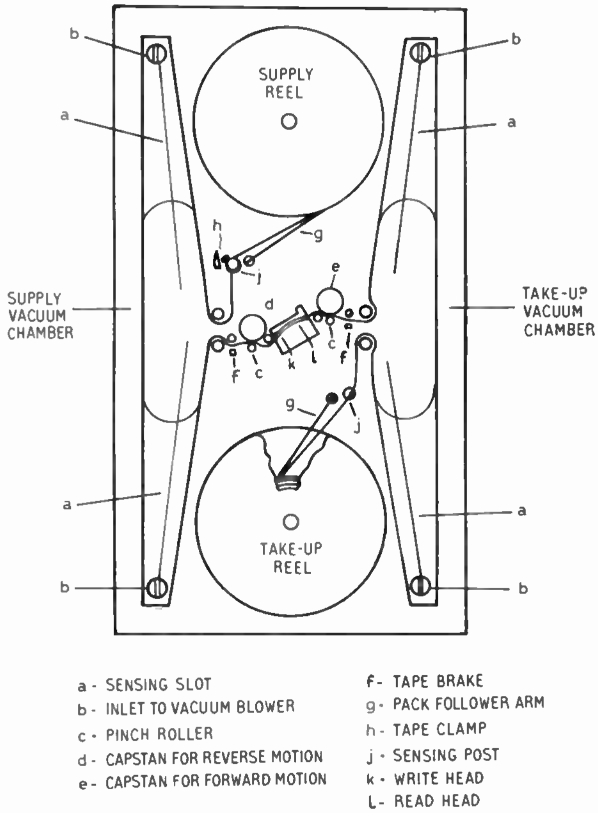


Fig. 2. The tape transport (Ampex TM.2).

The buffer store, parity and check-sum circuits are shared between all transferring channels in such a way that all channels may operate at one time, each channel either reading, writing or searching.

When a channel needs attention, for transfer of a stripe or word, it may have to wait until other channels are dealt with. The maximum time for which the channel can need attention and the transfer be held up, is termed crisis time. All transfers must take place at

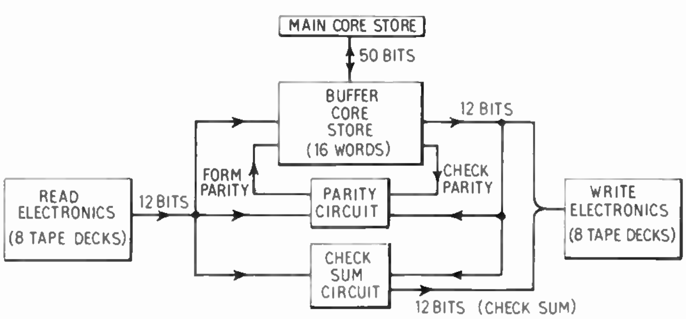


Fig. 3. Block diagram of the tape co-ordinator.

a high enough rate to ensure that not one transfer has to wait for longer than its crisis time. As each transfer takes place it is checked by circuits which also must operate at a sufficient rate.

1.4. *The Operating System*

Operations within the co-ordinator are controlled by its circuits once the block transfer has been initiated by a program in the central computer.

These programs come under the control of a master supervisor routine,[2] which is mainly contained within the fixed store. When a program instruction calls for a particular unit of data from tape the supervisor requests the operator to put the reel on a free deck. There is no need for the operator to select a particular deck or alter the address of a deck; as each new reel of tape is put on a deck, the supervisor reads the first block of information which contains the identifying factor. The supervisor keeps a directory of the contents of each deck and is able to search for data from tape as requested by the program. The transfer takes place autonomously including the separate parity checks on the transfers between core store and co-ordinator, and those between co-ordinator and magnetic tape are checked by a check-sum. If a fault is indicated the supervisor organizes a recovery process, for which it is convenient to read in both forward and reverse directions. The write-to-tape transfer takes place in the forward direction of tape motion so that the written data can be checked by a trailing read head. Data are written to tape blocks of fixed length. Each block has an address which is written during an initial testing and addressing operation.

2. Block Layout on Tape

2.1. *A Comparison of Fixed and Variable Block Length on Tape*

The aim of the design has been to give the programmer tape reliability throughout the life of a tape, without it being guaranteed drop-out free. The programmer prefers no restriction on the length of block on tape but finds certain limitations in practice. If the block is very short then the efficiency of tape use is poor, blocks less than 128 words being shorter than the interblock gap. The upper limit to block length is defined by the size of the main computer store. This could be 98 000 words on *Atlas*, but the tape is to be the means of communicating with smaller satellite computers with small main stores which cannot cope with blocks longer than 512 words.

If a block is written to a region of tape which is found to contain a permanent fault then the block must be rewritten further along the tape, wasting tape and time. The waste is proportional to the block length and it is economical to divide a large batch of data into blocks.

If variable block lengths are allowed then the same region of bad tape will be repeatedly encountered, leading to recurring waste. A fixed block length permits the tape to be pre-tested with its good regions marked and addressed. Tape which develops a fault can be taken out of service by erasing its marks and readdressing the remaining good tape. In this way the quality of tape used in transfers does not deteriorate with age, and time wasted in rectifying permanent tape faults does not increase.

2.2. *The Atlas Fixed Block*

The fixed block system makes it possible to rewrite new information over old without the risk of destroying good information in neighbouring blocks. A facility leading to the use of magnetic tape as an efficient backing store which is exploited for the internal organization of the computer by the supervisor routine. The main core store and drum both operate with fixed blocks of 512 words and the block on magnetic tape is fixed at this. Programmers are offered the variable block length facility through subroutines in the fixed store. The variable length blocks are assembled into fixed lengths by the supervisor routine.

When very short blocks are encountered these are first assembled and remain in a section of the core store until a full block of 512 words is available for transfer. This is slightly wasteful of core store space but is not thought to be significant.

Accurate location of data on tape is guaranteed by the system of block marks and permanent addresses. The address is checked as each transfer takes place. In particular the address is checked before write current is switched on in a write transfer to ensure that valuable data are not overwritten. The accurate transfer of data between the computer and magnetic tape is confirmed by means of a check-sum, which is written to tape as an additional half word at the end of the 512 word block.

2.3. *Layout of Tape*

The block layout is shown in Fig. 4. The separate read and write heads of the tape deck each have 16 tracks allocated as follows:

12 information
2 clock
1 block marker
1 reference marker

The information within the block is written as 2048 stripes of 12 digits. Each stripe is sub-divided into two 6-digit characters. Each character has a centrally located clock digit. The block addresses are written in the same way as information but distinguished by means of the mark on the separate block marker track. The mark consists of 13 reversals

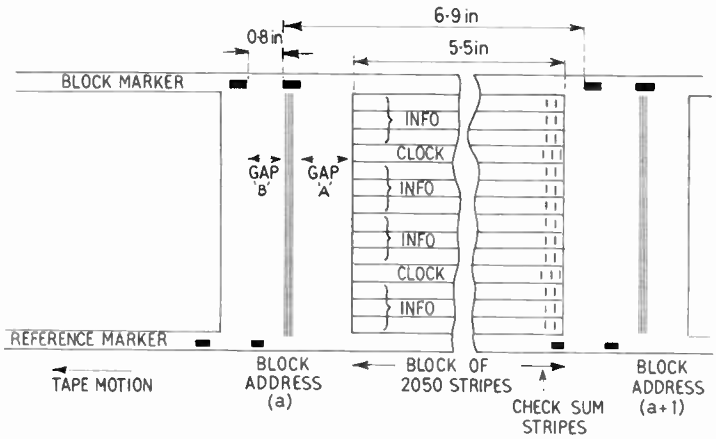


Fig. 4. Layout of magnetic tape.

of magnetization, three of which must be detected by an integrator circuit before the mark is indicated, to reduce the possibility of spurious signals on this track being mistaken for the mark. The marks themselves can be used for accurate location of information but it is necessary to have the additional check made by the address of each block.

The marks are protected from accidental erasure by switching out the relevant write amplifiers. It is not possible to protect the address in this way. In normal use the write current is switched on only when the tape is moving in the forward direction and the address has been checked. This accounts for a gap A of Fig. 4. It is made up of distance between read and write heads plus time to check the address plus time for which write current is turned on to produce a margin which erases previous data. The block of 2048 information stripes plus two check-sum stripes is then written to tape. A signal from the trailing block mark turns the write current off to ensure that the next address is not erased.

Thus, gap B of Fig. 4 between trailing block mark of one block and leading block mark of the next must be greater than 0·39 in, the distance between read and write heads. This gap length also depends upon the stop/start characteristics of the tape transport, which receives a stop command as a selected block mark is read. The tape is brought to rest so that the read head is in the inter-block gap and in such a position that on restart, in either direction, the tape is up to speed when a block mark is read. The gap B must not be smaller than twice the start distance. The start distance is greater than the stop distance and the stop command is given to the deck 1·5 milliseconds after the block mark has been sensed, so that the tape may be restarted in the opposite direction.

The manufacturers of the tape deck state that the maximum distance covered 2 milliseconds after a start command is 0·18 in. The tape has a tendency to flap away from the head at this distance unless the pinch roller and actuator assembly is carefully adjusted. It has been necessary to assume a safe start time of 3·5 milliseconds during which time the maximum tape movement is 0·380 in. This allows the tape transport adjustment to be less critical and reduces the amount of preventive maintenance.

The block marks exist not only to distinguish between address and information but also to stop the tape in the inter-block gap and define the regions of proved and tested tape which is available for information.

2.4. *The Addressing and Readdressing Process*

The block address is aligned in four stripes adjacent to the leading block mark. The mark and address are written during the initial addressing process and rewritten if the tape is readdressed. The location of the current region of tape to write a block mark is decided by the reference mark. During the initial addressing process this mark is written to tape and when it is read the leading write head puts the block mark to tape. The positioning of consecutive reference marks to define the inter-block and block lengths is decided by a crystal-controlled counter in the tape co-ordinator. In the readdressing process this mark is read and the write head puts the block mark to tape. If a block is faulty its reference marks are erased and no block marks can be written in the bad region when the tape is readdressed.

The addressing and readdressing processes are carried out on one of the channels of the tape co-ordinator under program control. This operation can proceed at the same time as transfers on the other seven channels. Each installation has its own mechanism for producing addressed tapes, which can operate when the computer tape system is not busy, making it independent of supplies from a distant source.

The tape is tested during the addressing process by writing a pattern of ones to tape in the region of the block. This pattern is checked by the trailing read head and a single fault condemns the block.

3. Write Electronics

Write current is turned on under program control after the block address has been checked. Even so valuable information may be overwritten under fault conditions. To prevent this a 'write permit' ring is supplied which must be added to the reel before its attachment to the deck to allow write current to occur. A reel which has just been written may require immediate protection. A 'write inhibit' button is provided on the deck which allows the operator to immobilize the write circuits by switching off the write power, even though the 'write permit' ring is present. Other logical interlocks are provided to prevent the write command being obeyed when tape is moving in the reverse direction or at fast rewind speed. Furthermore, every block mark signal resets the write command.

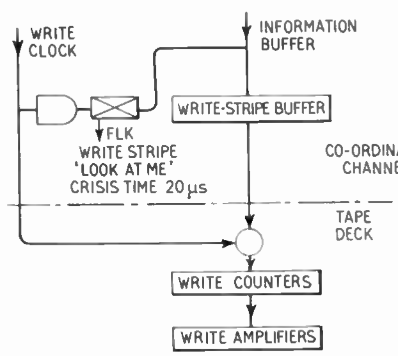


Fig. 5. Write electronics.

The write head has the following dimensions:

$$\text{gap width} = 0{\cdot}0005 \text{ in}$$
$$\text{track width} = 0{\cdot}031 \text{ in}$$

The write current needed to saturate the tape is 80 mA, but a write current of 120 mA is used to ensure adequate erasure of old information.

The digit representation on tape is modified non-return-to-zero in which each flux reversal represents the digit 1. This system is simple to use since a reversal of either polarity read from tape represents a 1. The write electronics must include a counter which changes state as it receives a pulse representing 1. The output of the counter governs the direction of current in the write head and is in the tape deck receiving signals from the distant tape co-ordinator as shown in Fig. 5.

Each channel has a write stripe buffer register which holds 12 digits. It is written from a word buffer store within the tape co-ordinator at a time when its output is not being gated. The write clock interrogates the output of the write stripe buffer every 22·2 μs, under the control of a crystal oscillator. This gating takes place in the tape deck so that all 12 write counters set at the same time, independent of differing cable delays between the co-ordinator and the deck. The write clock pulse is delayed by 2 μs to allow the write counters to set, before setting the write stripe 'look at me' flip-flop. This flip-flop indicates to the tape co-ordinator control circuits that this channel has just written a stripe to tape and demands the transfer of a new stripe into the write stripe buffer within 20 μs. This time is termed the 'crisis time' of the write stripe buffer.

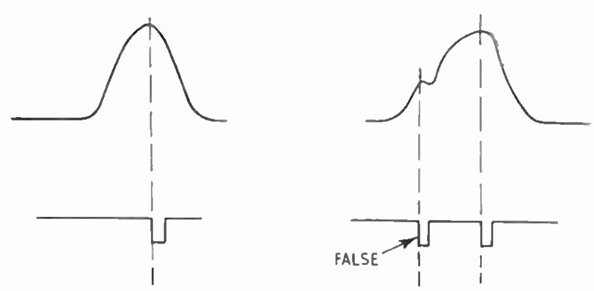
4. Reading from Tape

4.1. *The Detection of the Read Signal*

Each read head has the following dimensions:

$$\text{gap width} = 0{\cdot}00025 \text{ in}$$
$$\text{track width} = 0{\cdot}024 \text{ to } 0{\cdot}025 \text{ in}$$

It gives an output signal of approximately 15 mV peak to peak from 3 M's 499 magnetic tape. The amplitude will vary due to poor tracking, dust particles between the head and the tape causing the tape to lift off the head, and the tape flapping away from the head immediately after starting to move. The read signal is detected by a method which is independent of absolute amplitude, such as peak detection (Fig. 6(*a*)).



(*a*) Noise free. (*b*) Noise present.

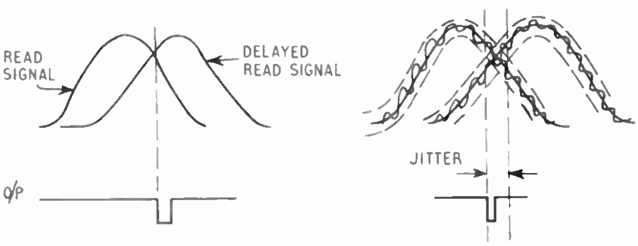
Fig. 6. Peak detection waveforms.

On installations which include a switching unit it is economic to put much equipment in the tape co-ordinator which can be shared between several decks. For this reason the detection circuits are in the tape co-ordinator and the tape deck contains simple circuits to amplify the read signals to a level which can be transmitted through a cable and switching unit.

Noise due to interference which is picked up by this cable causes distortion of the read signal. The high frequency noise superimposed upon the read signal can cause shifts in time of the peak and false signals as shown in Fig. 6(*b*). The peak detection circuit gives an output when the input voltage has zero rate of change. Noise can cancel the rate of change of the input signal and cause zero rate to occur and be detected away from the actual peak if the detector has a high-frequency response.

Thus the high-frequency noise must be removed by means of a low-pass filter. This measure seems to be adequate on the *Atlas* magnetic tape system where the maximum signal frequency is 45 kc/s. However, the magnetic drum system has a read signal of higher frequency (500 kc/s) closer to that of the noise (4 Mc/s) and a different detection scheme is used.

The principle of this angle detection scheme is illustrated in Fig. 7(*a*). A delayed version of the signal is subtracted from the signal itself and the point of zero difference is detected and used to indicate the presence of the signal. The detection point is a small angle away from the peak. It must be near the peak to reduce shift of the detection point due to waveform asymmetry. Noise causes shift of the peak but has much less effect upon the point of zero difference as illustrated in Fig. 7(*b*) and described in the Appendix.



(*a*) Noise free. (*b*) Noise present.

Fig. 7. Angle detection waveforms.

4.2. *Time Displacement of Read Signals*

The minimum time interval between reading stripes from tape is important in the design of the system and its calculation follows. When the tape is written its speed shortly after start may vary by $\pm 12\%$ from the nominal 120 in/s. The important case to consider for the purpose of this calculation is a slow speed on write of 105·6 in/s and a fast read at 134·4 in/s when the digits will be read at a nominal time interval of 17·4 µs. This time can be reduced by a combined jitter of write electronics and peak detection circuits accounting for a jitter of $\pm 1{\cdot}4$ µs, in the read signal. Thus the time between read signals on the same track can be as short as 14·6 µs.

The output signal from a peak detection circuit may be displaced in time from other signals of the same stripe due to jitter and also tape skew and variations in the mechanical construction of the recording heads. This displacement is greatest when the tape is written fast and read at slow speed on different decks with opposite variations.[3]

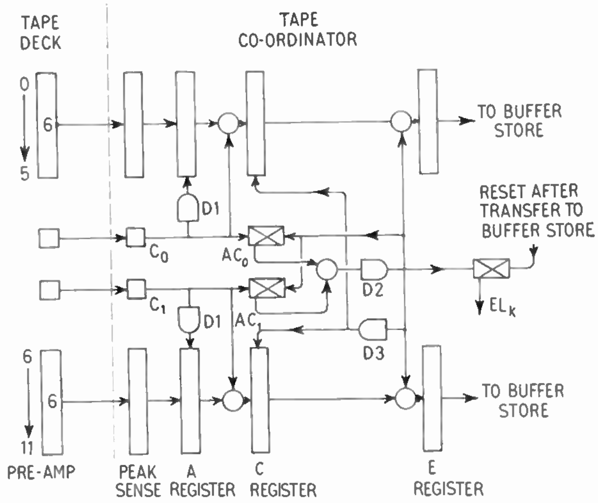


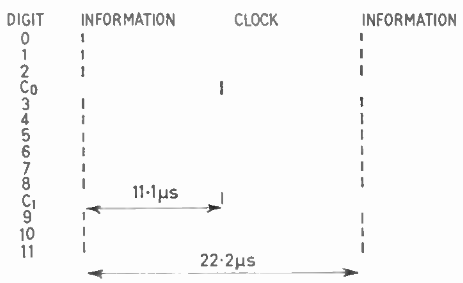
Fig. 8. Read electronics.

The time displacement of digits in the extreme tracks of a 6-digit character in the worst case is $\pm 5{\cdot}95$ µs about a mean time corresponding to an ideal clock signal written at the same time as the stripe. This time for the slow write/fast read extreme is $\pm 5{\cdot}05$ µs. Each read signal is stored on a flip-flop of the 'A' register (Fig. 8) until all digits of the character have been read; all 6 digits are then transferred simultaneously to a second register (the 'C' register) as shown in Fig. 8. It has been customary to use a delayed clock signal to strobe this transfer. If the clock signal is ideal the delay must be greater than 5·95 µs and less than $(17{\cdot}4-5{\cdot}05) = 12{\cdot}35$ µs to account for both extremes of tape speeds. The clock signal may be displaced in practice by $\pm 3{\cdot}05$ µs in the case of fast write/slow read, and $\pm 2{\cdot}9$ µs for the other extreme. Thus the delay must be greater than $(5{\cdot}95+3{\cdot}05) = 9{\cdot}0$ µs and less than $(12{\cdot}35-2{\cdot}9) = 9{\cdot}45$ µs requiring a delay circuit of 2% accuracy. An alternative solution is to write the clock digit halfway between stripes controlled by a crystal oscillator as shown in Fig. 9(*a*). This allows the direct output from the clock track peak detection circuit to be used as the strobe. The timing of this signal with respect to the digits of the character is shown in Fig. 9(*b*). It is seen that the earliest clock signal is later than the latest information signal and the latest clock signal is earlier by 0·75 µs than the information signals of the next character. This time increases for slower read speeds or faster write speeds.

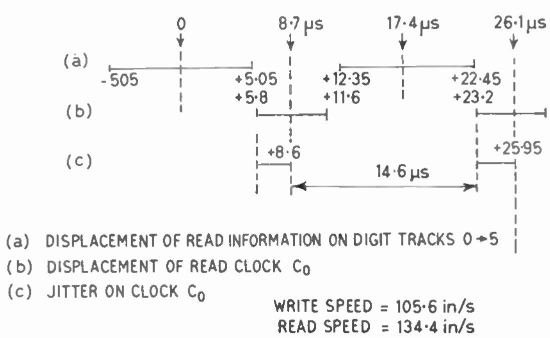
The jitter on the clock track causes the character to be transferred to the C register at intervals as short as 14·6 µs in which time the character must be cleared. There are 16 such registers in the tape co-ordinator and it is not possible to transfer from all of them in 14·6 µs without additional buffer storage. It is necessary to have a third register, the E register to which the stripe is transferred after the later of the two characters has been set in the C register, as shown in Fig. 8.

The clock signal C_0 sets a flip-flop AC_0 and transfers the character from the A to the C register. It is then delayed by D_1 to reset half the A register. This is performed by standard *Atlas* circuits and takes approximately 0·2 μs, much less than the 0·75 μs allowed. The output AC_0 is gated with a similar signal AC_1. The later clock pulse produces a signal from this gate which is delayed by D_2, to allow the C register to set and then transfers the twelve digits of C to E also resetting AC_0 and AC_1. After a further delay D_3 the C register is reset.

The flip-flop EL_k is set as the E register is set. It indicates to the control circuits of the tape co-ordinator that a stripe is ready for transfer to the buffer store; furthermore, the transfer must take place within 14·6 μs the crisis time before the next stripe is read into the E register.



(*a*) Timing of write signals.

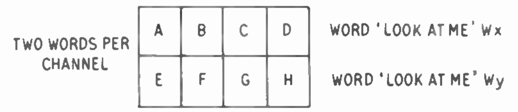


(*b*) Time displacement of read signals.

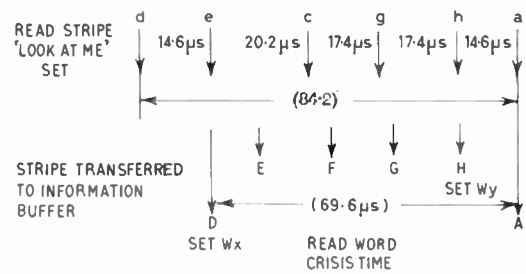
Fig. 9.

5. The Buffer Store

The buffer store consists of 16 words of 50 digits (48 digits for the information and two for parity) on magnetic cores[4] with a cycle time of 1 μs. Each channel has two words of storage and the stripes are packed as shown in Fig. 10. On reading from tape a word is filled from the stripe register and when full, a word 'look at me' is set. This word waits to be trans-



(*a*) Words in information buffer.



(*b*) Timing of read transfers: worst case.

Fig. 10. Read word crisis time.

ferred to the main core store whilst stripes are transferring into the second word of storage. Each word has its own 'look at me' so that for a short time two words can be waiting, as in the example shown in Fig. 10, when a tape of high packing density is being read at highest speed. The last stripe of word Wx is set into the E register at time d, assuming the tape co-ordinator control takes the longest permissible time to transfer this stripe it is not set into the buffer store and the word 'look at me' Wx set until time D, 14·6 μs later. The shortest time between d and setting the first stripe of Wx into the E register at a is $4 \times 17{\cdot}4 + 14{\cdot}6 = 84{\cdot}2$ μs as shown. The stripe may be transferred to the buffer immediately at A by which time the word must have been cleared to the main core store. The maximum permissible time between forming the word in the buffer and transferring to the main core store is $84{\cdot}2 - 14{\cdot}6 = 69{\cdot}6$ μs. This time is called the read word crisis time. The write word crisis time is much longer, being $(5 \times 22{\cdot}2 - T)$ when T is the longest time to deal with a write stripe crisis. If $T = 20$ μs then the write word crisis time is 91 μs.

6. Method of Time Division

The buffer core store is shared between eight channels. Its operations consist of stripe and word transfers which need not occur at the same time. Other operations which take place during a stripe transfer are: check on parity and formation of check-sum on writing to tape, and checking check-sum and formation of parity on reading from tape. These use circuits which are shared between the eight channels.

The method of sharing which is used is to divide time into discrete intervals each allocated to a particular channel. If the channel requires attention during this interval it is dealt with. If not, then the time is wasted. The duration of this time interval t can be estimated as follows: The worst case to consider is that of read transfers on all eight channels.

word crisis time $T_w = 69{\cdot}6\,\mu s$

stripe crisis time $T_s = 14{\cdot}6\,\mu s$

Then $$\frac{8t}{T_w} + \frac{8t}{T_s} < 1$$

Therefore $$t < 1{\cdot}5\,\mu s$$

This does not take into account the wasted time which reduces t to $1{\cdot}2\,\mu s$.

The core store has a cycle time of 1 μs, the adder circuit[5] used in the formation of the check-sum has an operation time of 0·2 μs and the parity circuit also operates in 0·2 μs. There is time enough for the operations to be completed within the time allowed and the synchronous system is relatively simple to commission and maintain.

Under fault conditions transfers can be missed. Stripe transfers which do not take place will be detected by the check-sum system, and missing words are discovered by a counting circuit. There is also a system to check that word transfers between co-ordinator and main core store do take place as requested, faulty main store transfers are detected by the parity circuits.

7. Check-sum

The check-sum consists of 24 binary digits. It is formed during the write transfer by pairing consecutive stripes into 24 digit half-words and adding them all up with an adder including end around carry. This sum is written to tape as two stripes after the block and also stored on a register in the tape co-ordinator. During the write transfer the trailing read head detects the stripes which are totalled to form a read check-sum. The check-sum achieved in this way is compared with the last two stripes read from tape and the stored check-sum. If all three agree then the transfer is successful.

8. Conclusions

The tape system will provide an efficient form of bulk storage and input/output medium for the *Atlas* computer. Its pre-addressed tapes of fixed block length will lead to little loss of computing time due to permanent tape faults. The time division method is simple to maintain. This has been achieved by working well within the limits of the circuit technique.

Though it is possible to achieve improved performance from the magnetic tape medium by using advanced circuit techniques, this could not be matched by the relatively poor reliability of the tape transport. It is important to improve the reliability of tape transports before the advanced performance can be exploited in a computer system.

The complete system has been in operation since December 1962 and is successfully completing its work load.

9. Acknowledgments

The authors wish to thank the Atlas Computer Group at Manchester University and Ferranti Ltd., for their assistance at all stages of this project.

10. References

1. T. Kilburn, D. B. G. Edwards, M. J. Lanigan and F. H. Sumner, "One-level storage system", *Trans. Inst. Radio Engrs (Electronic Computers)*, **EC–11**, p. 223, April 1962.
2. T. Kilburn, D. J. Howarth, R. B. Payne and F. H. Sumner, "The Manchester University *Atlas* operating system. Part I: internal organization", *Computer J.*, **4**, No. 3, p. 222, October 1961.
3. G. C. Ziman, "Maximum pulse-packing densities on magnetic tape", *Electronic Engng*, **34**, p. 521, August 1961.
4. D. B. G. Edwards, M. J. Lanigan and T. Kilburn, "Ferrite-core memory systems with rapid cycle times", *Proc. Instn Elect. Engrs*, **107B**, p. 585, November 1960. (I.E.E. Paper No. 3307M, November 1960.)
5. T. Kilburn, D. B. G. Edwards and D. Aspinall, "A parallel arithmetic unit using a saturated-transistor fast-carry circuit", *Proc. Instn Elect. Engrs*, **107B**, p. 573, November 1960. (I.E.E. Paper No. 3302M, November 1960.)

11. Appendix: Angle Detection

The aim of a peak detection circuit is to determine the zero-rate-of-change condition of the input waveform with a minimum amount of jitter. One important property of such a system is that amplitude variations in the input signal can occur without significantly affecting the detection time. However, if noise is present then deviations from the peak occur even for small amplitudes of noise. Furthermore, if the detection circuit has a high-frequency response then detection can occur even at positions significantly away from the peak (Fig. 6).

In the angle detection scheme a delayed version of the input signal is subtracted from itself. This can be done quite effectively using a delay line, and detection occurs when the output signal crosses the base line. This in fact will be at a time delayed from the peak. Now any high frequency whose period $T \ll T_d$ (the signal delay) can be biased off by the amplitude of the signal and is not magnified as in the peak detection scheme by its high rate.

To compare the peak detection and angle detection methods consider a system incorporating a peak detection circuit which has a jitter of $\pm T$ about the peak. Now consider that this is replaced by an angle detection circuit with a delay of $2T$. The waveforms are shown in Fig. 11. One half cycle of the read signal of amplitude E is shown and the signal delayed by $2T$. Each of these sinusoids has superimposed noise represented as a shift in the sinusoid of $\pm a$ shown by

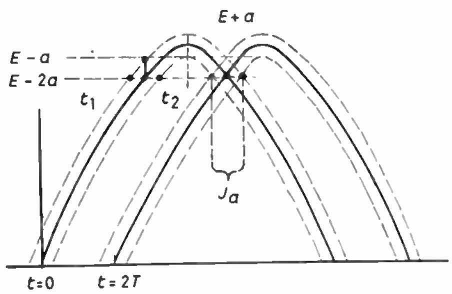


Fig. 11. Description of angle detection.

the dotted lines. The movement of the point of zero difference is shown as distance J_a.

Assume $\quad E\cos\omega T = E-2a \quad \ldots\ldots(1)$

Therefore $$\omega T = \cos^{-1}\left(1-\frac{2a}{E}\right)$$

Also $\quad E\sin\omega t_1 + a = E-2a$

Therefore $$\omega t_1 = \sin^{-1}\left(1-\frac{3a}{E}\right)$$

and $\quad E\sin\omega t_2 - a = E-2a$

Therefore $$\omega t_2 = \sin^{-1}\left(1-\frac{a}{E}\right)$$

But $\quad J_a = t_2 - t_1$

Therefore
$$J_a = \frac{1}{\omega}\sin^{-1}\left(1-\frac{a}{E}\right) - \sin^{-1}\left(1-\frac{3a}{E}\right)$$

Jitter due to peak detection circuit is J_P

where $\quad J_P = 2T$

Therefore
$$J_P = \frac{2}{\omega}\cos^{-1}\left(1-\frac{2a}{E}\right) \text{(substituting from (1))}$$

Improvement in reduction in jitter $= \dfrac{J_a}{J_P}$

$$\frac{J_a}{J_P} = \frac{\sin^{-1}\left(1-\frac{a}{E}\right) - \sin^{-1}\left(1-\frac{3a}{E}\right)}{2\cos^{-1}\left(1-\frac{2a}{E}\right)}$$

Assume $\quad a = \dfrac{E}{10}$

Therefore $\quad \dfrac{J_a}{J_P} = \dfrac{1}{3\cdot7}$

This improvement of approximately four to one has been confirmed in practice on the magnetic drum system.

Some noise is unavoidable in any big system where peak detectors are situated remotely from the actual storage mechanisms so that they can be shared between several switched systems. In these cases it is important to use angle detection, but in any case there is always this factor of improvement to be gained since an angle detection circuit is no more expensive than peak detection. In practice the longer the delay the bigger the improvement, but the circuit then becomes sensitive to asymmetry which can exist in the waveforms due to the digit patterns stored. The maximum delay is then a compromise between these two conditions.

Manuscript first received by the Institution on 15th August 1963 and in final form on 31st October 1963. (*Paper No. 878/C65.*)

# of current interest . . .

Microelectronics—First Steps to an Agreed Terminology

Welcome progress is reported towards an agreed terminology in microelectronics as a result of a meeting in Bad Kreuznach, Germany, of Technical Committee No. 47 (TC 47) of the International Electrotechnical Commission. Countries represented on the Working Group of the Committee dealing with this subject were Belgium, France, Germany, Italy, Japan, Norway, Netherlands, Sweden, United Kingdom and the U.S.A.

The following terminology was agreed:

Microelectronics: That entire body of technology which is associated with or applied to the realization of electronic circuits with a degree of miniaturization greater than that usually obtained with conventional methods and/or parts.

NOTE: Component densities greater than 3 per cm^3, or 50 per in^3, may be considered to be included in this category.

Microstructures: A structure of high component density the parts of which may be assembled or may be integrated and which for the purpose of commerce and specification is considered indivisible.

The term and definition agreed for the section of microstructures dealing with assemblies of discrete components was agreed as follows:

High-density Assembly: A microstructure in which the various components and devices are realized and tested separately before being assembled and packaged.

Integrated Circuit: A microstructure in which a number of circuit elements are inseparately associated on or within a continuous body.

NOTE: This definition covers circuits realized within one or more blocks of semiconductor and those obtained by deposition of thin layers or films or a combination of both.

It was further agreed that the integrated-circuit approach could be further subdivided and the following terms and definitions were agreed.

Thin-film Integrated Circuit: An integrated circuit composed entirely of thin-film circuit elements and interconnections deposited on supporting material.

Semiconductor Integrated Circuit: An integrated circuit composed of circuit elements and interconnections, the circuit elements being realized entirely within one or more blocks of semiconductor material.

The Working Group appreciated that there would be combinations of the above techniques and the following definitions of hybrid assemblies were agreed.

Hybrid Integrated Circuit: An integrated circuit using a combination of thin-film and semiconductor techniques.

Hybrid Microstructure: A microstructure consisting of one or more integrated circuits in combination with one or more discrete devices or components.

In spite of doubts that the Working Group may have proposed terms and definitions which may be too restrictive it was agreed that the terms and definitions as recorded above should be circulated as a Secretariat document for discussion by National Delegations. It was further agreed that examples of devices or assemblies in each section or area should be included for information.

The British delegation to the Committee was led by Mr. P. A. Fleming, technical secretary of the Electronic Valve and Semiconductor Manufacturers' Association.

U.K. Assists I.A.E.A. Training Course in Ceylon

The United Kingdom has provided £9,000 worth of modern nuclear electronic instruments and has made a Harwell scientist available for a period of four months to direct an International Atomic Energy Agency course on the repair and maintenance of nuclear electronic equipment which is opening in Ceylon on January 13. The course is being held as part of the United Nations Expanded Programme for Technical Assistance and it is designed to help alleviate the acute shortage of nuclear instrument repair and maintenance experts in that part of the world. The course will last six months and will train 20 participants from Africa and Asia.

The scientist concerned, Mr. K. Kandiah of the A.E.R.E. Harwell Electronics Division, will assist the host country, the Government of Ceylon, by directing the course for the Agency. Mr. Kandiah, who was born in Ceylon, is distinguished for his work on the design of electronic instruments for use in modern nuclear physics research; he was awarded an Institution Premium for a paper on counting tubes presented in the 1954 Brit.I.R.E. Convention. He is Group Leader of the Instrumentation for Physics Research Group in the Electronics Division.

The electronics instruments have been given partly by the Atomic Energy Authority, through the Office of the Minister for Science (approximately £7,500 worth) and partly by British industry (approximately £1,500 worth). The instruments include scalers, rate meters, power supply units, amplifiers and a wide range of related equipment and replacement components.

U.D.C. 621.396.96: 681.39

Digital Data Processing Considerations in Radar

By

P. J. CHILD, B.Sc.(Eng.)†

Presented at the Symposium on "Processing and Display of Radar Data" in London on 16th May 1963.

Summary: The paper describes various aspects of radar processing techniques which are being studied in order to determine their effect on overall system performance. Many data extraction systems make decisions regarding the presence or absence of targets by comparing the amplitude of the envelope of the demodulator output with a threshold which depends usually on the prevailing noise level. Studies have been made to determine if this is an optimum procedure, by considering other functions of the demodulator output to be used in the detection process. Energy per unit time is a possible alternative as a detection parameter. Statistical methods are used to compare performance, involving the determination of probability distribution of the parameter used in the detection process. The distribution of some functions cannot be obtained by normal analytical methods; a computer method is described which determines an approximation to the required distribution from which an estimate of performance can be made. The use of sampling theory is indicated in the study of optimum sampling of the decision parameter and comparisons are made between equipment and performance of systems using the conventional sampling rate and systems using the value indicated by the sampling theory. The effect on performance of a digital integrator of quantizing the decision parameter into two or more levels is studied and compared with results using analogue waveforms.

List of Symbols

B	Bandwidth of receiver (c/s).
T	Length of transmitted pulse (seconds).
$I_0(\ )$	Bessel function of first kind with imaginary argument.
$S^2/2$	Premodulator signal/noise ratio.
V	Envelope of demodulator input, normalized.
V_j	The amplitude of the jth (of a set of N to be integrated) output from a linear detector.
V_1, V_2	Envelope amplitudes τ seconds apart.
$p(V_1 V_2;\tau)$	Probability that V lies between V_1 and $V_1+\delta V_1$ at time t and between V_2 and $V_2+\delta V_2$ at time $t+\tau$.
$p(V)$	Probability that V lies between V and $V+\delta V$.
x	Dummy variable for generating function.
$K_1\ K_2\ K$	Thresholds used in production of a discrete probability distribution from a continuous distribution.
$W_1\ W_2$	Weights assigned to events occurring in the regions $K_1 \leqslant V \leqslant K_2$ and $K_2 \leqslant V \leqslant \infty$ in the production of a discrete probability distribution from a continuous distribution.
$\binom{i}{j}$	Binomial coefficient $\dfrac{i!}{j!(i-j)!}$
L	Second threshold in double threshold detection process.
N	Number of radar sweeps integrated.
$w(f)$	Power spectrum
a_n, b_n	Normally distributed random variables.

1. Introduction

Analogue/digital conversion of radar data is becoming attractive for many reasons. The intent of this paper is to discuss some of the alternative data processing methods that are available when the addition of an automatic digital detection system is to be made to a typical pulse radar equipment. It is postulated that such a radar system would have a bandwidth B c/s such that $BT \simeq 1$ where T is the length of the transmitted pulse in seconds. This is the well known compromise used in practice in order that the radar receiver channel is 'matched'[1] to the transmitted pulse.

† Electronic and Electrical Engineering Department, University of Birmingham.

Further it is assumed that the point at which digital data processing begins is at the output from the envelope demodulator in the receiver channel.

Of prime importance to the pulse radar equipment designer considering the use of digital data processing techniques is the effect on detection and position determination that such techniques will introduce. Range determination accuracy and resolution will be affected by the size of the basic range quantum used in the digital equipment. The size of this range quantum cannot be specified arbitrarily but is affected by the length of the r.f. pulse transmitted. The determination of azimuth depends on the effective beamwidth of the antenna and the p.r.f., and the manner in which the target echoes are processed. Any bias in the position estimates introduced by the digital techniques should be relatively independent of signal and/or noise level.

2. Methods for Comparing Performance of Proposed Digital Systems

In order to make a theoretical comparison between the detection efficiencies of different data processing systems the Neymann-Pearson observer is often assumed. This observer accepts a certain possibility of false alarm (i.e. that he will decide that a target is present when it is actually absent) and takes his observations in such a way that the probability of detection is maximized. It can be shown[2] that in order to implement these techniques it is sufficient to form a likelihood ratio. This likelihood ratio is the ratio of the joint probability that the samples of data upon which the decision of target presence or absence is being made, have come from the distribution of the signal plus noise to the joint probability that they belong to a noise distribution.

An alternative method used to compare the performance of different systems is to form the video signal/noise ratio.[1, 3] The mean video level in the presence of signal will increase and so it is often convenient to define a video signal/noise ratio as being the difference between the means of the distribution of signal plus noise and noise alone, divided by the standard deviation of the integrated noise alone. This method has to be used with care since this definition of a signal/noise ratio is not always very significant because a simple ratio cannot inherently specify the cumulative probabilities of two different distributions, for example, the distribution of signal plus noise and the distribution of noise alone at the demodulator output. However this definition of signal/noise ratio is useful in some cases.

When considering the design of a digital system it is often useful to compare its performance with the optimum system using analogue waveforms. Using likelihood ratio techniques outlined above it can be shown[2, 4] that the optimum procedure is to use a log $I_0(SV)$ envelope demodulator. If integration techniques are then used to enhance the probability of detection the detector is followed by a linear integrator. For very small signals ($S \ll 1$) simplifications can be made to the above specification so that the optimum detection procedure can be implemented by using a square law demodulator followed by an integrator which weights each V_j^2 by its expected power signal/noise ratio $S_j^2/2$. It is usually impossible to predict the expected signal/noise ratio for each measurement so that non-optimum weighting has to be used in practice. The difference in performance between optimum and non-optimum weighting is small so that Cooper[5] was able to construct an analogue detector 0·9 dB below the performance which would have been obtained using the optimum weighting specified by the likelihood ratio. This equipment comprised a reiterative loop integrator with exponential weighting. By the addition of an extra reiterative loop the performance was improved by a further 0·6 dB. Thus the analogue detector performance has been developed virtually to the maximum possible, without a tremendous degree of complexity. Practical difficulties arise due to imperfections in the delay lines used; also changes in p.r.f. are not easily made and temperature changes affect the transmission time in the delay lines and this has to be compensated.

3. Choice of Detection Parameter in Digital Detection Systems

In an analogue detector, the optimum method of processing is to form a square-law demodulator integrator combination. This uses a continuous waveform, (the envelope of the received signal plus noise waveform) as a detection parameter. In a digital detection system this envelope must also provide the information upon which the decision of target presence or absence and position must be made. Many systems use a time quantization equal to a pulse length and quantize in amplitude to one of two levels, dependent on whether the envelope is above or below a certain threshold in the time quantum under consideration. These systems count the number of times the threshold is crossed in a given time interval (usually the number of repetition periods that could possibly contain an echo from a given target) and produce a 'presence' signal each time this count exceeds a second threshold. Studies are being carried out to determine the performance of systems incorporating different types of sampling and sample utilization.

3.1. *Sampling Methods*

These methods define the manner in which the demodulator output is processed in order to produce inputs for the digital signal detector. The inputs to the signal detector are in the simplest case, a series of 0's or 1's. The 1's could, for instance, be generated using almost any combination of the following criteria:

(*a*) If at any time during a range quantum the demodulator output exceeds a threshold a '1' would be generated.

(*b*) If at any time during a range quantum the demodulator output *rises* through a threshold a '1' would be generated. This avoids the double detection that would occur using criterion (*a*) if the demodulator output rose above the threshold in one range quantum and did not fall below the threshold until the following range quantum.

(*c*) If the amplitude of a very short (much less than the transmitted pulse length) sample of the demodulator output exceeded a threshold a '1' would be generated.

(*d*) If the average energy of the demodulator output exceeded a threshold a '1' would be generated. The averaging period could be as long as or shorter than the basic range quantum.

The demodulator law has not been specified as yet. In practice it would probably be an approximate linear or square-law device depending to some extent on signal input level. In order to predict the performance using any of the criteria listed above, the probability distribution of the particular parameter being quantized must be known. In many cases it is impossible to determine an exact expression. For instance, predictions of performance for a system using criterion (*c*) can be made using an exact probability distribution due to Rice,[8] assuming that the samples are essentially statistically independent. The probability that the demodulator output will exceed a threshold at some time during a period $T > 0$ is obviously greater than the probability that the demodulator output will exceed the same threshold at an instant say at the beginning of this interval T. The exact probability distribution required to predict performance using criterion (*a*) is not known to the author, although the distribution function due to Rice[8] can probably be used without incurring a large error.

An approximation to the desired result may be obtained in several ways amongst them being the following:

(*a*) It is sometimes possible to calculate some of the moments of the required distribution. These moments can be used to form an expression approximating to the required distribution by using Hermite[6, 8] or Laguerre polynomials.[4, 6, 7]

(*b*) It is usually possible to measure the probability distribution experimentally. This method is however usually expensive in time and equipment. Also the equipment so built is usually only suitable for the measurement of a specific distribution.

(*c*) A more versatile method which has become very attractive with the advent of high-speed digital computers involves the specification of the parameter whose probability distribution is required in the form of a mathematical model. This method is described for a particular case in the Appendix. As before, the moments of the distribution can be determined and used in a general approximation.

A further alternative method of sampling the demodulator output is suggested by the Sampling Theorem, which states that any time varying waveform which is known to contain no frequency component above a frequency W c/s is defined by a set of samples taken at a rate greater than $2W$ samples per second. The receiver bandwidth B in a radar receiver is usually such that $BT > 1$ where T is the transmitted pulse length. This condition normally exists as a practical attempt at producing a matched filter.[1] This condition then restricts the spectral components of the demodulator input to the band 0 to B so that the Sampling Theorem requires samples to be taken at least twice per pulse length. The application of this double sampling method requires the knowledge of the joint probability distribution of the demodulator output. This can be shown to be

$$p(V_1 V_2;\tau) = \frac{V_1 V_2}{1-C_0^2}\exp - \left[\frac{V_1^2+V_2^2+2S^2(1-C_0)}{2(1-C_0^2)}\right] \times$$
$$\times \sum_{n=0}^{\infty} (2-\delta) I_n(a) I_n(b) I_n(c)$$

$$C_0 = \frac{\sin X}{X} \qquad X = B\pi\tau \qquad a = \frac{SV_1}{1+C_0}$$
$$b = \frac{SV_2}{1+C_0} \qquad C = \frac{C_0 V_1 V_2}{1-C_0^2} \qquad \delta = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

Using several approximations and assuming small predemodulator signal/noise ratios it can be shown,[9] using the likelihood ratio method, that the best way to use two adjacent samples of the demodulator waveform spaced τ seconds apart is to find the square of the sum of the amplitudes of each sample pair, to sum these quantities and then require that the sum exceeds a certain value, set by false alarm conditions. This may be achieved in an analogue system as illustrated in Fig. 1.

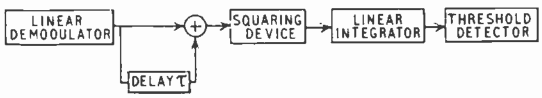


Fig. 1. Block schematic of a typical analogue system.

It is to be noted here that this result prescribes the use of a linear demodulator whereas the optimum

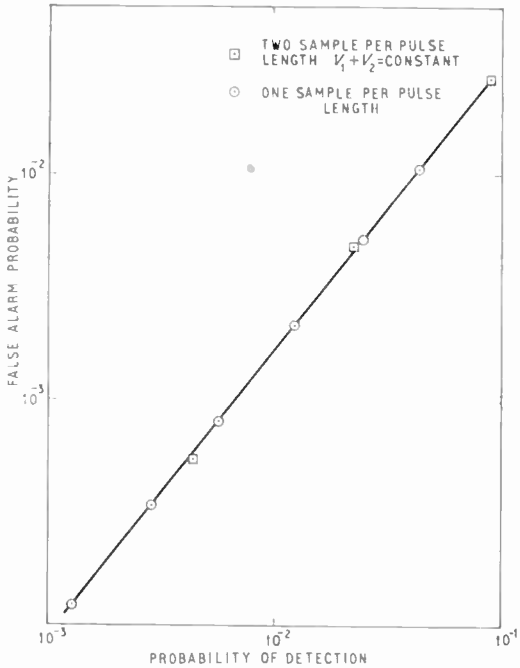


Fig. 2. Comparison of probabilities of detection by the conventional method and by the method suggested by the sampling theorem.

system using first order statistics prescribed a square law demodulator/integrator combination. In this latter case it is usual to assume the use of a square law demodulator followed by a linear integrator.

It is useful to imagine corresponding values of V_1, V_2 and $p(V_1 V_2; \tau)$ being plotted in orthogonal directions. The optimum procedure outlined above for analogue processing prescribes a threshold which is defined by V_1+V_2 = constant. For a system in which V_1 and V_2 are quantized, the optimum conditions cannot be achieved, but by introducing many quantizing levels it can be approached as closely as may be desired. These extra quantizing levels are expensive in terms of equipment and would not be warranted unless performance improvements were considerable.

It has been shown that in the analogue case the implementation of this optimum technique does not improve the performance by a significant amount compared with the commonly used system which omits the delay shown in Fig. 1. These results are shown in Fig. 2 in which the probabilities of detection obtainable by the conventional method (one sample per pulse) and the method suggested by the sampling theorem (two samples per pulse) are compared for similar false alarm rates. Detection on a single range sweep basis was assumed (i.e. $N = 1$) since the integrated probability distributions are not known in each case. It can be seen that the performance of the two systems is virtually identical over the range studied.

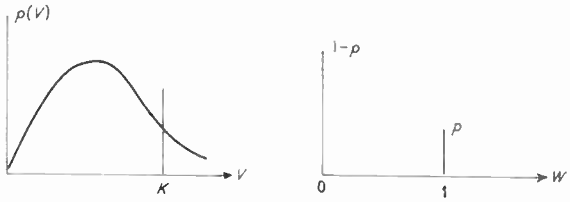
Also it has been shown[9] that by using the criterion that a target is considered present if V_1 and/or V_2 exceed a given threshold results in a degradation of performance of approx. 1·5 dB. This criterion would be easy to implement using digital techniques and is of course, essentially the result predicted using a single threshold and first order statistics. These estimates of performance have all assumed that the predemodulator signal/noise ratio is constant. In practice this is not true since the received pulse is stretched due to the receiver bandwidth. Thus if periodic instantaneous sampling of the demodulator output is to be used and a more accurate prediction of performance is required the probability distribution assuming a varying signal/noise ratio appropriate to the stretched pulse is required.

The presence of a target echo effects the video waveform over a period of approximately two pulse lengths. In practice the samples in the region would be considered to indicate the presence of a target if all or any of them exceeded a threshold.

4. Amplitude Quantization Effects

4.1. *Two Level Quantizing*

The amplitude of the decision parameter can be quantized into several levels. Most analyses have however assumed quantization into two levels. The effect of this quantization on the distribution can be seen in Fig. 3.



(*a*) Prior to quantizing. (*b*) After quantizing.

Fig. 3. Effect of two level quantization on distribution.

Figure 3(*a*) shows a typical probability distribution prior to quantizing. If a threshold is set at K and cumulative probabilities are calculated the distribution function that characterizes the quantized waveform consists of two lines of magnitude $1-p$ at zero and p at 1 where

$$p = \int_K^\infty p(V)\,\mathrm{d}V$$

Many digital detectors can be postulated (as will be described later). Let us at this time consider what is known as the sliding window detector which keeps a count of the number of 1's in the previous N radar repetition periods for a given range quantum. The probability that the count is L or greater is simply

$$\sum_{i=L}^{N} \binom{N}{i} p^i (1-p)^{N-i}$$

Then if p_0 is the probability that one sample from the noise waveform exceeds the threshold K and p_1 is the probability that one sample from the signal plus noise distribution exceeds K we have

$$\text{false alarm probability} = \sum_{i=L}^{N} \binom{N}{i} p_0^i (1-p_0)^{N-i}$$

$$\text{probability of detection} = \sum_{i=L}^{N} \binom{N}{i} p_1^i (1-p_1)^{N-i}$$

The problem now is to maximize the probability of detection for a given false alarm probability. This involves the optimization of L. This has been done[10] and in the usual range of interest i.e. $N \simeq 6$ to 10 the optimum value of L/N is $\simeq 0{\cdot}5$. We may now compare the performance of a digital integrator using sliding window method with the optimum analogue system. For $N = 6$ $L = 3$ and a false alarm rate of 10^{-8} and with a signal to noise ratio such that the probability of detection is 50% of the performance falls short of ideal by about 2 dB.

4.2. *Multi-level Quantizing*

As above, multi-level quantizing changes a continuous probability distribution to a discrete distribution. Figures 4(*a*) and (*b*) show the effect of quantizing to three levels. In these figures

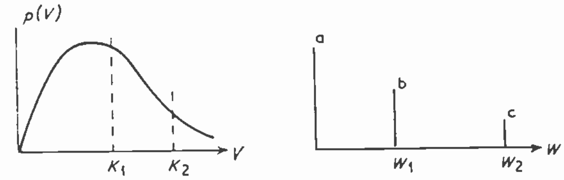
$$a = \int_0^{K_1} p(V)\,\mathrm{d}V \qquad b = \int_{K_1}^{K_2} p(V)\,\mathrm{d}V \qquad c = \int_{K_2}^{\infty} p(V)\,\mathrm{d}V$$

In the multi-level quantizing several degrees of freedom are available in that the relationship between the thresholds, in this case K_1 and K_2 and the weights assigned to events occurring in the regions

$$0 \leqslant V < K_1, \quad K_1 \leqslant V < K_2 \quad \text{and} \quad K_2 \leqslant V \leqslant \infty$$

has not been specified. It is not obvious how the optimum relationship could be determined analytically so that various possibilities are being tried, their choice being guided by considerations of resultant equipment economy and the results obtained for analogue data processing. For instance if $K_2 = rK_1$, then analogue processing considerations would suggest that $W_2 = r^2 W_1$, the exponent 2 being the power law of the optimum analogue demodulator integrator combination. On the other hand, assuming that the weights W_1 and W_2 would be summed as in a sliding window detector, the weights should be small so that the counter associated with each range quantum may be as small as possible.

The probability distribution of N samples taken from this discrete distribution may be found by standard methods theoretically. However, in practice an exact solution is only convenient for relatively low



(*a*) Prior to quantizing. (*b*) After quantizing.

Fig. 4. Multi-level quantization.

values of N. By forming the generating function for this discrete distribution defined by

$$g(x) = (a + bx^{W_1} + cx^{W_2})$$

the probability distribution for N samples can be found by forming $g^N(x)$ and extracting the coefficients of the different powers of x in the resultant expansion. Thus we may write

$$g^N(x) = \sum_{i=0}^{N} \sum_{j=0}^{i} \binom{N}{i} \binom{i}{j} a^{N-i} b^{i-j} c^j x^{W_1(i-j)+W_2 j}$$

The probability of the detector counter containing L after N samples have been taken from the distribution shown in Fig. 3(*b*) is then $p(L)$ where

$$p(L) = \sum_{i=0}^{N} \sum_{j=0}^{i} \binom{N}{i} \binom{i}{j} a^{N-i} b^{i-j} c^j$$

with $\qquad W_1(i-j) + W_2 j \geqslant L$

The extension of this method to the determination of the probability distributions after quantizing to more than three levels is obvious.

The evaluation of $p(L)$ and thus the determination of the cumulative probabilities can best be performed in a digital computer. These computations have been executed for the following cases:

(*a*)	$K_2 = 2K_1$	$W_1 = 1$	$W_2 = 4$
(*b*)	$K_2 = 2K_1$	$W_1 = 1$	$W_2 = 2$
(*c*)	$K_2 = \sqrt{2}K_1$	$W_1 = 1$	$W_2 = 2$

It may be noted that case (*c*) was an attempt to use optimum weighting (presumed to be square law from analogue data processing considerations) whilst at the same time minimizing the weights assigned to the different amplitude levels. These calculations were made assuming $N = 8$ and for comparison, similar calculations were made using the expressions relevant to two level quantization.

As in the case of two level quantizing an optimum second threshold was obtained. This optimum second threshold is not independent of signal/noise ratio (for a given false alarm probability) but increases as the signal to noise level increases, apparently between two limits. For very low signal/noise ratios, the probability that the decision parameter will exceed K_2 is low compared with the probability that it will exceed K_1.

Thus the maximum count is only a little greater than NW_1. The optimum second threshold under these conditions is, as might be expected, only a little higher than the value obtained by Swerling.[10] For higher signal/noise ratios the optimum second threshold increases and for sufficiently high ratios the count is primarily dependent on the weight assigned to crossings of K_2. Thus it appears that the upper limit is in the region of NW_2 multiplied by the optimum ratio determined by Swerling.[10]

The performance of a sliding window detector using case (*a*) and case (*b*) are very similar, the difference appearing to be less than 0·1 dB for the low signal/noise ratios considered (−12 to 2 dB). The method utilizing linear weighting, i.e. case (*b*), appears to be the best one to use in practice since the maximum of the curve of probability of detection versus second threshold for a given false alarm rate (here assumed to be 10^{-4}) is the flattest. Thus the degradation of performance through not using exactly optimum second threshold for a given signal/noise ratio would be minimized. Also the equipment necessary to implement the second threshold detector would be minimized owing to the lower counts required.

A comparison of these results with those for two level quantizing shows that, for similar false alarm rates, equal probability of detection can be achieved with approximately 0·5 dB smaller signal/noise ratio.

5. Digital Detector Implementation

A digital detector is required to accomplish several tasks. It must be sensitive and yet not produce too many false alarms. It should produce estimates of position of a target. These position estimates relate primarily to azimuth and range and will in general be biased. The bias of the estimates should be insensitive to changes in signal or noise level in order that they may be removed readily. The inputs to the detector are the samples of the decision parameter derived from the envelope demodulator output discussed above and it is obvious that many alternatives are available insofar as the use of these samples to determine presence and position of a target is concerned. Only a few of the many possibilities will be discussed. In general it will also be assumed that the way in which the video waveform is sampled will only produce a zero or a one, in order to facilitate the descriptions.

5.1. *Sliding Window Detector*

This is probably the most commonly used digital detector. Its function is to keep a tally of the number of successes (1's) in the previous N radar repetition periods for each range quantum. It may be implemented by having an N stage shift register for each range quantum the inputs to which are the 0's and 1's generated in the respective range quantum. Logic circuitry is required to monitor these shift registers in order to detect when any of these registers contains more than a specified count or second threshold. As mentioned earlier the first threshold is the reference used in the generation of the 0's and the 1's and is set by false alarm considerations. The optimum value of the second threshold is not influenced to any great extent by false alarm rate. Alternatively, counters could be used to replace the shift registers but in most cases the total amount of equipment involved is greater using counters than shift registers. Also the ability to estimate target centre is reduced. Azimuth estimation is accomplished by noting the time at which the count of 1's exceeds a target start threshold and the time at which the count falls through a target and threshold. This time difference defines the width of the target. The bias in this case is fairly large being approximately equal to half the target width.

5.2. *Coincidence Detector*

This detector produces an indication of target presence when 2 or more successive repetition periods produce a 1 in the same range quantum. This detector was studied by K. Endresen and R. Hedemark[12] who concluded that it was efficient in its ability to combat unsynchronized pulse-type interference, although its detection sensitivity was reduced by about 1 dB relative to the sliding window detector for a typical integration time. The bias and variance of target position estimate for this type of detector is being studied.

5.3. *Sequence Detector*

It is possible to design a detector that produces an indication of target presence based on a count of time between the occurrence of a given sequence (of 0's and 1's) or one of several sequencies. If the time between the occurrence of this (or any of several) sequences exceeds a threshold an indication of target presence is made and the counter is reset. Two such detectors are described by Dineen and Reed[13] who found by a computer simulation study that the azimuth bias produced by a detector of this type could be less than the sliding window detector. The variance of the estimate was approximately the same as for the sliding window detector.

5.4. *Sequential Detector*

An alternative method of using the quantized video waveform is to increase a count by W (dependent on quantizing level) when a 'success' is determined by the quantizer and to reduce the count by an amount R when a failure is determined by the quantizer. The value of R could be set in several ways, perhaps being controlled by average noise level. A 'target present' pulse is generated when the count exceeds a threshold, when the count is reset. Whenever the count falls below a negative threshold a 'noise present' pulse is generated and the counter is reset. Target width is measured from the 'target present' pulse to the first 'noise present' pulse following it. Dinneen and Reed[13] describe two possible detectors of this type.

5.5. *Digital Computer Utilization as a Digital Detector*

Additional position estimation accuracy can usually be obtained when, as is often the case, a general purpose digital computer is an integral part of the automatic detection and data processing equipment. Thus, if detections by the previously described detectors are called 'events', then all events occurring in a predetermined volume element in the space under surveillance by the radar can be correlated and the different components of the position estimates can be averaged.

In practice an estimation of range rate (velocity along line-of-sight) and time of occurrence of the events would also be made and used in this correlation technique.

6. Conclusions

The work described in this paper was carried out in an attempt to review methods of digital detection being used and to study in detail some aspects of the signal processing in order to determine if and where improvements can be made. The exact performance analysis of some systems in use at this time is impossible due to lack of knowledge of the appropriate probability distributions.

It has been shown that the performance of a sampling detector (using demodulator-output samples of vanishingly-small time duration) which samples twice per pulse length is not practically superior to a detector sampling once per pulse length if the signal/noise ratio is considered constant at all sampling instants. If the shape of the video pulse due to finite receiver bandwidth is considered, improvement may result from sampling more than once per pulse length.

The addition of an extra level in the amplitude quantization process appears to allow for a reduction in pre-modulator signal/noise ratio of approximately 0·5 dB. Little difference in performance due to the different weighting alternatives available is evident.

Initial estimates of the p.d.f. of envelope energy per unit time indicate that it should be advantageous (on the basis of a comparison of video signal/noise ratio) to use this quantity as a decision parameter rather than short samples of the output of a square-law demodulator in a digital detector. Quantitative estimates of performance are being obtained from improved estimates of the distributions. Also the effect of averaging time is being studied. A direct comparison with the system commonly used (i.e. a '1' occurs at quantizer output if at any time during a range quantum the demodulator output exceeds a threshold) is not possible at this time as the form of the necessary p.d.f. is not known. This may be obtained by the method described in the Appendix.

Of all the detectors considered, the sliding window appears to give the best compromise between performance requirement and equipment complexity. The recovery of the 1·5 dB to 2 dB loss due to the introduction of coarse (two level) quantization seems possible only with considerable additional complexity.

7. Acknowledgments

The author wishes to thank Dr. D. C. Cooper for helpful advice during the studies described above.

This work was made possible by support from the Ministry of Aviation.

8. References

1. J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals", M.I.T. Radiation Lab. Series, Vol. 24, (McGraw-Hill, New York, 1950).
2. C. W. Helstrom, "Statistical Theory of Signal Detection", (Pergamon Press, Oxford, 1960).
3. J. V. Harrington and T. F. Rogers, "Signal-to-noise improvement through integration in a storage tube", *Proc. Inst. Radio Engrs*, **38**, p. 1197, October 1950.
4. J. I. Marcum, "A statistical theory of target detection by pulsed radar", *Trans. Inst. Radio Engrs* (*Information Theory*), **IT-6**, No. 2, p. 59, April 1960.
5. D. C. Cooper, "Integration and Detection in Pulsed Radar and Sonar Systems", Ph.D. thesis, University of Birmingham, September 1961.
6. T. C. Fry, "Probability and its Engineering Uses", (D. Van Nostrand, New York, 1928).
7. H. Cramer, "Mathematical Methods of Statistics", (Princeton University Press, 1946).
8. S. O. Rice, "Mathematical analysis of random noise", *Bell Syst. Tech. J.*, **23**, p. 282, July 1944, and **24**, p. 46, January 1945.
9. P. J. Child, "Determination of Optimum Data Processing Techniques for use in Digital Detectors", M.Sc. qualifying thesis, University of Birmingham, December 1962.
10. P. Swerling, "Probability of detection for fluctuating targets", *Trans. Inst. Radio Engrs* (*Information Theory*), **IT-6**, No. 2, p. 269, April 1960.
11. D. C. Cooper and J. W. R. Griffiths, "Video integration in radar and sonar systems", *J. Brit.I.R.E.*, **21**, No. 5, p. 421, May 1961.
12. K. Endresen and R. Hedemark, "Coincidence techniques for radar receivers employing a double-threshold method of detection", *Proc. Inst. Radio Engrs*, **49**, No. 10, p. 1516, October 1961.
13. G. P. Dinneen and I. S. Reed, "An analysis of signal detection and location by digital methods", *Trans. Inst. Radio Engrs* (*Information Theory*), **IT-2**, No. 1, p. 29, March 1956.

9. Appendix

Determination of Approximations to Probability Distributions

The determination of the moments of probability distribution of energy per unit time of the output from a linear envelope demodulator will be considered as an example. The method to be described is however applicable to many other problems and can be considered to be an extension of the Monte Carlo method used in other fields of engineering.

In order that the method should be usable, the waveform to be studied must be formulated so that representative samples of it can be calculated. The output from a linear envelope demodulator for example may be derived as follows.

Let us consider a period of time length T, during which we observe a waveform which consists of a normally distributed noise voltage. This waveform may be analysed by Fourier methods and the amplitude $V(t)$ be written as

$$V(t) = \sum_{n=0}^{\infty} (a_n \cos \omega_n t + b_n \sin \omega_n t)$$

a_n, b_n are normally distributed random variables with a standard deviation of $\sqrt{(w(f_n)\Delta f)}$. $w(f_n)$ is the power spectrum of the noise voltage, i.e. $w(f_n)\Delta f$ is the average power which would be dissipated by those components of $V(t)$ which lie within the frequency band f to $f+\delta f$ if they were to appear across the terminals of a 1 Ω resistance. Since Δf is the frequency band associated with the nth component of the noise waveform, $w(f_n)\Delta f$ is the energy which would be dissipated if the voltage

$$a_n \cos \omega_n t + b_n \sin \omega_n t$$

were to be connected across 1 Ω, this average being taken over all possible values of a_n and b_n.

Thus

$$w(f_n) = \overline{a_n^2} \cos^2 \omega_n t + \overline{2a_n b_n} \sin \omega_n t \cos \omega_n t + \overline{b_n^2} \sin^2 \omega_n t$$
$$= \overline{a_n^2} = \overline{b_n^2}$$

Since a_n and b_n are independent random variables and thus the average $\overline{a_n b_n}$ vanishes. In the following discussion it will be assumed that $w(f)$ is equal to unity within a band of frequencies $(2M+1)\Delta\omega$ wide, centred on $\omega_m = m\Delta\omega$ and zero outside this band. This restriction is only made in order to obtain results for an ideal bandpass filter as was assumed earlier. The method can however, deal with any power spectrum whose shape can be defined in terms of analytic or statistical quantities.

With the above assumptions we may write after some rearrangement

$$V(t) = \sum_{n=-M+m}^{M+m} (a_n \cos \omega_n t + b_n \sin \omega_n t)$$
$$= X \cos \omega_m t + Y \sin \omega_n t$$

where
$$X = \sum_{n=-M}^{M} (a_n \cos n\Delta\omega t + b_n \sin n\Delta\omega t)$$
$$Y = \sum_{n=-M}^{M} (-a_n \sin n\Delta\omega t + b_n \cos n\Delta\omega t)$$

Now let a signal $S \cos \omega_m t$ be added to this noise waveform. We then obtain

$$V(t) = (S+X) \cos \omega_m t + Y \sin \omega_m t$$

The envelope of this waveform is given by

$$R(t) = \{(S+X)^2 + Y^2\}^{\frac{1}{2}}$$

Let us now consider a function $g(t)$ where

$$g(t) = f[R(t)]$$

We may study the statistical behaviour of $g(t)$ by determining its probability distribution. This can be done analytically for some simple functions of $R(t)$ but for the most part this is impossible.

An approximation to the probability distribution of $g(t)$ can be obtained by calculating (with t constant throughout) $g(t)$ for a large number of a_n's and b_n's. By classifying these samples of $g(t)$ a histogram approximating the distribution may be obtained. The method may be extended to determine estimates of joint probability distribution in an obvious manner, the limitation to the extension of this process being that the rate at which the number of calculations required increases as the order of the required probability distribution increases and/or the accuracy required of the approximation. In the case of first order probability distributions, the expressions defining the distributions can be determined by calculating the moments of the distribution from the histogram and using these to write down a suitable series in terms of either the Hermite polynomials[4,6,7] or Laguerre polynomials.[4,7] A slight improvement in moment estimate can be made in general by incorporating Sheppard's corrections.[6]

As an example, if

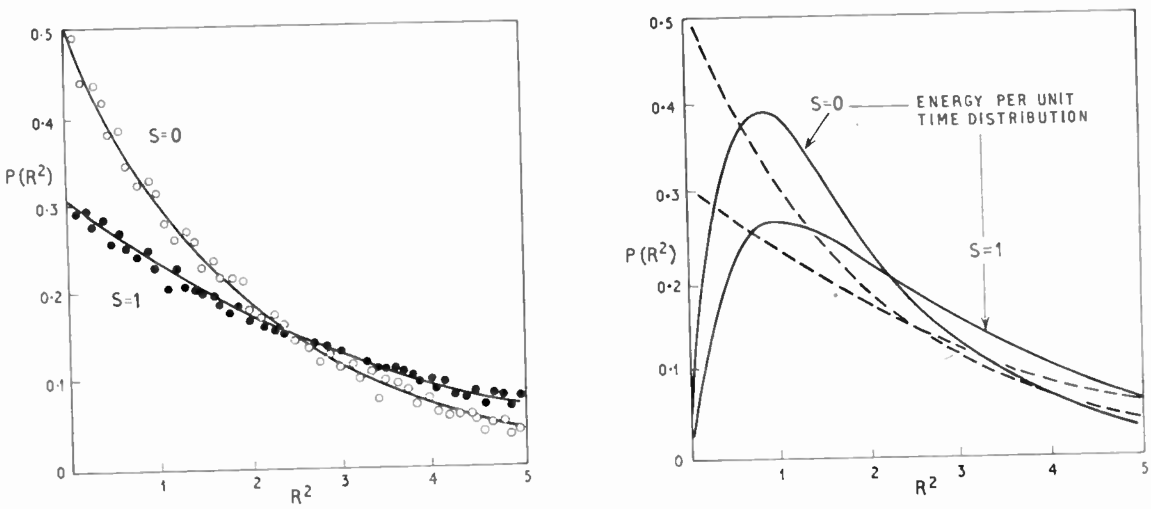
$$g(t) = \frac{1}{T} \int_0^T R^2(t)\,dt$$

i.e. energy per unit time in the envelope of signal plus noise, we may write

$$R^2(t) = \left[S + (2M+1)^{-\frac{1}{2}} \sum_{n=-M}^{M} (a_n \cos n\Delta\omega t + b_n \sin n\Delta\omega t)\right]^2 + \left[(2M+1)^{-\frac{1}{2}} \sum_{n=-M}^{M} (-a_n \sin n\Delta\omega t + b_n \cos n\Delta\omega t)\right]^2$$

The factor $(2M+1)^{-\frac{1}{2}}$ appears as it is convenient to generate the a_n's and b_n's with a standard deviation of unity. This factor then reduces the mean square amplitude of the noise to unity.

However a further rearrangement of $R^2(t)$ is preferred in order to save time in the calculation of the samples of $g(t)$. Thus it can be shown[9]



(*a*) Output from a square law envelope demodulator.

(*b*) Energy per unit time distribution.

Fig. 5. Probability distribution estimates.

$$R^2(t) = S^2 + 2S(2M+1)^{-\frac{1}{2}} \sum_{n=-M}^{M} (a_n \cos n\Delta\omega t + b_n \sin n\Delta\omega t) + (2M+1)^{-1} \sum_{n=-M}^{M} (a_n^2 + b_n^2) + $$
$$+ 2(2M+1)^{-1} \sum_{n=1}^{2M} \left\{ \cos n\Delta\omega t \sum_{i=-M}^{M} (a_i a_{i+n} + b_i b_{i+n}) + \sin n\Delta\omega t \sum_{i=-M}^{M} (a_i b_{i+n} - a_{i+n} b_i) \right\}$$

This expression can be integrated to give

$$g(t) = \frac{1}{T} \int_0^T R^2(t)\, dt$$
$$= S^2 + 2S(2M+1)^{-\frac{1}{2}} \sum_{n=-M}^{M} (a_n S_n + b_n C_n) + (2M+1)^{-1} \sum_{n=-M}^{M} (a_n^2 + b_n^2) +$$
$$+ 2(2M+1)^{-1} \sum_{n=1}^{2M} \left\{ S_n \sum_{i=-M}^{M} (a_i a_{i+n} + b_i b_{i+n}) + C_n \sum_{i=-M}^{M} (a_i b_{i+n} - a_{i+n} b_i) \right\}$$

where

$$S_n = \sin n\Delta\omega T/(n\Delta\omega T);\ C_n = (1 - \cos n\Delta\omega T)/(n\Delta\omega T)$$

As would be expected the expression for energy per unit time approaches the expression for $R^2(0)$ as $T \to 0$. The probability distribution estimates derived by this method for the output from a square law envelope demodulator i.e. $R^2(t)$ and for energy per unit time are shown in Fig. 5(*a*) and (*b*) where T is equal to the reciprocal of the bandwidth $(2M+1)\Delta\omega$ corresponding to the usual time quantization. The solid curve in Fig. 5(*a*) is the analytic form due to Rice, namely

$$p(R^2) = \tfrac{1}{2} \exp -\tfrac{1}{2}(R^2 + S^2)\, I_0(SR)$$

These curves are repeated for comparison purposes on Fig. 5(*b*). Ideally M would be very large. However for the curves shown M was made equal to five. Good agreement is apparent between the analytic and computer synthesized estimates for the distribution of $R^2(t)$. Similar good agreement has been found for other distributions, for instance that for $R(t)$ and also integrated distributions using this value of M. Also the values of a_n and b_n although generated with a standard deviation of unity were in fact generated in the interval -6 to $+6$ and not in the interval $-\infty$ to $+\infty$. This range could easily be extended. The computer time involved in the generation of these estimates is quite short, being only a few minutes in each case.

From an initial estimate of the moments of this p.d.f. of energy per unit time, it would appear that a slightly higher video signal/noise ratio can be achieved using energy as a decision parameter than by using the envelope directly in the form of short samples.

Manuscript received by the Institution on 11th April 1963. (Paper No. 879/RNA27).

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied. Information on translating services will be found in the Institution publication "Library Services and Technical Information".

LOGIC CIRCUITS

The general requirements to be fulfilled by logic circuits in large logic systems lead to an interesting discussion on the principles of diode-transistor logic circuits. A new type of diagram (p–n) is recommended as permitting a better assessment of the properties of the digital circuit than that obtainable from an on–off diagram. Numerical optimization with the aid of a computer has been performed in two cases—a d.t.l. circuit of NOR type with germanium semiconductors and of NAND type with silicon semiconductors.

"Optimal d.t.l. circuits", Ø. Gjessvag. *Ericsson Technics*, **19**, No. 1, pp. 57–88, 1963.

MAINS INTERFERENCE IN COAXIAL CABLES

New problems arise in connection with mains interference due to the introduction into communications engineering of miniature coaxial pairs with remote d.c. feeding. It is shown in a German paper that, in spite of isolated outer conductors, the currents induced in the inner conductors can assume values which are inadmissably high for transistor amplifiers. Theoretical considerations show that the extraneous currents in the inner conductors can be substantially reduced by the connection of inductors between the outer conductors and earth at a few points along the cable link if one fixes the magnitude of these inductors according to certain resonance principles. Practical results verifying this theory and methods of determination of the resonance inductors are given.

"A contribution to the protection of remotely-fed coaxial cable systems against mains interference", H. Seil and E. Widl. *Nachrichtentechnische Zeitschrift*, **16**, No. 11, pp. 569–577, November 1963.

COLOUR TELEVISION ENCODER

The electrical and mechanical realization of an encoder for a colour television system with quadrature modulation of the colour carrier components is described in a Czech paper. The transmission of the colour information is effected by difference components of the R—Y, B—Y type transmitted by equal bandwidths of 1·6 Mc/s and using one colour carrier 4·429687 Mc/s. The respective parts of the equipment are described and the results and parameters attained are given.

"An experimental encoder for a colour television system with quadrature modulation", M. Ptáček. *Slaboproudý Obzor*, **24**, No. 5, pp. 272–279, May 1963.

CORRELATION DEMODULATORS

Whenever a signal is transmitted through a non-ideal communication channel, the received signal will be somewhat different from the transmitted signal. This error may be brought about by spurious signals created within the channel (noise), or by distortion of the input signal, or by both. A frequently used criterion for the quality of a channel is the signal/noise ratio at the output. It is defined as the average signal power divided by the total average power of the spurious components. This criterion is particularly suitable in cases where an observer at the output of the channel has to decide whether a known signal is being transmitted or not. Another measure is the mean-square error of the output. It is defined as the average square of the difference between received and transmitted signals. This error criterion appears to be preferable when the observer knows that a signal is present and he has to decide what its amplitude, frequency, etc., are. Unlike the signal-to-noise ratio the mean-square error criterion is not a dimensionless 'figure of merit' in the sense that it is a direct measure for the performance of a single given communication channel. It is, however, valuable for comparing several channels under identical input conditions. Taking this criterion as a yardstick, the channel with the smallest mean-square error will be the 'best' channel. This principle has been applied to compare several methods of retrieving the intelligence from an amplitude-modulated carrier in the presence of noise.

"Mean-square error in correlation demodulators", B. F. Ludovici. *Archiv der Elektrischen Übertragung*, **17**, pp. 278–88, June 1963.

WIDE-BAND AERIAL FOR AIRCRAFT

Some theoretical investigations have been made by three Japanese engineers on the circuit properties of impedance-loaded folded unipole antennæ. As a result, a one-eighth wave broad-band blade antenna has been developed which covers the v.h.f. communication band of 118–144 Mc/s. The input voltage standing-wave ratio on a 50-ohm coaxial line is less than 2 : 1 over this frequency range. It has been designed especially for use on aircraft where low drag, light weight, high structural strength and broad-band impedance characteristics are required.

"A one-eighth-wave broad-band blade antenna", T. Kitsuregawa, Y. Takeichi and M. Mizusawa. *Mitsubishi Denki Laboratory Reports*, **4**, No. 2, pp. 189–98, April 1963.