ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

VOL. 18, No. 1, 1984

ATR is published twice a year (in May and November) by the Telecommunication Society of Australia. In addition special issues may be published.
ATR publishes papers relating to research into telecommunications in Australia.

CONTRIBUTIONS: The editors will be pleased to consider papers for publication. Contributions should be addressed to the Secretary, ATR, c/- Telecom Australia Research Laboratories, 770 Blackburn Rd., Clayton, Vic., 3168.

RESPONSIBILITY: The Society and the Board of Editors are not responsible for statements made or opinions expressed by authors of articles in this journal.

REPRINTING: Editors of other publications are welcome to use not more than one third of any article, provided that credit is given at the beginning or end as: ATR, the volume number, issue and date. Permission to reprint larger extracts or complete articles will normally be granted on application to the General Secretary of the Telecommunication Society of Australia.

SUBSCRIPTIONS: Subscriptions for ATR may be placed with the General Secretary, Telecommunication Society of Australia, Box 4050, G.P.O., Melbourne, Victoria, Australia, 3001. The subscription rates are detailed below. All rates are post free. Remittances should be made payable to the Telecommunication Society of Australia, in Australian currency and should yield the full amount free of any bank charges.

ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

ISSN 0001-2777

VOLUME 18, NUMBER 1, 1984

Contents

Challenge . . .

For most people, the mention of "1984" conjures up thoughts of "Big Brother is Watching You", "newspeak" and "double think", even if they have not read the book. The mention of "1984" does not evoke thoughts of the VIII Plenary Session of the CCITT to be held in October of this year. This is a pity because at the Plenary, various Recommendations relating to the evolving ISDN will be finalised, and this framework of standards will be the basis for the communication network to serve the information age.

The Orwellian totalitarian state required, for its maintenance, total control over communications, with ubiquitous two-way television and means to rewrite history. The technology available today with the synergy of communications and computers is equal to this task in a way which Orwell did not foresee. Remember how Orwell had teams of people sanitising written text by hand; imagine what that same team could do for newspeak and history with a user friendly word processor system.

Today we are striving for technical excellence in the setting of standards for the communications network of the information age; the challenge is to be equally excellent in setting social and political standards and safeguards for their use.

A.J. GIBBS

An Overview of the MELBA Automatic Code Generation Project

C.J. FIDGE
G.J. CAIN
L.N. JACKSON
R.S.V. PASCOE
Royal Melbourne Institute of Technology

A review of the history and current status of the MELBA automatic telecommunications programming project is provided. The MELBA system aims to automate the job of converting high level specifications into programming language code for a variety of telecommunications applications. The system thus frees the programmer from the burden of semantic translation, allowing its users to concentrate on design. The system is also ideally suited to rapid prototyping.

The present discussion is particularly timely since the MELBA project is presently nearing the end of a significant development phase which will see the system fully functional.

1. INTRODUCTION

This brief paper provides an overview of the MELBA research project, including a historical perspective and a description of the current status of the system.

MELBA is an ambitious project involving the design and implementation of a software system for automatically converting program specifications into compilable high-level language code for a wide variety of telecommunications applications, in effect performing the work normally done by a programmer. Although currently oriented towards telecommunications, the theoretical output from MELBA has applicability to the entire field of automatic code generation.

The MELBA project is currently in its fifth year. The project is based in the Department of Communication and Electronic Engineering, Royal Melbourne Institute of Technology, with assistance from the Department of Computing during the final two-year development period.

As an appendix to this paper a detailed example of the automatic code generation process illustrates some of the capabilities of the current version of the MELBA system.

2. BACKGROUND

The MELBA project began in earnest in 1979. Its aim was to allow automatic conversion of SDL specifications to compilable CHILL programming language code.

2.1 SDL and CHILL

SDL and CHILL are two standard international languages designed by the CCITT (International Telegraph and Telephone Consultative Committee, the telecommunications affiliate of the United Nations). Both languages were formally specified in 1980, however revised specifications are being prepared for 1984. Both languages are mainly intended for SPC (Stored Program Control) telecommunications switching applications, such as telephone exchanges and packet switching nodes.

SDL (Specification and Description Language) provides a graphical means of specifying the behaviour of telecommunications software. It can be used as a software development tool or as a documentation technique. Based primarily on state transition diagrams, SDL combines traditional flowcharting features such as operations (tasks) and decisions with the concept of signals, allowing communication and synchronization between concurrently executing program units. Each SDL diagram is used to describe a single concurrent process. SDL can be used in either of two forms: the more common graphical form or the equivalent program-like form (known as SDL/PR).

CHILL (the CCITT High Level Language) is a new programming language based primarily on Pascal, PL/I and Algol-68. Reviewing existing programming languages, the CCITT found that none were entirely suitable for SPC use and proceeded to develop a new language specifically for this purpose (precisely the same situation that prompted the U.S. Department of Defence to produce Ada for its embedded applications). CHILL is a language with capabilities very similar to Ada, however its syntax is vastly different (Refs. 26,24). CHILL differs from most of its predecessors mainly in the extensive concurrency features included, as well as in its modularisation constructs.

The advantages of a system such as MELBA can now be readily appreciated. SDL and CHILL are both standard languages developed by the same body with many conceptually similar features (the CCITT is emphasising compatibility between SDL and CHILL). Also both languages are (or will be) familiar to the majority of telecommunications engineers and will be widely used in the near future. If current trends continue, system development and documentation will commonly be performed using SDL while the system itself will be programmed using CHILL. A technique for automatically converting SDL specifications to CHILL code will greatly enhance programmer productivity and increase system reliability. Also documentation and code will always match since one is derived from the other.

2.2 Other Automatic Code Generation Systems

The overall aim of MELBA was therefore to allow a user to draw SDL diagrams on a graphics terminal and automatically produce a file containing the equivalent CHILL code. The impending definition of SDL and CHILL in the late seventies resulted in several groups throughout the world starting work on similar projects. In fact MELBA was inspired by SX1, one of the first of such projects and certainly the most experienced automatic code generation system known to the authors (Refs. 11,15).

The SX1 system represented the state-of-the-art when the MELBA project began in 1979. It used a textual form of input which was processed to produce an SDL-like diagram and PO-CORAL code. Nevertheless, despite being well developed and tested, SX1 (now known as CADOS) seems to be less ambitious than the current version of MELBA, since it requires the user to include target language statements in the graphical symbols. MELBA aims to be target language independent, the SDL user working purely with English-like statements.

Since the beginning of the MELBA effort a number of "rival" systems have surfaced including the "SDL to CHILL skeleton transformer" (Ref. 14). As the name implies, this system is capable of generating the outline of a CHILL program but cannot generate complete, compilable code. This system appears to be remarkably similar in aim to the 1979 version of MELBA.

Other software development systems based on SDL include the "System Generator for ESS" (Ref. 18) and CROSS (Ref. 17), both of which relieve much of the programmer's burden. They both use SDL-like input.

A number of automatic code generation systems for business applications have also appeared in recent years including the (heavily promoted) Last One, Pro IV and Pearl (Refs.19, 16). In general these systems are used to produce BASIC programs by linking a pre-defined set of subroutines together. Systems of this sort are often similar to aspects of MELBA, however they are restricted to business applications, just as MELBA is designed for telecommunications.

The main user interface to the MELBA system is a utility allowing the user to draw SDL diagrams at a graphics terminal by manipulating pre-defined symbols with a cross-hair cursor. A large number of automated SDL drafting tools with similar capabilities have been developed recently (Refs.21,22,23). Many of these provide drawing facilities much more sophisticated than those available in MELBA, however very few have considered the question of automatically converting the diagrams thus produced into actual code and none appear to have made serious inroads into this research area. The most sophisticated SDL drawing aid known to the authors is the one developed by Pierre de Chazal of the CSIRO Department of Computing Research, Canberra, Australia in 1981-82.

Finally, papers submitted to CCITT conferences often contain intriguing, but brief, mentions of SDL to CHILL translation systems developed by various organizations throughout the world, but since this is still very much a research area the details of such projects are often closely guarded. To the authors' knowledge the MELBA system, while often restrictive, is still the most ambitious automatic telecommunications code generation project under development. Certainly MELBA seems to be the only system to use the self-documenting qualities of SDL to full advantage. Nevertheless it should be emphasised that MELBA is still a research tool rather than a commercial product. User experience is very limited and as a result the practical viability of the system under industrial conditions remains to be proven.

3. EARLIER VERSIONS OF MELBA

The development of the MELBA system can be divided into three distinct phases. The first two, described in this section, resulted in the foundations of the SDL to CHILL conversion process and the development of some prototype software. The third phase, currently nearing completion, has seen the development of a complete code generation system and is discussed in the next section.

In 1979 the MELBA group performed a one year study of the problem, funded by the Telecom Australia Research Laboratories. This work established a correspondence between SDL constructs and features of the CHILL language (although minor semantic differences between similar concepts caused problems). Software was also developed (in BASIC and Pascal) to allow SDL diagrams to be drawn on a graphics terminal, which could then be converted into the SDL/PR form and into rudimentary CHILL code, providing the outline of the CHILL program but without complete detail; no data structure information could be generated and since the SDL diagrams only described one process at a time, each process was generated separately and needed to be manually linked together to produce a single, coherent CHILL program. Generation of SDL/PR as an intermediate language provides target language independent feedback to the users; error messages refer back to the original SDL diagram rather than statements in the CHILL code.

Encouraged by the results thus far, Telecom Australia funded a further two year study contract covering the period 1980-1981. The reason for the inability of the initial system to generate complete code was analyzed and found to have an obvious explanation: an SDL diagram simply does not contain as much information as the equivalent CHILL program implementing it. Techniques were required to supplement SDL to provide the additional input. The second phase of the project saw the development of the Module Structure Diagram (MSD). This was a new type of graphical input enabling the user to describe the block structure of a proposed system in a series of hierarchical diagrams. The position of processes could be

shown relative to the entire system and the communication between processes was also explained by illustrating the flow of signals in the system. Further software was produced to enable MSD diagrams to be drawn and converted into CHILL.

Thus the net result of the first two phases of the project was a system enabling the generation of CHILL code based on two forms of graphical input: MSD for describing the overall program structure and SDL for providing the detailed actions of each concurrently executing process. A methodology for users of the system was also adopted in which the development of a proposed system is under the control of a "system designer" who produces the MSD diagrams and therefore controls the overall system structure. The SDL diagrams for individual processes are relegated to "process designers" who draw SDL process specifications using the interfaces specified in the MSD's.

The system still had one large gap, however; it could not generate data structure information. A by-product of this problem was the inability to generate detailed statements for assignment, etc. since the MELBA system could not tell whether an identifier was meant to be an integer, character or even a more complicated structure such as an array. An attempt was made to incorporate data structure definitions in MSD's by adding some sort of Data Structure Diagram but the results were inconclusive and the solution to this problem was postponed until the next phase of the MELBA project.

4. MELBA: METHODOLOGY

The third MELBA study contract period covers 1982-1983, again funded by Telecom Australia. The final problem requiring a solution before the development of the complete system was the ability to acquire and use data structures.

4.1 Introduction of Data

Data structure information can be obtained where necessary from a library of data structures and operations manually coded and incorporated in the MELBA software system. The library approach was used since it relieves the process and system designers from the burden of defining data structures, something that is not normally done when using MSD and SDL. To increase reliability and stop this library from becoming a random collection of arbitrarily designed routines, MELBA adopts a very strict approach to the way library entries are stored. All information must be stored as specific, well-designed abstract data types. Further, the full advantages of the abstract type concept are used by allowing types to be parameterised (the same as a "generic package" in Ada (Ref. 25)). To the MSD or SDL user these types (from the "permanent" abstractions library) appear to be recognised by the system and can be used as if they were built-in types in a typical programming language. Their parameterisation allows them to act as building blocks for creating an even wider range of types. Strong typing is adhered to within the abstract type system. To achieve this each library entry consists of two main items: the data structure supporting the type and the operations that can be applied to objects of this type.

Although it is hoped that a thorough library of abstractions can be constructed it would be unrealistic to assume that this library, no matter how large, will be suitable for all MELBA user's needs. To overcome this it is possible for an "abstraction designer" to create a separate temporary library for any particular project under development (the "project" abstractions library) containing any special types required. This practice is frowned upon, however, and should be avoided if at all possible since the creation of abstract types is inherently complex and poorly designed types can lead to an unreliable software product.

4.2 Module Structure Diagram (MSD) Subset Used

MSD's were originally defined in 1980, when much of the MELBA system (and certain aspects of SDL) remained undefined. The Module Structure Diagram was critically reviewed in early 1983 and a subset was defined for use in the new system since the original MSD specification contained a number of features that were made redundant by subsequent developments in the MELBA project.

The MSD user (i.e. the system designer) can define the structure of a proposed system in a series of hierarchical diagrams consisting of modules and processes. Modules are a means of restricting visibility. Each module can be decomposed into further modules and/or processes. A process cannot be decomposed in an MSD diagram; its functions are described using SDL. Fig. 4 provides a typical example of an MSD.

Communication between processes is shown in MSD by drawing all signals that are sent to/from each process (solid lines in Fig. 4). These must correspond to the signals used in the SDL diagram for the particular process. Two processes in the same module can communicate directly but if a process needs to send a signal to a process in another module, the signal must go via a port. A port can be described as a hole in a module boundary through which signals can pass, while a channel is viewed as a uni-directional pipe used to link ports (dotted lines in Fig. 4). In this way processes that cannot "see" each other due to visibility restrictions caused by modules can still communicate if necessary by sending external signals (i.e. signals which travel through one or more ports).

4.3 SDL Subset Used

MELBA uses a subset of the 1980 CCITT SDL specification (recent developments from the CCITT have been examined but no attempt will be made to incorporate them until a new recommendation has been formally defined and approved in 1984). The optional "pictorial elements" are not used and the "functional block diagrams" (which as pointed out by L.M. Ericsson (Ref. 27) were rather ill-defined) are replaced by the Module Structure Diagram.

Apart from this there are two differences between MELBA's SDL subset and the 1980 standard:

(i) Signals in MELBA are always persistent (i.e. if a signal arrives at a process and is not immediately used it waits in a queue until required). The SDL standard was very vague in this area, however the "save" symbol (not used in MELBA) was introduced by the CCITT for when the SDL user desires persistent signals (Ref. 28). SDL signals are therefore non-persistent by default, i.e. if not immediately accepted upon arrival they disappear without trace. It was decided to make MELBA's signals persistent since it makes the implementation in CHILL much easier and is considered to be safer (Ref. 30).

(ii) MELBA allows a third type of signal apart from the usual external and internal types: the "encapsulated" signal. Encapsulated signals solve two problem areas that were both left undefined in the SDL standard; they can be used to implement signals that do not follow the usual semantic rules (e.g. conditional signal acceptance, continuous signals, signals with priorities, etc.) and they cover the concept of shared data between processes, solving the mutual exclusion problem that normally results from attempting to share variables between two or more process instances. Encapsulated signals were represented graphically by 3 bars on input and output symbols.

In fact all of the modifications required by the MELBA system are necessary to rigidly define areas that were not included in the 1980 CCITT SDL specification. All of the above problems are currently under review by the CCITT but until the 1984 CCITT SDL definition is available the MELBA system must provide its own solutions to these difficulties.

4.4 CHILL Generated

MELBA currently generates a subset of CHILL as output (many of the more esoteric features of the language are simply not required). Although the users of the MELBA system should not need to examine the generated code, it is still desirable to make the code as readable as possible. For this reason all identifiers used are left intact and there is a direct correspondence maintained between the MSD, SDL and library input and the CHILL output.

The code generated by MELBA also allows for potential differences in CHILL compilers wherever possible. For example, the CHILL Language Definition states that a CHILL implementation need not support the lower case alphabetical characters. Therefore a "switch" is available in the MELBA software to allow for the optional generation of code containing only upper case characters.

One feature of the generated code that has caused a surprising amount of debate is the use of goto's. Two SDL features, the "nextstate" and "join" are both converted into CHILL goto's and the authors were bemused when this approach was criticized by purists who argued that any use of goto's should be avoided. This technique is justified in the case of MELBA for four reasons:

(i) all goto's generated are jumps to user-defined names and are thus always self-documenting,

(ii) the nextstate and join in SDL are goto's at an intuitive level and it is only logical that they should be converted into the equivalent structure in CHILL,

(iii) an attempt to achieve the same results without the use of goto's results in convoluted code that is larger and less efficient than the equivalent goto implementation (this is done by Farnetani, *et al* (Ref. 14),

(iv) the goto's appear only in the generated code, and as such have no effect on the structure of the original system design (the designer does not examine the generated code).

4.5 Other Aspects of MELBA

One of the most important features of the system from the user's point of view is the self-documenting text employed in SDL. The SDL user actually works solely with high-level operations defined in the abstractions library, however the syntax for these operations has been defined so that it looks more like English than a procedure call. With the use of meaningful identifiers for local variables the process designer can draw SDL diagrams containing text that reads like natural language and is fully self-documenting. This conforms with SDL's main aim of being a specification rather than programming language.

Also, since MSD and SDL users work with types and operations available from the library, they do not need to be aware of the target language and in fact do not even need to know which language is being generated. Target language independence and self-documenting diagrams were both given high priority during the design of the MELBA system.

The three-level hierarchy of MELBA inputs is shown in Fig. 1.

5. MELBA: IMPLEMENTATION

The first two phases of the project saw prototype software developed in a number of languages including BASIC and Pascal on a variety of computers ranging from a micro to a very large mainframe. Unfortunately, the diversity of languages made the software difficult to maintain and hardware problems caused the system to be unreliable at times. At the beginning of the third phase of the project the group obtained access to new hardware, far more reliable than the original system. As a result it was decided to abandon the original software and develop a totally new system, this time written almost exclusively in ISO Pascal to give a degree of portability.

5.1 Hardware

The MELBA hardware configuration for the 1982-83 period is shown in Fig. 2. MSD and SDL

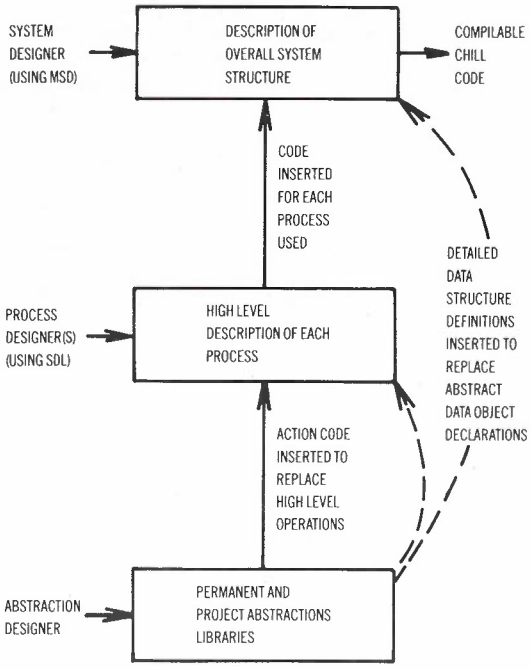


Fig. 1 - the hierarchy of MELBA inputs. Notice the way that gaps in high-level descriptions are filled with more detailed descriptions from lower levels. Items are inserted textually to produce the complete CHILL program.

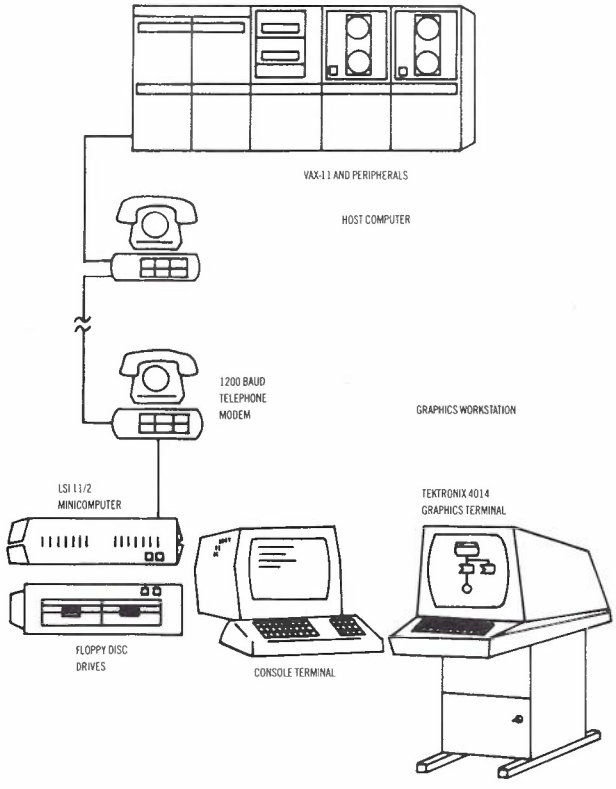


Fig. 2 - physical architecture of the MELBA system.

diagrams are prepared on a microprocessor and associated hardware forming a "graphics workstation". The principal hardware items in the workstation are a DEC compatible LSI 11/2 minicomputer and a Tektronix 4014 graphics display terminal. The minicomputer has 60 kilobytes of usable memory, dual double sided double density floppy disc drives, two serial I/O ports, one parallel I/O port and a Centronics type printer port. Software on the minicomputer is prepared and executed under the RT-11 operating system. The serial I/O ports are used for the console terminal and connection to a 1200 baud telephone modem. The parallel port is a DRV-11 interface and is used to provide a high speed connection to the Tektronix graphics display. All commands and responses to/from the system are via the console terminal. The 4014 is simply a facility for graphics output and cursor position input. The floppy disc system has a storage capacity of 1.25 megabytes for each 8 inch disc drive giving a total capacity of 2.5 megabytes. All disc transfers use DMA.

The 1200 baud modem is used to connect the workstation to a DEC VAX-11/780 mainframe where the off-line code generation programs reside. This is a 32-bit machine using the VMS operating system and is powerful enough to handle all of the larger programs in the system. Communications software developed on the LSI 11/2 allows the user at the graphics workstation to send files directly to the VAX, process them and receive the results, all in a single workstation session.

5.2 Software

As shown in Fig. 3, the MELBA software suite consists of seven main units. All code is written in Pascal. Overall the system design was divided into two areas to allow for independent (and concurrent) software development. The three software units on the left of Fig. 3 form an interactive graphics facility while the remaining items represent the off-line code generation programs. This structure has been useful in enabling swift software development despite the small size of the MELBA team.

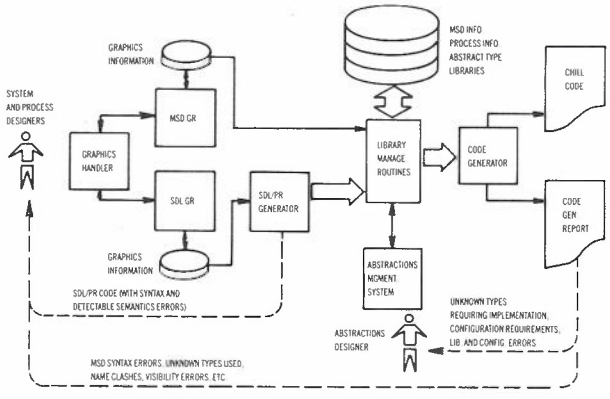


Fig. 3 - MELBA software structure.

The interactive graphics facility is supported by three programs: SDL.GR, MSD.GR and the Graphics Handler. The Graphics Handler implements a subset of the Core graphics standard (Ref. 29) and provides a sophisticated

interface between the other two graphics programs and the graphics terminal hardware. MSD.GR and SDL.GR manage the development of MSD and SDL diagrams respectively. They are both menu driven and easy to operate. Symbols are placed and connecting lines drawn by menu selection and manipulation of a cross-hair cursor on the graphics terminal. This facility provides the main user interface to the system. At the end of each MSD.GR or SDL.GR session a file is automatically produced describing the diagram created. This file is processed by the off-line programs to generate code.

The SDL/PR Generator accepts the intermediate description of an SDL diagram produced via the graphics software. Based on this it produces the equivalent SDL/PR description of the process. This listing will contain error messages for all syntax errors in the diagram as well as detectable semantic errors. This serves as immediate feedback to the process designer who can then modify the SDL diagram accordingly. When error free the SDL/PR code is saved for later processing by the Code Generator. The SDL/PR Generator also creates a cross reference for each diagram processed, enabling the Code Generator to perform various error checks without the need to re-examine the SDL/PR code.

The Library Management Routines are a collection of routines held in library form that give a standard interface between the applications programs and the file structures, thus helping to make the programs and files more modifiable.

The Abstractions Management System is a utility for the creation and modification of the abstractions libraries. It allows the so-called abstraction designer to create, enquire about, delete and modify library entries. Only privileged users may gain access to this utility. The Abstractions Management System is also used to configure a generated system for a particular application by specifying, for example, the maximum length of queues, depth of stacks, etc.

The Code Generator processes the complete project description to produce final CHILL code. This is the only target language dependent program in the system. It uses all stored files to construct and generate CHILL code. It also does complete error checking before the final code generation (listed in the Code Generation Report). Possible errors detected at this stage include problems with the MSD diagram, use of abstract types that do not exist in the library or misuse of existing types, identifier visibility errors or identifier clashes, etc. The program can also run in a checking mode where only individual items (e.g. MSD's, processes or library entries) are examined and reported on, without code generation. This enables feedback to the system users without the need to perform a complete code generation run.

Not shown on the system diagram are a number of small software units that provide support for the main programs in the system. These include the File Management Routines which isolate code that relies on non-standard Pascal (e.g. the use of random-access files) thus aiding system portability and the Communications Software which enables direct communication between the micro-computer and the mainframe used.

5.3 Current Status

At the time of writing (November 1983), five software units are basically complete: SDL.GR, the Graphics Handler, the Library Management Routines, Abstractions Management System and the SDL/PR Generator. All remaining software has been designed and is under development with an anticipated completion date of late 1983/early 1984.

6. CONCLUSION

This paper has presented an overview of the MELBA automatic code generation research project.

The end of 1983 will see the completion of the MELBA system. Following this it is expected that a one or two year study of the viability of the MELBA concepts under industrial conditions will be made, possibly resulting in a commercial product in a few years time.

7. ACKNOWLEDGEMENT

The authors would like to acknowledge the assistance and guidance of Telecom Australia Research Laboratories who funded the MELBA research programme.

8. REFERENCES

MELBA

1. Cain, G.J., Fidge, C.J., Jackson, L.N., and Pascoe, R.S.V.: "Computer-Aided CHILL Code Generation, Report No. 8", RMIT Research and Development Memorandum No. 112 028M, July 1982

2. Cain, G.J., Jackson, L.N., Vesetas, R., Walter, A., and Yong, W.B.: "Computer-Aided Software Generation (the MELBA system for generating CHILL code)", IEE SETSS Conference, University of Warwick, U.K., July 1981

3. Fidge, C.J., Jackson, L.N., and Pascoe, R.S.V. "Automatic Generation of CHILL Code From SDL Specifications - The MELBA Automatic Code Generation System", CCITT SDL Implementors Forum, Florence Italy, September 1982.

4. Jackson, L., Cain, G., and Yong, W.B.: "Automatic CHILL Code Generation (From CSTDs and Other Pictorial Inputs), Report No. 5", RMIT Research and Development Memorandum No. 112 024M, June 1981

SDL

5. CCITT: "Functional Specification and Description Language (SDL), Recommendations Z.101-Z.104", CCITT Yellow Book, Vol. VI.VII, Geneva, 1981

6. Gerrand, P.: "SDL and Its Use in Australia and Overseas, Part I: Introduction to SDL", Telecommunication Journal of Australia, Vol. 31, No. 3, June 1981

7. Gerrand, P., and Bierman, E.: "An Overview of SDL, the CCITT Specification and Description Language", ITU Telecommunication Journal, Vol. 49, V/1982

CHILL

8. CCITT Study Group XI: "Introduction to CHILL (May 1980)"

9. CCITT Study Group XI: "CHILL Language Definition" CCITT Recommendation Z.200, March 1982

10. Rekdal, K.: "CHILL - The Standard Language For Programming SPC Systems", Telecommunication Journal, Vol. 49, V/1982

Automatic code generation

11. British Telecom: "Computer-Aided Design for Software", British Telecom Research Laboratories Review 1980/81

12. Corker, M.J., and Coakley, F.P.: "Automatic Code Generation for SPC Call-Processing", IEE SETSS Conference, Salzburg, February 1976

13. D'Issernio, J.P.: "An Automatic Software Production for SPC Switching Systems", IEE SETSS Conference, Helsinki, June 1978

14. Farnetani, F., Giarratana, V., and Modesti, M.: "An SDL to CHILL Skeleton Transformer", CCITT SDL Implementors Forum, Florence Italy, September 1982

15. Hill, A.J.: "The SX1 Automatic Programming Project", SX 02 01, Issue 3, January 1978

16. Johnson, R.C.: "Hands-Off Production of User Software Gaining Ground", Electronics, April 1981

17. Kurosaki, T., Watanabe, T., and Oishi, T.: "Software Development Support System for Communication Systems", Hitachi Review, Vol. 31, No. 5, 1982

18. Mizuhara, N., Osamu, T., Hiyama, K., and Kurosaki, T.: "A System Generator for ESS Based on the Distributed State Transition Method", IEEE Globecom, 1982

19. Smith, K.: "Package Turns Software Specs into Fully Coded Bug-Free BASIC Program", Electronics, June 1981

20. Weaver, R.: "A Guide to the SX1 Graphics System Version 5", University of Essex Telecommunication System Group, Report No. 147, January 1978

SDL drafting tools

21. Gerrand, P.H., and Nguyen, K.D.: "CADDIE: A Computer Graphics Aid for the Behavioural Specification of New Communications Facilities for the Telecommunications Network", Australian Telecommunication Research Journal, Vol. 15, No. 1, 1981

22. Koning, H.: "CASA, a Computer Aid for SDL Applications", CCITT SDL Implementors Forum Contribution, Florence Italy, September 1982

23. Macalleno, G.: "Drawing SDL Diagrams Is Not So Easy ...", CCITT SDL Implementors Forum, Florence Italy, September 1982

Miscellaneous

24. Boute, R.T., and Jackson, M.I.: "A Joint Evaluation of the Programming Languages Ada and CHILL", IEE SETSS Conference, Warwick, U.K., July 1981

25. US Department of Defence: "Ada Reference Manual (July 1980)", Springer-Verlag, New York, 1981

26. Fidge, C.J., and Pascoe, R.S.V.: "A Comparison of the Concurrency Constructs and Module Facilities of CHILL and Ada", Australian Computer Journal, Vol. 15, No. 1, February 1983

27. L.M. Ericsson: "Views on SDL Related to Extension with Structural Concepts", CCITT Study Group XI Contribution, April 1981

28. L.M. Ericsson, Plessey Telecommunications, Swedish Administration, and Swiss PTT: "Persistent Signals", CCITT Working Party XI/3 Contribution DXI/3-107, Q.7/XI, Geneva, 21 June-1 July, 1982

29. Michener, J., and Van Dam, A.: "A Functional Overview of the Core System with glossary", Computing Surveys, Vol. 10, No. 4, December 1978

30. Swiss PTT Administration: "Considerations on the Safeness of SDL", CCITT Study Group XI Contribution No. 99, Q.7/XI, September 1981

31. Tanenbaum, A.S.: "Computer Networks", Prentice-Hall, 1981

APPENDIX

EXAMPLE

This example is based on the simple "par" data link protocol from Tanenbaum (Ref. 31). Although it makes some unrealistic assumptions it was chosen since it is reasonably small and illustrates the major features of MELBA.

The aim of the project is to specify and implement the ISO data link layer connection between two adjacent network nodes.

The first step is for the system designer to specify the overall system structure. Part of the MSD diagram to do this is shown in Fig. 4.

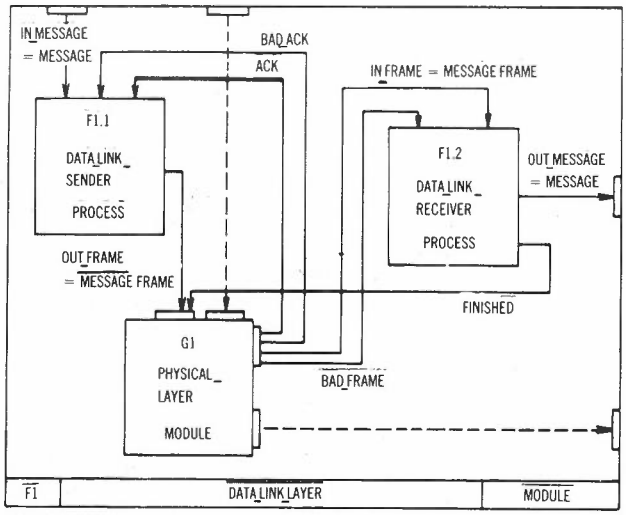


Fig. 4 - MSD diagram for the data link layer protocol. Details of the "higher" network layer and "lower" physical layer protocol would be provided on accompanying diagrams in a complete system.

Drawn using the MSD.GR graphics utility this diagram shows that at the data link level there are two concurrent processes, a sender and a receiver (we have made our protocol full-duplex rather than Tanenbaum's original simplex model). The details of the physical layer protocol are not visible in the current diagram, but could be made so by "exploding" G1. Lines connecting processes and ports represent signals. Some signals carry information, for example the signal in_message carries an object of type message. Notice that "messages" are converted into "frames" by the data link sender and transmitted to the physical layer (and vice versa for receiving). Finally, there are two channels attached to the physical layer module. Signals may pass through these channels but are not relevant to the data link layer and therefore cannot be seen from within module F1.

Having edited the MSD diagram until satisfied, the system designer then assigns the task of implementing the individual processes to one or more process designers. We will examine the sending process. The process designer first studies the MSD diagram to determine the interfaces that must be used. Then the SDL.GR graphics utility is employed to create the SDL diagram to implement the process as shown in Fig. 5. Again this diagram would be iteratively modified until correct (the process designer may even request a change in the interface from the system designer). Notice the use of English-like text in the diagram. The various statements in the graphical boxes do not imply any particular target language. The output could be CHILL, Ada or any other suitable HLL with no change to the SDL (or MSD) diagrams. Also note how the signals used correspond to the signals in the MSD diagrams. The process operates by reading a message into "buffer", creating a frame "s" using the message by attaching a sequence number, transmitting the frame to the physical layer and waiting for an acknowledgement or timeout. Of special interest

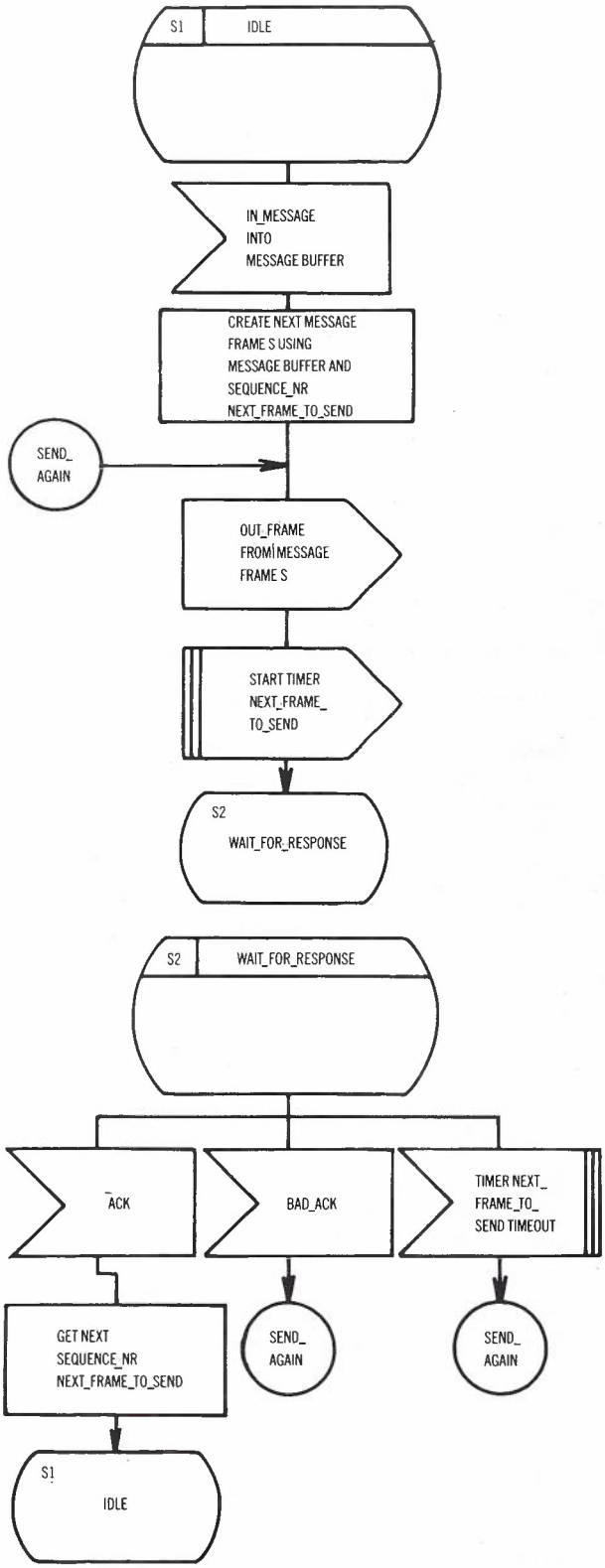


Fig. 5 - SDL diagram for the data link sender process.

here is the use of an encapsulated signal "timer". The diagram is processed by the SDL/PR Generator to produce the SDL/PR code shown in Fig. 6.

Finally, when both MSD and all SDL diagrams are complete, the results are all processed by the Code Generator to produce CHILL code as shown in Fig. 7. Code is generated for the

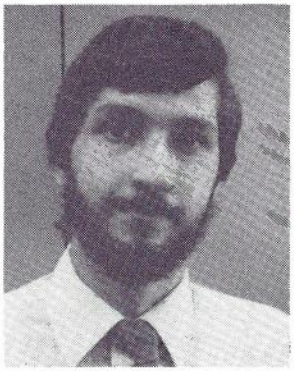
```
/* MELBA AUTOMATIC CHILL CODE GENERATION SYSTEM        */
/* SDL/PR GENERATION PROGRAM - VERSION 5.1, APRIL 1983 */

PROCESS data_link_sender;

 STATE S1 'idle';

    INPUT 'in_message INTO message buffer';
      TASK 'CREATE NEXT message frame s USING message buffer
      AND sequence_nr next_frame_to_send';
      send_again:
      OUTPUT 'out_frame FROM message frame s';
      OUTPUT 'START timer next_frame_to_send' ENCAPSULATED;
      NEXTSTATE S2 'wait_for_response';

 ENDSTATE S1 'idle';

 STATE S2 'wait_for_response';

    INPUT 'ack';
      TASK 'GET NEXT sequence_nr next_frame_to_send';
      NEXTSTATE S1 'idle';

    INPUT 'bad_ack';
      JOIN send_again;

    INPUT 'timer next_frame_to_send TIMEOUT' ENCAPSULATED;
      JOIN send_again;

 ENDSTATE S2 'wait_for_response';

ENDPROCESS data_link_sender;

/* 0 ERROR(S) FOUND IN SDL/GR DIAGRAM */
```

Fig. 6 - SDL/PR code generated for the data link sender by the SDL/PR Generator. This code is equivalent to Figure 5.

```
    .
    .
    .
ABSTRACT_TYPE_TIMER:
REGION
  SEIZE ABSTRACT_TYPE_SEQUENCE_NR ALL;
  SEIZE ABSTRACT_TYPE_CONDITION ALL;
  DCL STARTING_TIME ARRAY (SEQUENCE_NR) INT;
  DCL CLOCK_STATE ARRAY (SEQUENCE_NR) CONDITION;
  DCL CURRENT_TIME INT;
  DCL TIMEOUT_INTERVAL INT := 15;

  START_TIMER:
  PROC (NUM SEQUENCE_NR INOUT);
    STARTING_TIME (NUM) := SYSTIME;
    CLOCK_STATE (NUM) := ON;
  END START_TIMER;

  STOP_TIMER:
  PROC (NO SEQUENCE_NR INOUT);
    CLOCK_STATE (NO) := OFF;
  END STOP_TIMER;

  TIMER_TIMEOUT:
  PROC (NUMBER SEQUENCE_NR INOUT) (BOOL);
    CURRENT_TIME := SYSTIME - STARTING_TIME (NUMBER);
    CASE CLOCK_STATE (NUMBER) = ON OF
     (TRUE):
       CASE CURRENT_TIME >= TIMEOUT_INTERVAL OF
         (TRUE):  CLOCK_STATE (NUMBER) := OFF;
                  RETURN TRUE;
         (FALSE): RETURN FALSE;
       ESAC;
     (FALSE):
       RETURN FALSE;
    ESAC;
  END TIMER_TIMEOUT;

  GRANT START_TIMER;
  GRANT STOP_TIMER;
  GRANT TIMER_TIMEOUT;

END ABSTRACT_TYPE_TIMER;


ABSTRACT_TYPE_MESSAGE:
MODULE
  SEIZE ABSTRACT_TYPE_BIT ALL;
  NEWMODE MESSAGE = ARRAY (1:100) BIT;
  GRANT MESSAGE;
END ABSTRACT_TYPE_MESSAGE;


ABSTRACT_TYPE_SEQUENCE_NR:
MODULE
  SEIZE ABSTRACT_TYPE_ANSWER ALL;
  NEWMODE SEQUENCE_NR INT (0:1);
  DCL NULL READ SEQUENCE_NR := 0;

      GET_NEXT_SEQUENCE_NR:
      PROC (A SEQUENCE_NR INOUT);
        CASE A = 1 OF
          (TRUE):  A := 0;
          (FALSE): A := A + 1;
        ESAC;
      END GET_NEXT_SEQUENCE_NR;

INITIALIZE_SEQUENCE_NR:
PROC (B SEQUENCE_NR INOUT);
  B := NULL;
END INITIALIZE_SEQUENCE_NR;

  IS_SEQUENCE_NR_EQUAL_TO_SEQUENCE_NR:
  PROC (C SEQUENCE_NR INOUT, D SEQUENCE_NR INOUT)(ANSWER);
    CASE C = D OF
      (TRUE):  RETURN YES;
      (FALSE): RETURN NO;
    ESAC;
  END IS_SEQUENCE_NR_EQUAL_TO_SEQUENCE_NR;

  GRANT SEQUENCE_NR;
  GRANT GET_NEXT_SEQUENCE_NR;
  GRANT INITIALIZE_SEQUENCE_NR;
  GRANT IS_SEQUENCE_NR_EQUAL_TO_SEQUENCE_NR;

END ABSTRACT_TYPE_SEQUENCE_NR;

    .
    .
    .
SIGNAL IN_MESSAGE = (MESSAGE) TO DATA_LINK_SENDER;
SIGNAL OUT_MESSAGE = (MESSAGE) TO NETWORK_LAYER_PROTOCOL;


DATA_LINK_LAYER:
MODULE
  SEIZE IN_MESSAGE;
  SEIZE ABSTRACT_TYPE_TIMER ALL;
  SEIZE ABSTRACT_TYPE_MESSAGE ALL;
  SEIZE ABSTRACT_TYPE_MESSAGE_FRAME ALL;
  SEIZE ABSTRACT_TYPE_SEQUENCE_NR ALL;
          .
          .
          .
  SIGNAL BAD_ACK TO DATA_LINK_SENDER;
  SIGNAL ACK TO DATA_LINK_SENDER;
  SIGNAL OUT_FRAME = (MESSAGE_FRAME) TO PHYSICAL_LAYER;

  DATA_LINK_SENDER:
  PROCESS ();
    DCL BUFFER MESSAGE;
    DCL S MESSAGE_FRAME;
    DCL NEXT_FRAME_TO_SEND SEQUENCE_NR;

    INITIALIZE_SEQUENCE_NR (NEXT_FRAME_TO_SEND);

    S1/* IDLE */:
    RECEIVE CASE
      (IN_MESSAGE IN BUFFER):
        CALL CREATE_NEXT_MESSAGE_FRAME_USING_MESSAGE
             AND_SEQUENCE_NR (S, BUFFER, NEXT_FRAME_TO_SEND);
        SEND_AGAIN:
        SEND OUT_FRAME (S);
        CALL START_TIMER (NEXT_FRAME_TO_SEND);
        GO TO S2 /* WAIT_FOR_RESPONSE */;
    ESAC;

  S2 /* WAIT_FOR_RESPONSE */:
  RECEIVE CASE
    (ACK):
      CALL GET_NEXT_SEQUENCE_NR (NEXT_FRAME_TO_SEND);
      GO TO S1 /* IDLE */;
    (BAD_ACK):
      GO TO SEND_AGAIN;
    ELSE
      IF TIMER_TIMEOUT (NEXT_FRAME_TO_SEND)
        THEN
          GO TO SEND_AGAIN;
        FI;
      GO TO S2 /* WAIT_FOR_RESPONSE */;
  ESAC;

END SENDER_3;
```

Fig. 7 - part of the code generated for the data link protocol. Notice the insertion of abstract type modules from the library.

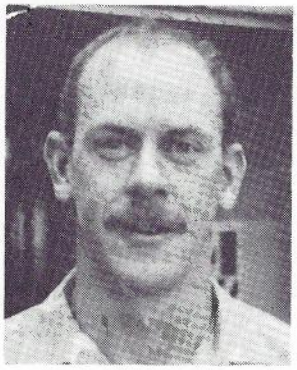
entire system based on the MSD structure with SDL processes and abstraction library entries inserted at appropriate points. This code is ready for compilation. Runtime errors may result in the modification of MSD and SDL diagrams and repetition of the code generation process.



BIOGRAPHIES

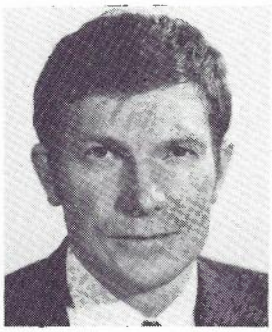
COLIN FIDGE received the B.App.Sc. degree in Computer Science (with distinction) from the Royal Melbourne Institute of Technology in 1981. He is presently completing a Masters Degree in Computer Science, again with the Department of Computing, RMIT. This paper is one of the outcomes of his research work which is primarily concerned with the data structures required to support the process of automatic HLL code generation.

Mr Fidge joined the Telecom Australia Research Laboratories, Software Engineering Research Section, in 1983.



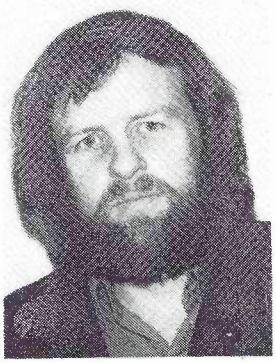
LINDSAY JACKSON received the B.E.E. Degree from Melbourne University in 1958 and the M.Eng. Degree from RMIT in 1980. He has had industrial experience in Australia and West Germany and is currently a Senior Lecturer in the Department of Communication and Electronic Engineering at RMIT.

Mr Jackson has consulted to industry in communication systems and has also been involved in extensive research activity in the areas of switching systems, software, protocols, and specification. Since 1979 he has been project leader of the MELBA automatic code generation team.



BOB PASCOE is Senior Lecturer in Computer Science at the Royal Melbourne Institute of Technology. His research interests are high level languages, high level language compilers and programming methodologies. As a regular member of CCITT Study Group XI, Bob has been formative in the recent development of the SDL language, and is currently responsible for the specification of SDL's data description aspects.

He has a Bachelor of Applied Science (Computer Science) from RMIT.



DR GREG CAIN graduated from the University of Western Australia with a BSc (Honours) in 1969 and was awarded a PhD at Monash University in 1973. He has worked as a software consultant for the CSIRO Division of Computing Research (1974-77) and has since started his own software consultancy.

Greg has been involved in consulting and software supervision on the MELBA Project since 1979.

Propagation at 500 MHz for Mobile Radio

B.R. DAVIS
R.E. BOGNER

Department of Electrical and Electronic Engineering
The University of Adelaide

The basic characteristics of radio paths relevant to mobile radio communication at U.H.F. are reviewed. Data gathered in a comprehensive survey of urban and suburban Melbourne were analysed to test for consistency with several simple models: Rayleigh envelope, a power spectrum model, log-normal distribution of mean, and the effects of distance and slope of terrain. Factors limiting the validity of some of the models are discussed. Other factors investigated were repeatability under varying weather conditions, the averaging required for reliable representation of local field strengths, and distributions associated with slow and fast fading patterns. A study of the applicability of space diversity showed that a saving of at least 3/4 of the transmitter power, for a given grade of service, could be made if diversity were used.

Because of the number and complexity of the factors influencing the propagation to a vehicle, and because weather and time of day were found to be not significant influences, it was concluded that experimental assessment is probably the cost-effective method for determining accurately a service area. Examples of suitable data gathering plans are given.

1. INTRODUCTION

The investigations described formed part of the preparations of Telecom Australia for the introduction of F.M. mobile radio services in Australian cities. Much previous work has been reported on certain aspects of U.H.F. propagation for mobile communication (Refs. 1-11) but a considerable amount of uncertainty existed as to its range of applicability, and no definite conclusions could be drawn from it as to some aspects of the design of the systems required.

The experimental work was carried out by Telecom Australia personnel, using a specially designed mobile logging system (Refs. 9, 11). From the analysis of preliminary observations made on a limited number of receiving locations a plan of data gathering was prepared. The routes and conditions are summarised in Fig. 1, and a more detailed specification is given in Appendix IV.

Specific questions that the data gathering and analysis were designed to answer included:

(a) Does one traverse of a route provide data that are adequately representative of the route, under other conditions of time and weather etc.?

(b) Do the distributions of fast variation of field strength satisfy the widely used Rayleigh model?

(c) Do the slow fading distributions satisfy the log-normal model?

(d) Do the spectra of the envelopes of the received signals support the hypothesis of a

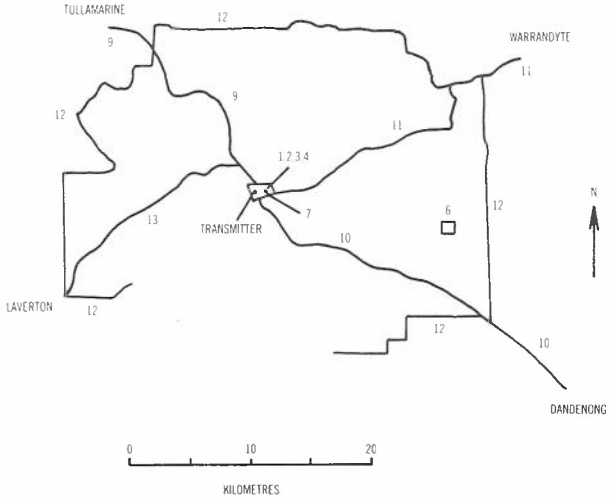


Fig. 1.1 - Location of test routes

Test No. (As used by Telecom Australia)	*Route*
1,2,3,4	*. Melbourne: Flinders St, Spring St, Victoria St, King St, Flinders St.*
1	*. 4 traverses between 10 am and noon on dry day*
2	*. As for 1, but 3 pm to 5 pm*
3	*. As for 1, but 10 pm to midnight*
4	*. As for 1, but on a rainy day*
6	*. Jordanville (Map 61): Highbury Rd, Huntingdale Rd, Outlook Rd, Quaintance St, Jubilee Rd, Highbury Rd - 4 traverses.*
7	*. Stationary, close to a busy city intersection*
9	*. GPO to Tullamarine Airport via main route*
10	*. GPO to Dandenong via main route*
11	*. GPO to Warrandyte PO via High St, Doncaster Rd, Williamsons Rd, Warrandyte Rd.*
12	*. A circumferential route at a radius of about 15 to 20 km from the transmitter.*
13	*. Laverton Railway Station, via Princes Highway.*

large number of independent scatterers (which leads to the Rayleigh fading model)?

(e) Is there a simple model that describes the variations of received signal strength with distance?

We are interested in propagation between a stationary transmitter or receiver, the "station", and a mobile vehicle-carried receiver or transmitter, the "mobile". The propagation is assumed to be linear. Time-variations due to motion of the mobile are very slow compared with the velocity of propagation so that field distributions are, observation-by-observation, effectively the same as if the mobile were stationary. Consequently propagation between station and mobile is reciprocal. Both directions of propagation are adequately characterised by study of the station-to-mobile propagation, and for practical convenience the transmitter was at the station, and the receiver at the mobile.

1.1 Basic Concepts

As the mobile moves, the propagation paths to the receiving antenna vary. Often there is no direct line of sight path and the signal arrives by a multiplicity of paths caused by reflection and diffraction from surrounding objects. Even with the mobile stationary, the propagation characteristics may vary with time due to the motion of other vehicles or objects.

To characterise fully the propagation it is necessary to obtain the time varying impulse response. This has been done by some investigators, using either pulse transmission or pseudorandom binary modulation (Ref. 1). Important features of the impulse response are the number of different paths and the time delay spread between them and the rate of variation of the impulse response.

Let an impulse applied at time t give an output $h(\tau,t)$ at time $t+\tau$. In the case of n distinct propagation paths of gain a_i and delay τ_i we have:

$$h(\tau,t) = \sum_{i=1}^{n} a_i \, \delta(\tau-\tau_i)$$

If a_i and τ_i are not time varying, then $h(\tau,t)$ is time invariant (i.e. independent of t).

The frequency response $H(f,t)$ defined such that $H(f,t)\, e^{j2\pi ft}$ is the response to an input $e^{j2\pi ft}$ is given by:

$$H(f,t) = \int_{-\infty}^{\infty} h(\tau,t-\tau)\, e^{-j2\pi f\tau}\, d\tau$$

The properties of $h(\tau,t)$ in both the τ and t domains are important. The rate at which it varies with t determines the rate of change of $H(f,t)$. At a particular frequency f, $H(f,t)$ varies with time and its variations correspond to the fading observed in practice.

The important property of $h(\tau,t)$ in the τ domain is the time delay spread, which is the range of values of τ for which $h(\tau,t)$ is not negligible. Inversely related to this is the coherence bandwidth, which is the range of frequencies over which the fading is not frequency selective. (A useful formal definition is the frequency separation of two sinusoidal components such that the cross correlation of their envelopes is 50%.)

The mobile field intensity recording receiver used effectively takes time samples of $|H(f_o,t)|$ where f_o is the carrier frequency. No information regarding phase or the time delay spread or coherence bandwidth is available.

1.2 Phase Information

The phase information $\theta = \arg H(f_o,t)$ is of some interest, particularly if frequency modulation is used. Phase variations give rise to so-called random frequency modulation or "FM noise". If f_m is the maximum doppler frequency caused by motion of the vehicle with velocity v,

$$f_m = \frac{v}{c} f_o$$

Then it has been shown that for $f > f_m$ and a large number of scatterers, the power spectrum of $\dot{\theta} = \frac{d\theta}{dt}$ is given by

$$S_{\dot{\theta}}(f) \approx 2\pi^2 F^2/f$$

where F is the radius of gyration of the spectrum of the in-phase (and quadrature) components of the received carrier in Hz. If there is a large number of scatterers uniformly distributed around the vehicle then $F = f_m/\sqrt{2}$ (Ref. 3).

If the demodulated signal is then band-limited to between W_1 and W_2 Hz then the random FM noise in this band is

$$N = 2\pi^2 f_m^2 \ln \frac{W_2}{W_1}$$

For an FM system with peak frequency deviation f_d, the signal output for maximum sinusoidal modulation is

$$S = 2\pi^2 f_d^2$$

The signal to random FM noise ratio is therefore

$$SNR = \frac{f_d^2}{f_m^2 \ln \frac{W_2}{W_1}}$$

For a vehicle travelling at 20 m/s (72 km/hr) with f_o = 500 MHz, the doppler

frequency is f_m = 33.3 Hz. For a narrowband FM system with f_d = 3 kHz and W_1 = 300 Hz and W_2 = 3000 Hz, the signal to FM noise ratio is 35 dB.

The FM noise power increases quadratically with vehicle speed and carrier frequency.

1.3 Coherence Bandwidth

The time delay spread for propagation in urban and suburban areas has been studied by others (Refs. 1,3,4), and while significantly different results have been obtained, it seems that values in excess of 10 μsec are comparatively rare. Since this time corresponds to a free space path difference of 3 km this seems a reasonable result, as most reflections might be expected to be in the vicinity of the mobile if the stationary antenna is well elevated.

In the case where the time delays are exponentially distributed with mean σ sec, the coherence bandwidth is given by $1/2\pi\sigma$ (Ref. 5, Chapter 1) and is 100 kHz or greater if $\sigma \leq 1.6$ μsec (the probability of delay being 10 μsec or greater is then 0.2%). A narrowband FM system of 10 - 20 kHz bandwidth would experience essentially non-selective fading.

Time delay spread (or coherence bandwidth) would, however, be a significant factor in any wideband channel (which might apply if digital modulation or time division multiplex were used).

1.4 Field Strength Models

A study of field strength properties has two main aims. Firstly, it enables the quality of service in an area where measurements have been made to be described, mainly in statistical terms. Secondly, by studying the statistical consistency of the results, a determination of the minimum number of measurements needed to adequately characterise the propagation statistics in that area can be made. If sufficiently good models are obtained little or no further measurement may be required. Should the statistics be found to behave consistently in the sense that their dependence on certain parameters can be established, then in effect a good model has been obtained.

1.4.1 The structure of fast fading

The ideal theoretical model for fast fading is based on the assumption of a large number of independent propagation paths, which by the central limit theorem leads to a Rayleigh distribution for the received envelope. Hence the measurements may be tested for goodness of fit to a Rayleigh distribution using a chi-squared test (Ref. 12). As a mobile moves, however, the surrounding topography changes. The fast fading characteristics might be expected to change at a rate such that significant variation occurs if the vehicle moves a distance λ/4 (0.15 m at 500 MHz) whereas the topography probably changes its gross features over a significantly larger distance.

Hence it is desirable to study the distance over which the fast fading is statistically stationary (whether it has a Rayleigh distribution or not).

1.4.2 Structure of slow fading

By slow fading is meant the change in signal strength due to topographical changes rather than that due to multipath. The multipath (or fast) fading will be superimposed multiplicatively on the slow fading.

The slow fading can therefore be thought of as a time varying mean about which the fast fading occurs. Studies (Ref. 5, Chapter 2) have shown that the slow fading is approximately distributed according to a log-normal law (i.e. the signal strength in decibels has a gaussian distribution) and this can also be justified theoretically by postulating the slow fading to be caused by a large number of successive but independent attenuations.

1.4.3 Effect of street orientation

The propagation to a mobile moving along a street oriented radially with respect to the base transmitter may be significantly different from that in a street oriented transversely. Also, and particularly in the latter case, the propagation might be expected to vary from one side of the street to the other (e.g. it might be worse on the side nearest the base due to shielding by buildings along the street).

1.4.4 Dependence on range and terrain

In addition to the inverse square law attenuation of field strength with distance that occurs in free space propagation, the field strength is influenced by the intervening topographical features, buildings and vegetation. Diffraction around intervening obstacles and reflections make major contributions in most situations that cannot be predicted in detail without extensive detail of the physical environs and extensive calculations.

1.4.5 Effect of diversity

Space diversity techniques can make significant savings in the transmitter power required to cover a specified area (Ref. 5, Chapter 6). Alternatively the use of diversity can extend the area of coverage or improve the quality of reception in the same area.

The penalty paid is in terms of increased system complexity and costs of receivers (including associated antenna structures). It is of course feasible to have a system in which both diversity and non diversity transmissions can be used simultaneously, and with this sort of compatibility, diversity receivers could be introduced as and when desired.

Although it is possible to have diversity provided at the sending end, in practice it seems likely that the simplest solution is to have it at the receiving end. For the mobile this implies two or more antennas and a more

complex receiver. The antenna spacings (Ref. 5, Chapter 6) need only be of the order of half a wavelength.

At the base station however, the spacing required is more like 20 to 30 wavelengths and this may require a significantly more complex antenna arrangement. Also at least 3 base antennas would be desirable to avoid orientation-dependent performance which might be obtained with only two.

2. REPEATABILITY

In order to test whether a set of measurements of field strength at a particular location could be taken as representative of that location, several traverses of the same path were made at different times of day and under different weather conditions. Because the position of the mobile would not be exactly the same on each traverse and the number and position of other vehicles and objects would also be different, it is not to be expected that the signal strength profiles for different traverses will be the same. Clark (Ref. 3) shows that if the field consists of many random contributions the correlation of received signals separated spatially by d is:

$$\rho \approx J_o^2 (2\pi d/\lambda) \qquad (2.1)$$

where ρ is the correlation coefficient of the signal envelopes and λ is the wavelength. At 500 MHz, the first zero of the Bessel function gives zero correlation at $d = 0.38\lambda$ which is 23 cm.

Thus the test of repeatability is concerned with whether the local mean is the same for each traverse. The local mean is expected to vary slowly with distance since it should be a function of the gross physical dimensions of buildings, trees and other reflecting and obscuring objects.

We note that all measurements are made with respect to distance x and are taken at equal distance increments. The variable t only appears because of the motion of the vehicle. Hence we shall consider the signals as the functions of time which would arise if the mobile were driven at constant velocity.

The signal strength received may be modelled as:

$$s(t) = m(t) + v(t) \qquad (2.2)$$

where s(t) is the measured field strength, m(t) is the local mean and v(t) is the multipath fading, all measured in decibels. The additive model (in decibels) is arrived at on the basis that the local mean m(t) is a factor describing the attenuation of each signal path to the mobile, whereas v(t) describes the interference and reinforcement between the various signals arriving at the mobile along different paths. These effects should be purely multiplicative (in voltage) and hence additive in decibels.

On two separate traverses of the same path, separated by a metre or more, the measured signal strengths may be written as:

$$s_1(t) = m_1(t) + v_1(t) \qquad (2.3a)$$

$$s_2(t) = m_2(t) + v_2(t) \qquad (2.3b)$$

where it is expected that there will be negligible correlation between $v_1(t)$ and $v_2(t)$, the multipath fading components, and the object is to determine whether $m_1(t)$ and $m_2(t)$ are the same. Because accurate location of the mobile is not possible from the data supplied, the signal strengths measured may be offset in time, and we need to test for equality between $m_1(t)$ and $m_2(t-T)$ where T is the time offset.

Since m(t) is expected to vary slowly compared with v(t), an estimate of m(t) can be formed by low-pass filtering s(t). If the samples of $s_1(t)$ and $s_2(t)$ are designated $s_1(i)$ and $s_2(i)$, $i = 1, 2 \ldots N$, then j estimates of the means formed are:

$$\hat{m}_1(r) = \frac{1}{k} \sum_{i_1(r)+1}^{i_1(r)+k} s_1(i) \qquad (2.4a)$$

$$\hat{m}_2(r) = \frac{1}{k} \sum_{i_2(r)+1}^{i_2(r)+k} s_2(i) \qquad (2.4b)$$

where

$$1 \leq r \leq j$$

$$i_1(r) = (r-1)k + i_{base}$$

$$i_2(r) = (r-1)k + i_{base} + i_{shift}$$

$$-i_{base} \leq i_{shift} \leq i_{base}.$$

This is shown diagrammatically in Fig. 2.1. The means $\hat{m}_1(r)$ and $\hat{m}_2(r)$ are the estimated means of the rth block of data, each block being of length k data samples. The value of i_{shift} is varied, and best agreement between the the means is expected when i_{shift} corresponds to the time offset T. The number of data points N must satisfy:

$$N \geq jk + 2i_{base}. \qquad (2.5)$$

To compare the estimates, the root mean square difference was formed:

$$\sigma = \frac{1}{j}\sqrt{\sum_{r=1}^{j} [\hat{m}_1(r) - \hat{m}_2(r)]^2} \qquad (2.6)$$

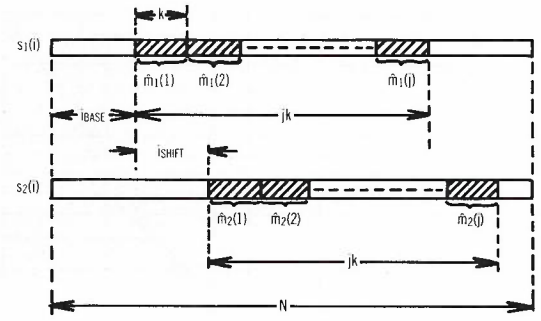


Fig. 2.1 - Samples used to calculate local means for matching of records to test repeatability

Filtering reduces the effect of v(t) but does not eliminate it. An estimate of the effect of filtering can be made as follows. The standard deviation of v may be shown to be 5.57 dB assuming v(t) has a Rayleigh distribution. The samples are taken at 5 cm intervals and are correlated such that in a set of k consecutive samples there are roughly 0.2 k independent samples (since the first zero of the correlation function is at d = 23 cm, i.e. about 5 times the spacing). Hence the standard deviation after averaging over k samples should be reduced to approximately $5.57/\sqrt{.2\ k} \approx 12/\sqrt{k}$ dB. A more exact analysis is presented in Appendix II which indicates that this approximate result is sufficiently accurate for our purposes. Hence for k = 100, the standard deviation of $\hat{m}$ due to v should be of the order of 1 dB. Hence we would not expect the agreement between m_1 and m_2 as expressed in the value of σ to be much better than about 1.5 dB for k = 100.

2.1 Observations

Several runs were made over the routes of Tests 1,2,3,4 and 6, the repetition occurring within two hours in each case. Field strength data were compared for pairs of runs by the above minimum mean squared error (mmse) process with records of 20,000 samples. This length of record, corresponding to 1 km of route, was chosen to ensure that spurious matching of sections which did not correspond should be unlikely. The possibility of such spurious matching is shown in Fig. 2.2.

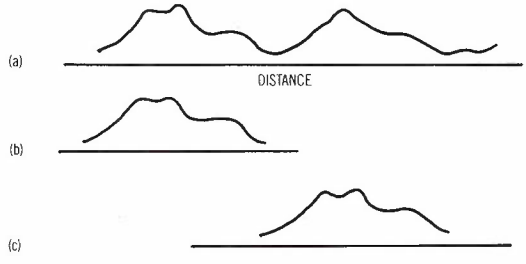


Fig. 2.2 - Possibility of spurious matching of profile of local means due to chance similarities of sections of data.
(a) complete data record for a street
(b) section of data record correctly matched
(c) section of data record with spurious matching

Typical results are summarized in Table 2.1.

The minimum rms difference is about 1 dB in all cases where the smoothing filter k was 1500. This shows that when the smoothing length is large compared with the coherence distance of the fast fading (about 23 cm) then the slow fading patterns are reliably reproducible. The agreement to 1 dB was observed irrespective of the differences of time of day or of weather conditions, and the residual difference thus is more likely to be attributable to retention of some contribution from mismatching of the fading patterns.

TABLE 2.1 - Repeatability under various conditions (Victoria Street, Melbourne)

Tests No.	Comparison	Filter length k	Average (over j)	rms diffce dB
1,3	Midday-midnight	150 1500	100 10	1.7 0.7
1,2	Midday-late afternoon	150 1500	100 10	1.8 0.5
2,3	Afternoon-midnight	150 1500	100 10	1.8 0.6
1,4	Dry day-rainy day about midday	150 1500	100 10	1.7 0.9
1,1	Self comparison	150 1500	100 10	.0 .0

The comparisons using smoothing lengths k of 150 samples show a significantly higher rms difference of about 2 dB, attributable to there being a greater contribution from the random fast fading.

Comparison of typical graphs of the rms differences for 150-sample smoothing and 1500-sample smoothing (see Fig. 2.3) shows the latter to be smoother, and in particular the dip is broader. This difference is consistent with the reduction of high-frequency detail by the longer smoothing window. Fig. 2.3 shows similar effects in the comparisons of runs taken on a fine day and a wet day.

A check of the process was made by comparing one record with itself (Fig. 2.3). The minimum rms error is zero as it should be, and the shape of each curve is consistent with that obtained in the other comparisons. The curve for the smoothing filter with length 150 samples has a sharper dip than that for 1500 samples, corresponding to the 150-sample filter having a broader passband. Both curves show that at distances remote from the notch the rms difference is still rising with distance, suggesting that the slow fading has a spectrum with strong low-frequency energy, at frequencies below one cycle per km.

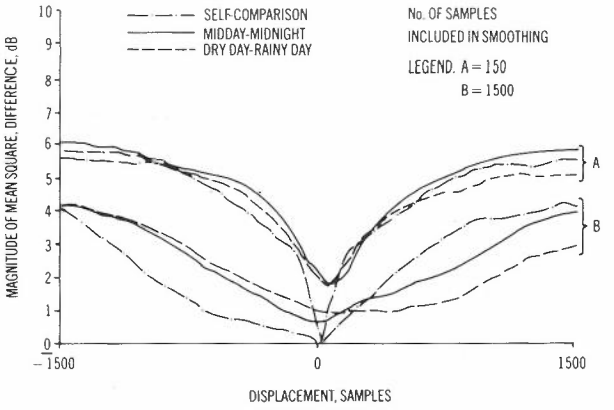


Fig. 2.3 - Comparison of field strengths at different times and weather conditions, using two different smoothing windows.

A couple of additional observations supported the repeatability of suitably-averaged measurements. In the long distance runs (Section 5) two points were each common to two routes, i.e. routes 9 and 13 and also 11 and 12. In these cases the 5000-sample averages, corresponding to 250 metres of route, agreed to within 1.24 dB. This was true both for averages of field strength in dB, and in voltage.

2.2 Conclusion on Repeatability

Field strength measurements are reliably repeatable to within about 1 dB when the fast nearly random fading due to multiple paths is averaged out. Suitable averaging distances are of the order of 250 metres.

3. DISTRIBUTION OF FIELD STRENGTH VALUES

Clarke's model (Ref. 3) of mobile radio propagation predicts that the statistical distribution of field strength values should follow a Rayleigh distribution. This is obtained by central limit theorem arguments based on the superposition of a large number of roughly equal strength components from different reflecting and diffracting objects. In practice this ideal situation does not exist. Firstly, the number of scatterers may be quite small, perhaps 3 or 4 for example, and secondly the strengths of the components may vary along the path of the mobile.

The effect of a small number of scatterers would be to cause the fast-fading multipath statistics to be other than Rayleigh. For example, Fig. 3.1 shows the comparison between the Rayleigh distribution and the rather extreme case of only two equal amplitude scatterers for which $r^2 = r_o^2 (1 + \cos\theta)$, where $r_o^2 = \langle r^2 \rangle$ and θ is the relative phase, assumed uniformly distributed over 0 to 2π.

The effect of variation in the component strengths is to cause the multipath fading effects to be superimposed on a more slowly varying local mean which depends on the gross physical dimensions of the scattering objects and which might be expected to vary only slightly over a distance of 10 m.

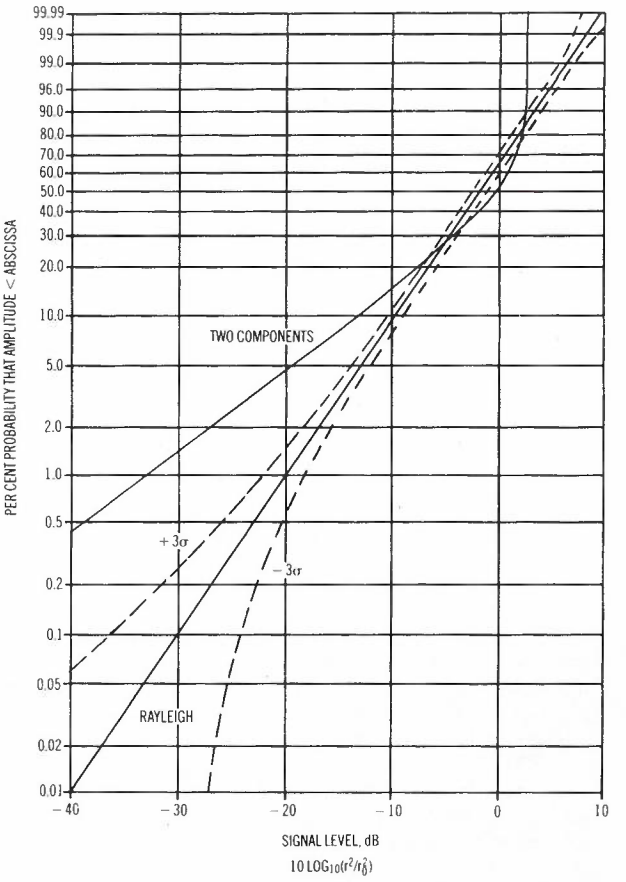


Fig. 3.1 - Comparison of Rayleigh distribution with distribution of envelope of two component multipath propagation

Since, as discussed in Section 2, the multipath fading effects and the local mean variations are expected to be additive in decibels and to vary at different rates, it was decided to try to separate them by filtering.

3.1 Separation of slow and fast fading

We model the measured signal strength (in dB) as:

$$s(t) = m(t) + v(t) \qquad (3.1)$$

where m represents the local mean and v the variation about that mean. The local mean is estimated by averaging the signal strength over 2 k+1 samples. This is similar to the averaging in Section 2 except that it is done continuously rather than in blocks and the averaging is symmetrical about the time index i. Then

$$\hat{m}(i) = \frac{1}{2k+1} \sum_{j=-k}^{k} s(i+j). \qquad (3.2)$$

This estimated local mean is subtracted from s(t) to give an estimate of the fast fading multipath factor

$$\hat{v}(i) = s(i) - \hat{m}(i). \qquad (3.3)$$

The distributions of $\hat{m}$ and $\hat{v}$ were obtained by examining 20,000 data points and finding the cumulative distribution at 0.5 dB increments. An estimate of the accuracy of the plotted graphs is desirable.

If $p = \text{Prob}(v<V)$ is the true cumulative distribution of v, then we estimate p by $\hat{p}$ where

$$\hat{p} = \frac{1}{N} \sum_{i=1}^{N} p_i \qquad (3.4)$$

where $p_i = 0$ if $v(i) \geq V$
$p_i = 1$ if $v(i) < V$.

Then if the samples are independent, the expected value of $\hat{p}$ is

$$\langle \hat{p} \rangle = p \qquad (3.5a)$$

$$\text{Var } \hat{p} = \frac{1}{N} p(1-p) \qquad (3.5b)$$

However here the samples are correlated and the effective number of independent samples is approximately 0.2N (based on a 5 cm sample separation with a correlation distance of 23 cm), and thus the number which should be used in estimating Var $\hat{p}$ is 0.2N. Rigorous analysis of this is considerably more complex and is not considered here.

The $\pm 3\sigma$ limits for N = 20,000 are shown in Figure 3.1 for the Rayleigh distribution.

3.2 Effects of various filters

The best filter length k to separate the local mean and multipath components is not obvious. A filter with k too small (large bandwidth) will result in $\hat{m}(t)$ containing m(t) plus a significant contribution from v(t). Unfortunately v(t) will have fairly large components at low frequencies due to the nature of the spectrum of the envelope. Similarly, the complementary signal $\hat{v}(t)$ will consist of v(t) with the low frequency components attenuated. It has not been possible to determine what effect this would have on the amplitude statistics in theory. A filter with k too large would result in the opposite situation, with m consisting only of the low frequency components of m(t), whereas $\hat{v}(t)$ would contain most of the frequency components of v(t) plus the higher frequency components of m(t). The power spectrum of v(t) for an ideal model is derived in Appendix I.

In many of the graphs plotted, a value of k = 100 was used, corresponding to an averaging distance of 10 m at a 5 cm sample spacing. This was chosen on the basis that this gives a low pass filter cut off frequency of approximately 0.06 times the doppler frequency* and hence $\hat{v}(t)$ should contain most of the frequency components of v(t), which extend up to roughly twice the doppler frequency.

A value of k = 1000, used in one test, corresponds to an averaging distance of 100 m. Here we would expect to find a significant component of m(t) in $\hat{v}(t)$ and $\hat{m}(t)$ only containing the lower frequency components of m(t). This was confirmed by the smaller variance obtained for $\hat{m}(t)$ compared with the same data with k = 100. However the effect on the distribution of $\hat{v}(t)$ is less obvious.

Figure 3.2 illustrates the effect on the fast fading component of changing the distance over which the mean is estimated. The upper distance of infinity corresponds to a common mean for all samples in the record of 20,000. As the averaging distance is increased above 25 m the departure from the Rayleigh distribution, which corresponds to the straight line, is seen to increase. This departure corresponds to the superposition of the effect of the varying local mean on the fine structure associated with multipath effects.

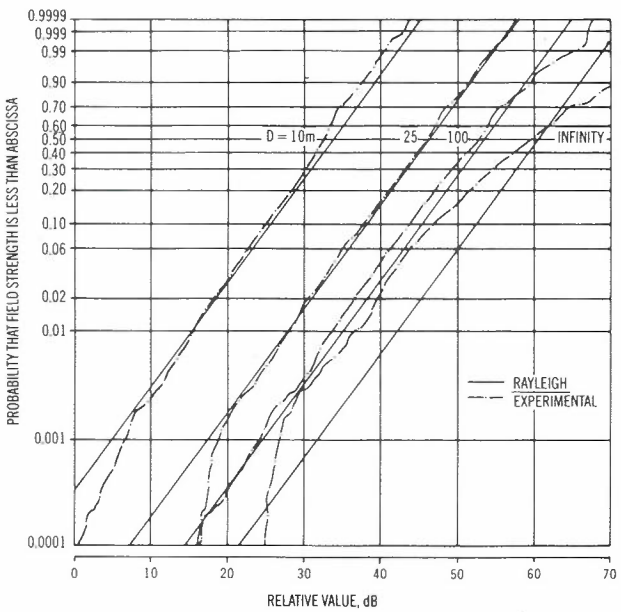


Fig. 3.2 - Departure of fast fading distribution from Rayleigh with increase of the distance D over which the local mean is estimated. The same data were used in each case, taken from a circumferential route. The straight lines correspond to ideal Rayleigh distributions. The horizontal shifts were introduced for clarity.

*The doppler frequency refers to the apparent shift in frequency of the carrier as experienced by a vehicle moving directly toward the transmitter under free-space conditions. Referring all motion-induced spectra to this frequency renders the frequency scales independent of velocity and of absolute frequency. An equivalent concept is relative spacial frequency, i.e. cycles per carrier wavelength.

Figure 3.3 shows the converse effect of averaging distance on the distribution of the estimates of the local mean. As the averaging length is increased the variance decreases as expected, the major decrease in variance being associated with the removal of fine structure due to fast fading. This corresponds to the difference between the distributions for 10 m averaging, and for 'no averaging".

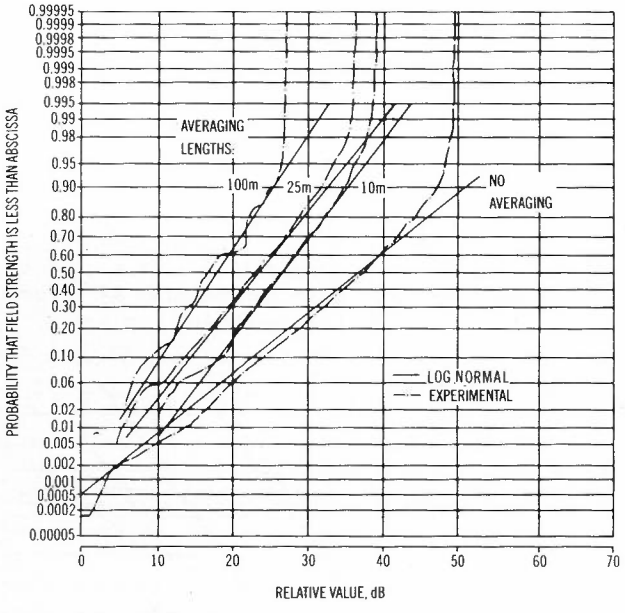


Fig. 3.3 - Distribution of local means $\hat{m}$ for various averaging lengths. The variance, which describes the spread of the distribution along the dB scale, decreases with increased averaging. The graphs have been separated by horizontal shifts for clarity.

3.3 Effects of street orientation

In Clarke's model, the street orientation relative to the base station is not a parameter, since the model assumes scatterers uniformly distributed around the mobile. Even if the scatterers are not uniformly distributed, this would not be expected to affect the multipath amplitude distribution which should be close to Rayleigh if a large number of scatterers is involved.

The street orientation would be expected to have the greatest effect on the local mean of the signal strength and the spectrum of the multipath fading. Streets running radially should have higher mean signal strengths than those running circumferentially because of reduced obstruction between the base and mobile. The spectrum corresponding to radial streets should have larger high frequency components due to enhanced doppler effects. Spectral effects are studied in Section 4. Here we are concerned with the amplitude distributions.

3.4 Stationary observations

When the mobile is stationary variations due to multipath effects should be reduced. If the scatterers remain fixed, then the signal strength would not vary at all. However with the other moving vehicles and objects, the scattering paths will vary with time and hence the signal strength will vary.

The variation in signal strength when the mobile is stationary depends on several factors. If the contribution from moving scatterers is only a small fraction of the total signal received, then only small variations in signal strength would be expected. However in the event the mobile was near a point where the contributions from fixed scatterers summed to zero, then the variation in signal strength would be large. In this case, however, the number of scatterers contributing to the signal strength variation might be small and the distribution not a good approximation to Rayleigh. Thus we see that in both cases the distributions would not be good fits to Rayleigh distributions, but for different reasons. The distribution obtained is shown in Fig. 3.4, with the reference being 0.1 $\mu V\ m^{-1}$.

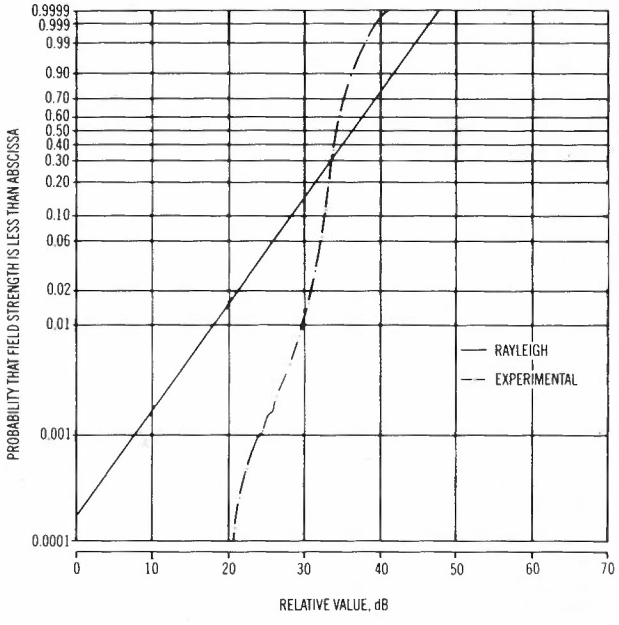
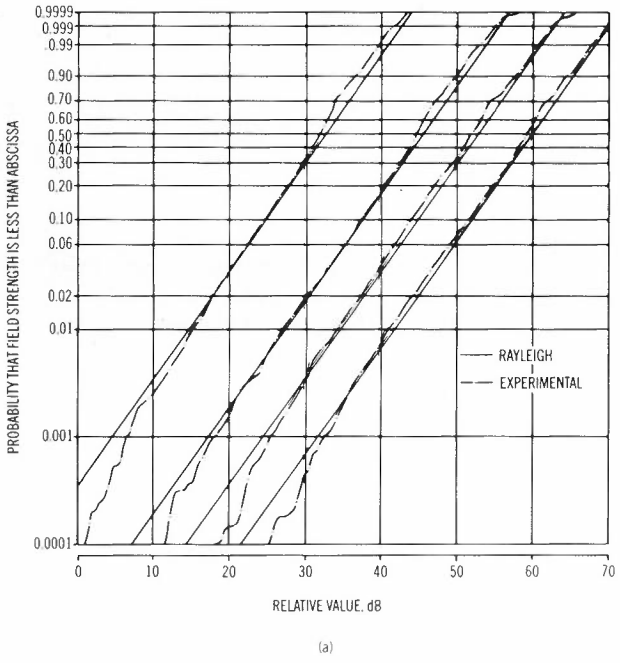


Fig. 3.4 - Cumulative distribution of field strength when vehicle was stationary.

3.5 Nature of the distributions

The cumulative probability distributions measured over N = 20,000 samples are shown in Figs. 3.5 and 3.6. The distributions of $\hat{m}$ are plotted on gaussian scaled axes, such that a gaussian distribution would appear as a straight line. This would correspond to a log-normal distribution for the envelope voltage. The distributions of $\hat{v}$ are plotted on Rayleigh scaled axes, such that a Rayleigh distribution



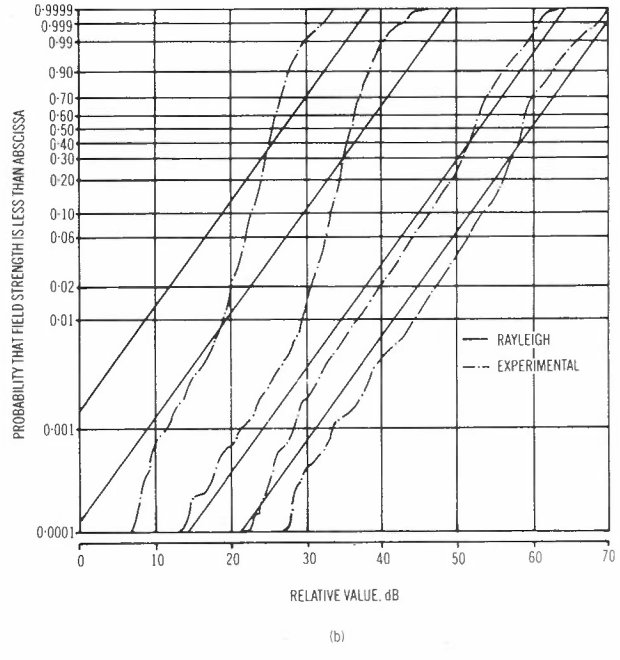
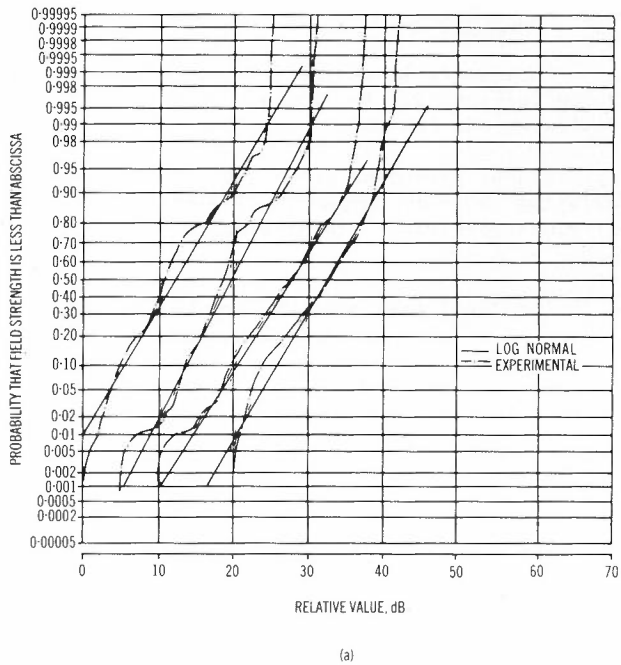


Fig. 3.5 - Distribution of fast fading component, relative to a local mean obtained by averaging over 10 m. The straight lines correspond to ideal Rayleigh distribution, and horizontal separation is introduced for clarity. Results for circumferential routes are shown in (a), and for radial routes in (b)



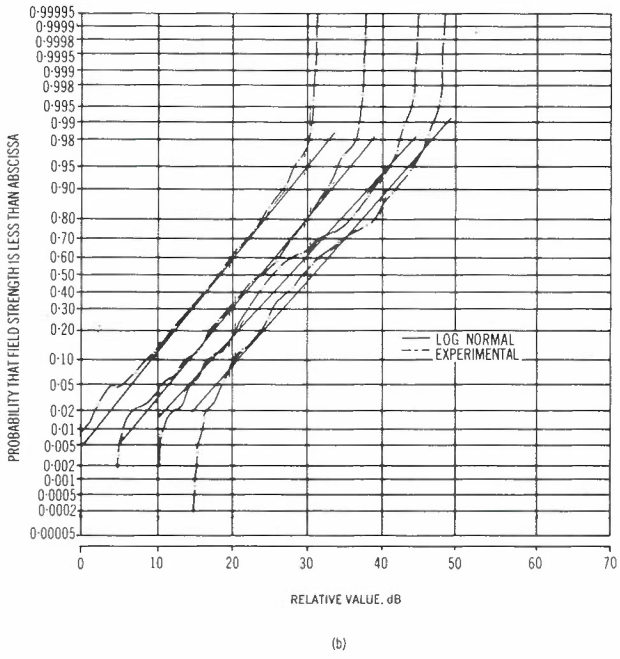


Fig. 3.6 - Distributions of means of 10 m runs of field strength measurements. The horizontal separations have been introduced for clarity. Circumferential routes (a) have variances of 7 to 8 dB while radial paths (b) have variances of about 5 dB.

of envelope voltage would appear as a straight line. Because the Rayleigh distribution is described by only one parameter, only straight lines of a particular slope represent a Rayleigh distribution. For the axes shown, this slope corresponds to the slope of the line drawn on the diagram, and any other slope corresponds to a Weibull distribution (Ref. 6).

It can be seen in Figs. 3.5 and 3.6 that radially oriented streets have $\hat{v}$ distributions which are worse fits to Rayleigh than for circumferentially oriented streets, probably due to line of sight conditions resulting in a strong direct ray. This idea is supported by the fact that the local mean $\hat{m}$ has a higher value and about 3 dB less variance for radially oriented streets.

3.6 Conclusions

From analysis of the data, of which some is shown in Figs. 3.1 to 3.6, we can draw the following conclusions:

(1) Signal strengths are higher along streets oriented radially compared with those oriented circumferentially. The difference is of the order of 6-8 dB.

(2) The local mean as estimated by low pass filtering is approximately log normal in the vicinity of the mean with a standard deviation between 5 dB and 8 dB. Radially oriented streets tend to have the lower spread and circumferentially oriented streets the larger values. However, the tails of the distribution are somewhat compressed as compared with the normal distribution.

(3) The multipath fading as estimated by high pass filtering tends to be approximately Rayleigh, but several examples show considerable divergence from this. However in every case the Rayleigh distribution would give a larger proportion of low signal values. It is the worst case, and hence the assumption of a Rayleigh distribution would result in conservative design estimates.

(4) The signal strength varies significantly even when the mobile is stationary. However the variation is less than if the mobile is in motion. The variation in signal strength is probably a function of the type and density of moving vehicles in the vicinity and gives the advantage that the effects of stopping at a point of extremely low signal strength are thereby reduced.

(5) It is not possible to deduce the time delay spread and hence the coherence bandwidth from these measurements because a particular received field pattern at one frequency may be associated with many different patterns of differential time delay. However this should not be a significant factor if channels of 20-30 kHz bandwidth are used.

4. SPECTRUM OF FIELD STRENGTH ENVELOPE

The spectrum of the envelope (or envelope-squared) gives an indication as to the location of the principal scatterers* and to some extent the number of scatterers involved. Dominant scatterers tend to give line components in the spectrum, and the scatterers in the line of vehicle motion tend to contribute to the high frequency end of the spectrum, whereas those located to the side tend to contribute to the lower frequency parts of the spectrum.

The decision to use the envelope-squared was influenced by two considerations. Firstly, the envelope-squared should be bandlimited at twice the doppler frequency, thus simplifying aliasing problems associated with the discrete sampling, and secondly the theoretical result for the ideal case of many scatterers uniformly distributed can be computed analytically (Ref. 3).

4.1 Sampling interval

For a mobile moving at velocity v, the maximum doppler frequency f_m is given by

$$f_m = \frac{v}{\lambda} \text{ Hz}$$

where λ is the wavelength of the transmitted carrier. The samples are taken at intervals of d which corresponds to a sampling frequency of v/d Hz.

The requirements of the sampling theorem for envelope-squared are satisfied if:

$$\frac{v}{d} > 2(2v/\lambda)$$

i.e.

$$d < \lambda/4.$$

Since d = 5 cm and λ = 60 cm at 500 MHz, this requirement is satisfied and no aliasing effects should arise.

The spectrum is obtained from a fast fourier transform using 2048 samples; it yields spectral estimates spaced by

$$\frac{v}{Nd} \text{ Hz} = \frac{\lambda f_m}{Nd} = .00586\ f_m.$$

A smoothing of width 25 samples was used to reduce the standard deviation of the spectral estimates to 0.2 of the mean values, resulting in an effective resolution of 0.146 f_m.

4.2 Theoretical spectrum of envelope power

For Clarke's model, the power spectrum of the AC component of envelope power is (Ref. 3)

$$S(f) = \frac{s_o^2}{\pi^2 f_m} K\left(\sqrt{1 - f^2/4f_m^2}\right) \qquad (4.1)$$

where r(t) is the envelope, $s_o = \langle r^2 \rangle$, f_m is the maximum doppler frequency v/λ and K(x) is the complete elliptic integral of the first kind.

$$K(x) = \int_0^{\pi/2} [1 - x^2 \sin^2\theta]^{-\frac{1}{2}}\, d\theta \qquad (4.2)$$

The theoretical spectrum of envelope power as given by equation 4.1 has a sharp cut-off at $f = 2 f_m$ and has a logarithmic infinity at f = 0.

*We use the term scatterer to denote any object which by diffraction, reflection or refraction contributes to the received field.

Practical spectra should exhibit similar properties, with deviations from Clarke's model caused by a dominant scatterer or by a small number of dominant scatterers and by the necessarily finite number of data. Reflections from other moving vehicles could cause spectral components at frequencies greater than 2 f_m.

In Appendix I the power spectrum of the log envelope is derived. This is relevant to the filtering used to separate local mean and fast-fading components as is done in Section 3.

4.3 Effect of street orientation

If the received field is due only to scatterers which contribute equally from all directions, the orientation of the street is unimportant. However, if the scatterers tend to be predominantly in the direction of vehicle motion, then radially oriented streets would be expected to have greater high frequency components than circumferentially oriented streets. If the dominant scatterers are transverse to the vehicle motion then the reverse situation would apply. Where there is a significant direct ray to the mobile (where line of sight conditions exist for instance) then the former situation will apply, and of course line of sight conditions are perhaps more likely for radially oriented streets.

Examples of the results are shown in Fig. 4.1. This shows reinforced high frequency components for radially oriented streets, thus indicating either line of sight conditions or a predominance of scatterers in the direction of vehicle motion. Since the envelope distribution curves of the multipath fading components obtained for radially oriented streets tend to be a worse fit to Rayleigh than for the circumferentially oriented streets, this would seem to suggest that line of sight conditions prevail. Hence the dominant scatterers could still be to the side of the vehicle motion.

4.4 Stationary observations

If no other moving vehicles were present, it would be expected that the signal received while the mobile was stationary would be a constant envelope sinewave. However, the amplitude statistics show that variations of the order of ±10 dB occur, which is small compared with Rayleigh fading but not negligible. The spectrum obtained for the stationary test is shown in Fig. 4.2.

Here the samples are taken at real time intervals of 25 ms. The spectrum indicates a predominance of low frequency content, but otherwise does not seem to have any particular features. The sharp cut-off frequency effect observed for mobile situations is of course not present, due to the spread of the velocities of the moving scatterers.

4.5 Conclusions

(1) Radially-oriented streets tend to have a spectrum with dominant high frequency components. This is probably due to the combination of a direct signal from the transmitter with reflections back toward the transmitter. This signal and reflection would give the maximum and minimum doppler frequencies of any of the multipath components.

(2) Circumferentially-oriented streets, for which line of sight conditions would not normally be expected, tend to show a spectrum with weaker high frequency components than predicted by a uniform scattering model.

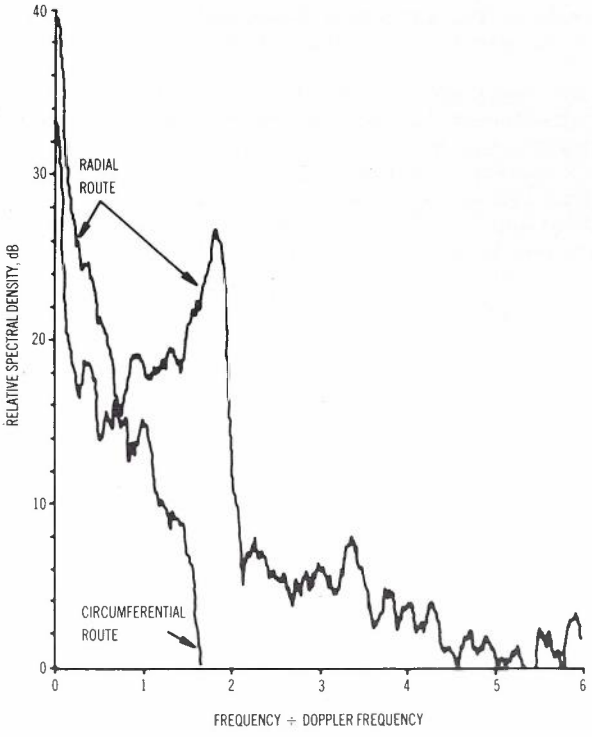


Fig. 4.1 - Spectra of squares of envelopes

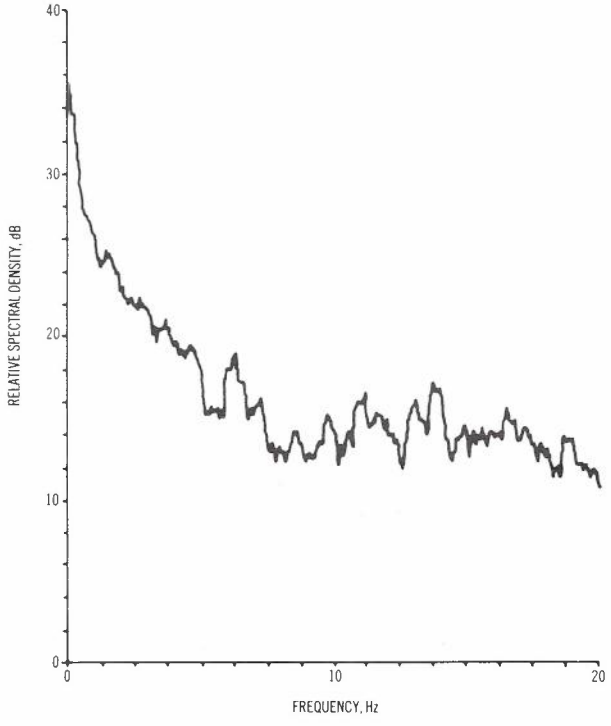


Fig. 4.2 - Spectrum of envelope squared. Stationary test, Test 7.

Possibly the strongest components are transverse to the direction of the street.

(3) The absolute values of spectral density in circumferential streets are considerably lower than in radial streets, by 20-30 dB. This difference is not the same as that of 8-10 dB observed for the mean field strengths (Section 3). This discrepancy has been studied and may be accommodated by the facts that:

a. The spectra under consideration are spectra of envelopes-squared. This would double the decibel range.

b. The radial-street spectra correspond to the interaction of a predominant (direct wave) component with scattered components while the circumferential-street spectra correspond to interaction of multiple small components.

(4) No significant frequency components were found above twice the doppler frequency, indicating that there were no dominant moving scatterers.

(5) The spectrum obtained when the mobile was stationary did not show any predominant frequency components, supporting the view that the field strength contains a dominant static component together with a wide spread of frequency shifted components from moving scatterers.

5. SLOW VARIATIONS - EFFECT OF DISTANCE

The received field at a particular location depends on that location through:

(a) distance from the transmitter

(b) transmitting antenna gain relevant to that location

(c) the presence and nature of intervening obstacles

(d) the nature of scattering by objects contributing multiple paths

(e) the transmitter altitude

(f) the receiver site altitude and ground slope

(g) street orientation (really a factor in (c) and (d)).

With the possible exception of (f) these have received considerable attention in the literature, and many models have been proposed to describe the effects. A fairly comprehensive survey of the models is given in Ref. 6. Many of the models attempt to describe the effects of distance by reference to knife edge diffraction produced by hypothesized intervening obstacles. Factors which compromise the applicability or the accuracy of the models include the multiplicity of diffracting obstacles, their departure from the ideal infinite knife edge in shape and extent, and their electrical properties. For example foliage is different from metal buildings. Empirical correction factors have been proposed (Refs. 5,7,10) to allow for shading by hills.

In the presence of such uncertainties, discrepancies between the prediction of models and observations often show standard deviations of about 20 dB.

The interpretation of measurements of received field to derive a value to represent the local field strength requires a decision. If the local variations are considered to represent Rayleigh fading about a local mean it appears reasonable to use the local root mean square field strength expressed in volts per metre. However, if the local distribution is truly Rayleigh it turns out that averaging the dB values gives a value 2.507 dB less than that obtained by averaging the voltages and then converting to dB (Appendix 1). If the distribution is not Rayleigh then this result is not exact and different values might be expected for the two averages, viz.

$$20 \log_{10} (v_{rms}) = 10 \log_{10}[\frac{1}{N} \sum_{i=1}^{N} v_i^2] \qquad (5.1)$$

where v_i is the field strength in volts/metre, and

$$dB_{ave} = \frac{1}{N} \sum_{i=1}^{N} dB_i \qquad (5.2)$$

where $dB_i = 20 \log_{10} v_i$ is the decibel value of the field strength.

The results of both methods (5.1) and (5.2) were calculated for a total of 70 points on 5 routes corresponding to Tests 9,10,11,12, and 13 (Fig. 1.1) and these are shown in Figs. 5.1 and 5.2. The averages were made over 4990 samples in each case, corresponding to a route distance of 250 m which was thought to be sufficient to remove irrelevant local variations. It was found that the values calculated from (5.2) were about 2 dB lower than those calculated from (5.1), with a standard deviation of the order of 0.5 dB for the residual difference. These quantities were estimated by visual inspection of the over-

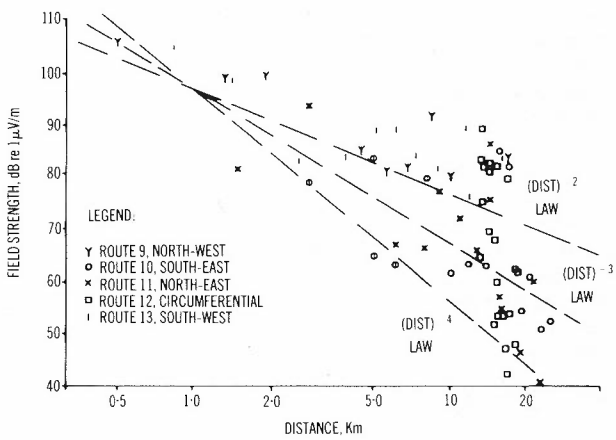


Fig. 5.1 - Field strengths at various distances, obtained by averaging decibel values

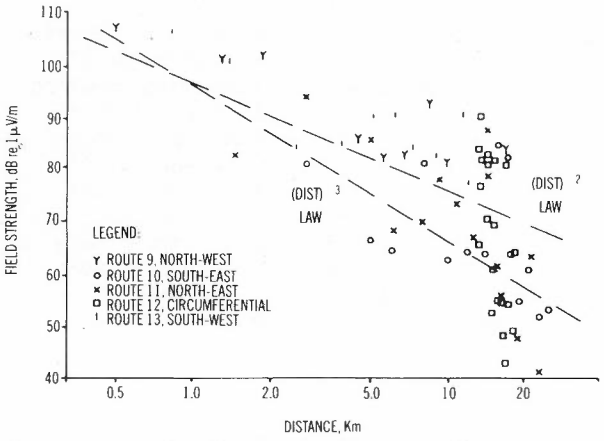


Fig. 5.2 - Field strengths at various distances, obtained by averaging voltages

laid plots. Thus it is valid to use either averages of dB values or averages of field strengths to represent local values. The data were plotted on logarithmic distance scales to facilitate comparison with power-law attenuation models. Free-space "attenuation" due to the inverse square law yields 20 dB per decade of distance, while other models predict a basic 30 dB or 40 dB per decade, corresponding to inverse-cube and inverse fourth power laws respectively. The latter is derived from diffraction theory.

A new model was devised, based on the idea that the received field is the superposition of many components scattered by objects within a characteristic local area. The total power entering this area is taken to be proportional to the incident energy flux as determined by an inverse square law, and to the projection of the area in the direction of the incident energy. Details of the model are given in Appendix III which shows that the projected area depends on the range, the relative altitudes of the transmitter and receiver, the slope of the ground near the receiver, and on the radius of the earth. Under most conditions the predominant term contains the altitudes and the inverse of the cube of the range; thus it is an inverse-cube law. Topographic data for application of this model were obtained from 1:50000 scale contour maps of Melbourne which permitted determination of altitudes of the ground surface and slopes.

5.1 Analysis of results of long distance runs

Figs. 5.1 and 5.2 in which the raw data are plotted do not suggest any simple relationship. Absolute values predicted by an inverse square law (with 100 W transmitter power and 2 dB antenna gain) are also shown. The absolute agreement looks quite good for small ranges. Use of the illumination model of Appendix III to correct the observations to take into account the altitudes and slopes on all the data did not give any significant reduction in the scatter, although the field strengths predicted by the model do show a scatter diagram of similar character to the data observed (Fig. 5.3). Two modifications of the model were also incorporated, but did not give any apparent improvement in detailed predictions:

(a) allowance of 10 dB additional loss for locations which were separated from the transmitter by an intervening high land obstructing the line of sight (Ref. 5) and,

(b) elimination from consideration of those locations at which the local slope blocked direct illumination from the transmitter.

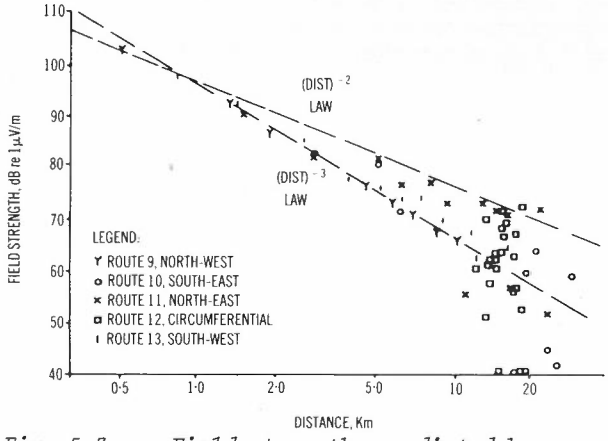


Fig. 5.3 - Field strengths predicted by illumination model. The values have been scaled to give an absolute value at 1 km equal to that predicted by free space conditions

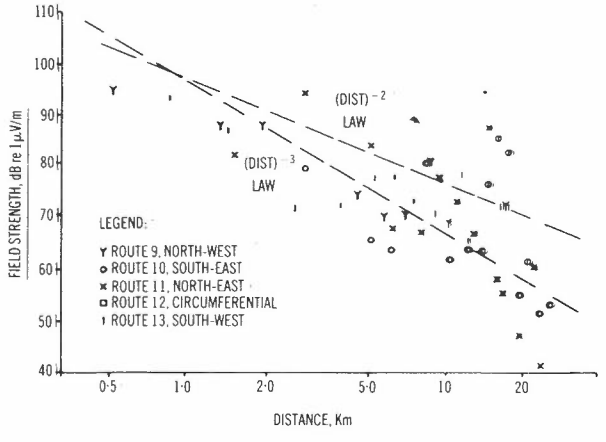


Fig. 5.4 - Superposition of field strengths of radial routes, with those of Test 9 (north-west) and Test 13 (south-west) attenuated by 12 dB

Inspection of the plots for the radial routes showed that correction by the addition of 12 dB attenuation to the northwest and the southwest route results gave improved clustering about trend lines (Fig. 5.4). A difference of this order between the results for different directions is consistent with the transmitter antenna having different gains in the different directions, possibly due to the effects of city buildings. Additional evidence for this effect is given by Fig. 5.5 in which the data for the large circumferential route, Test 12, are plotted against direction. This figure includes also points from other routes at similar radius.

The plots of Fig. 5.4 which include the compensation for different directions indicate a dependence on distance D^n where n is about

-2.5, i.e. significantly smaller than the values of -3 to -4 predicted by various models. The compensated plots show an overall pattern even more similar to that predicted by the illumination model (Fig. 5.3). The absolute values, based on a transmitter power of 100 W and omnidirectional gain of 2 dB (half-wave dipole), yield values in quite good agreement with the actual received fields (Fig. 5.2) although the dependence on distance of this model is greater than that observed.

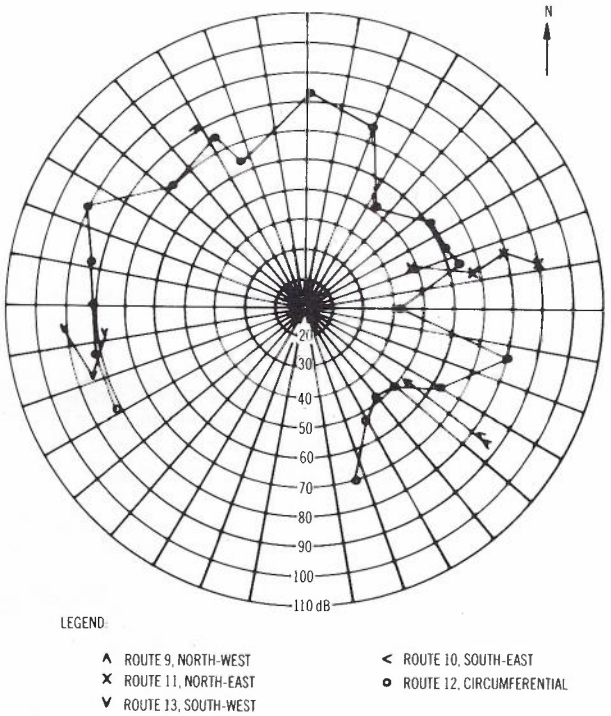


Fig. 5.5 - Polar plot of field strengths of points at about 18 km range

The scatter in the predictions is not as great as that observed. This is consistent with the model excluding factors which have significant contributions, for example street orientation. A consequence is that such a simple model would not permit good prediction of the proportion of locations with unsatisfactory performance even if it did give good prediction of dependence on distance.

5.2 Conclusions

The variations of field strength with receiver location depend in a complex manner on many factors (a) ... (g) as listed at the beginning of this section. Simple models of power law dependence do not fit well even when local terrain altitude and slope are taken into account. Each receiver location would need assessment with regard to the objects available for diffraction and reflection, including hills between the transmitter and receiver, the nature of vegetation and buildings on such hills, local street orientation, and also objects in the vicinity of the transmitter which might influence the radiation toward that location.

There is evidence that the radiation from the transmitter is about 12 dB different in different directions. This could be associated with the presence of tall buildings.

For distances up to about 10 km the expected value of the received field is quite well estimated by the illumination model of Appendix III. However, accurate determination of service area, including the proportion of locations in which service would be unsatisfactory, is not likely to be possible with models which do not include much more local detail than appears practical to include. The extraction of quite rudimentary terrain data from a contour map proved to consume so much time and to require so much judgement that direct measurement is strongly recommended.

6. DIVERSITY ASSESSMENT

A selection diversity operation was simulated by taking as the signal strength the larger of two field strength samples separated by half a wavelength. An assumption involved is that the field strength measured after the vehicle moves $\lambda/2$ is the same as would be measured on an antenna mounted $\lambda/2$ further forward on the vehicle. This will be true only if the vehicle itself does not substantially modify the field and that this is independent of the presence or absence of a second antenna. This assumption would only be approximately true in practice.

The results are shown in Figure 6. together with the distribution of the unmodified signal strength. The improvement is consistent with a theoretical prediction (Ref. 5).

It is seen that diversity gives a 6-10 dB improvement at the 1% level and 11-12 dB at the 0.1% level, compared with theoretical values of 10 dB and 15 dB based on independent Rayleigh fading signals.

7. OVERALL CONCLUSIONS

The measurements of field strength show good repeatability, independent of time of day and of weather conditions, provided local fast-fading effects are averaged out. With averages taken over 1500 samples, i.e. 75 m, the standard deviation is about 1 dB. It was found to be convenient and reliable to use averages over 5000 samples, i.e. 250 m to give estimates to characterise a locality.

Local means measured in streets oriented radially with respect to the transmitter were about 6-8 dB greater than those measured in streets oriented circumferentially, i.e. across the direction of arrival of the direct signal.

The effects of fast multipath fading were separated from slow local variations, such as those attributable to shading by buildings, by high-pass and by low-pass filtering. The fast-fading distributions were compared with those to be expected if the received signal were consistent with the Rayleigh fading model. While the distributions do show considerable similarity in many cases, in general there are significant departures, especially for streets running radially with respect to the transmitter. The lower and upper extremes of the observed distributions tend to be compressed as compared

with the Rayleigh model. While the discrepancy means that the Rayleigh model cannot be used to predict accurately the incidence of deep fades, it does give a pessimistic or worst case prediction.

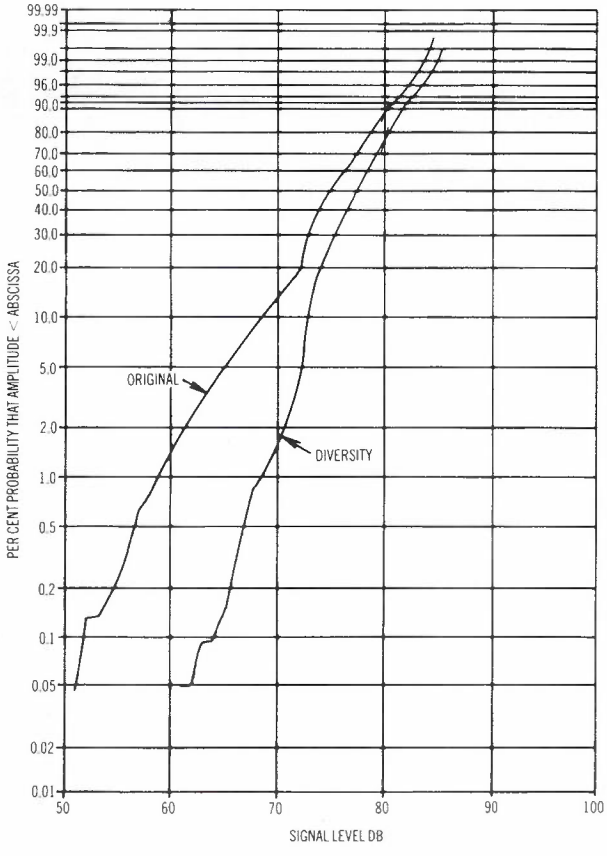


Fig. 6 - Effect of diversity in reducing probability of low signal amplitudes. Receiver antennas λ/2 apart

Likewise, the slow fading, which in simple models follows a log-normal distribution, was found to have compressed tails in practice. The departure from normality usually occurred for values which occurred with probability about 2% or less. Typical standard deviations were about 7 dB, but this must depend on locality. Use of the normal distributions with the same variances to predict the occurrence of small values would result in pessimistic estimates of system performance.

Spectra of the envelope-squared of the field strength were generally consistent with expectations. These spectra tended to be band-limited to twice the maximum doppler frequency, that is to a frequency corresponding to a spatial frequency with half the wavelength of the carrier. Spectra for radially oriented streets showed peaks near this frequency, corresponding to predominant propagation along the street in both directions. Spectra for circumferentially oriented streets tended to be attenuated, especially in the high frequency region. The differences between the two classes of spectra are attributable to the existence of a strong direct signal from the transmitter being present in the radial streets and not in the other streets. The spectrum of the envelope-squared of the signal received when the vehicle was stationary showed most energy at very low frequencies as expected. The existence of any energy at other than zero frequency is attributable to scattering from other vehicles in motion.

The effects of distance were compared with simple power-law models, and with a new "illumination" model. While the models do give trends and distributions similar to those observed, detailed predictions are prone to serious errors associated with local details such as the vegetation or material of buildings on intervening hills, street orientation, the number of scattering objects, the presence of distant reflecting paths, and the slope of the ground. The received signal strengths tended to depend also on the direction of the location of the receiver with respect to the transmitter, suggesting that city buildings could be contributing to a modification of the transmitting antenna pattern.

In the presence of such contributing factors it would require much more detailed description of the topography than is available and many times more effort than has been expended, to produce models adequate for predicting the quality of service to be expected at a given point. We do not think that this additional effort is justifiable for the objectives of Telecom Australia.

Overall the work has given a fairly good understanding of the nature of the received signal. However, the possibility of making reliable predictions of quality of service in realistically varied terrain appears to be fairly remote. The cost-effective way to evaluate the service area of a particular transmitter location appears at this time to be to carry out a program of measurements, including noise measurements. In the light of results on repeatability it is not necessary to carry out multiple traverses of any of the routes. It appears that the following survey plan would provide a reasonable indication of coverage:

(a) Four radial routes from the centre of population to be served, evenly spaced.

(b) A circumferential route at large radius from the centre of population. It is believed that a larger radius than that originally used could be preferable; about 30-40 km would be reasonable.

Significant improvements could be obtained by the use of receiver space diversity. To improve performance in areas with poor reception or transmission the use of multiple, synchronized transmitters at different locations could be useful, with or without space diversity in the mobile. Such a procedure would introduce additional multipath effects at most locations and would require further study, but the existing single-station systems already operate with multipath propagation.

The next generation of mobile radio systems is expected to use digital techniques, and a

number of different problems will arise with these. In particular, the effects of multipath propagation on pulse shapes and interference, and methods of combatting these will require study in the near future.

8. REFERENCES

1. D.C. Cox, "Delay-Doppler Characteristics of Multipath Propagation at 910 MHz in a Suburban Mobile Radio Environment", IEEE Trans. Ant. Prop. AP-20, September 1972, pp 625-635.

2. B.R. Davis, "Random FM in Mobile Radio with Diversity", IEEE Trans. on Comm. Tech. COM-19, December 1971, pp 1259-1267.

3. R.H. Clarke, "A Statistical Theory of Mobile Radio Reception", Bell Syst. Tech. J., Vol. 47, July 1968, pp 957 - 1000.

4. W.R. Young and L.Y. Lacey, "Echoes in Transmission at 450 MHz from Land-to-Car Radio Units", Proc. IRE 38, March 1950, pp 255-258.

5. W.C. Jakes, "Microwave Mobile Communications", John Wiley, 1974.

6. Allsebrook, K., "An investigation into the propagation of radio waves at VHF and UHF within certain British cities", British Aircraft Corp. Document Ref. No. ESS/ES 356 (Limited distribution; issued by British Aircraft Corporation, Electronic Systems Group, P.O. Box 77, Filton House, Filton, Bristol BS99 7AR England).

7. Okumura, Y., Ohmori, E., Kawano, T., Fukuda, K., "Field Strength and its Variability in VHF and UHF Land Mobile Service", Rev. Elec. Comm. Lab., Vol. 16, Nos. 9-10, Sept.-Oct. 1968, pp 825-873.

8. Allesbrook, K. and Parsons, J.D., "Mobile radio propagation in British Cities in the V.H.F. and U.H.F. Bands", Proc. IEE, Vol. 124 N2, Feb. 1977, pp 95-102.

9. Steel, J. and Court R.A., "Measurement of Field Strength for Mobile Services at V.H.F. and U.H.F. for the City of Melbourne", Electronics letters, Vol. 13 N24, 24 Nov. 1977, pp 710-712.

10. Bullington, K., "Radio Propagation Fundamentals", Bell Syst. Tech. J., Vol. 36 No. 3, 1957, pp 593-626.

11. Lawson, I.C., Hicks, P.R., and de Jong, H., "Field Strength Survey Techniques for Mobile Radio Services", ATR Vol. 15 No.1, 1981, pp 44-51.

12. H. Cramer, "Mathematical Methods of Statistics", Princeton U.P. 1946.

9. ACKNOWLEDGEMENTS

This work was supported in part by Telecom Australia and it is a pleasure to acknowledge the assistance of the staff of the Radio Systems Section (Research Department) in discussions and in data gathering. Valuable information was also received from The Plessey Co. Ltd., and Dr J.D. Parsons of the University of Birmingham.

APPENDIX I

AUTOCORRELATION FUNCTION OF THE LOGARITHM OF THE ENVELOPE

The received signal is of the form:

$$s(t) = r(t) \cos \{\omega_o t + \theta(t)\}$$

$$= x(t) \cos \omega_o t - y(t) \sin \omega_o t$$

where $r(t)$ is the envelope, $\theta(t)$ the instantaneous phase and $x(t)$ and $y(t)$ the corresponding cartesian components of the signal phasor. Clarke's model of mobile radio propagation has $x(t)$ and $y(t)$ jointly gaussian with correlation function $\psi(\tau) = J_o(2\pi f_m \tau)$. The joint distribution of $r_1 = r(t)$ and $r_2 = r(t+\tau)$ is (Ref. 5)

$$p(r_1, r_2) = \frac{r_1 r_2}{\sigma^4 \lambda^2} \exp\left[\frac{r_1^2 + r_2^2}{2\sigma^2 \lambda^2}\right] I_o\left[\frac{\psi r_1 r_2}{\sigma^2 \lambda^2}\right] \quad (I.1)$$

where

$$\sigma^2 = \langle x^2 \rangle = \langle y^2 \rangle$$

$$\psi(\tau) = \langle x(t)\, x(t+\tau) \rangle / \sigma^2 = \langle y(t)\, y(t+\tau) \rangle / \sigma^2$$

$$\lambda^2 = 1 - \psi^2$$

$I_o(x)$ = modified Bessel function of zero order.

The autocorrelation function of $r(t)$ has been derived (Ref. 5) and is given by

$$R_r(\tau) = \langle r(t)\, r(t+\tau) \rangle$$

$$= \frac{\pi}{2} \sigma^2 \left(1 + \frac{\psi^2}{4} + \frac{\psi^4}{64} + \frac{\psi^6}{256} + \ldots\right) \quad (I.2)$$

and the coefficient of ψ^{2k} is $\left[\frac{\Gamma(k-\frac{1}{2})}{2\Gamma(\frac{1}{2})k!}\right]^2$.

The mobile field intensity recording receiver makes measurements of envelope in decibels:

$$v(t) = 20 \log_{10} r(t) = 2c \ln r(t) \text{ decibels}$$

where

$$c = 10/\ln 10 = 4.343.$$

The autocorrelation function of v(t) is given by:

$$R_v(\tau) = \langle v(t)\, v(t+\tau)\rangle$$

$$= \int_0^\infty \int_0^\infty 4c^2 \ln r_1 \ln r_2\, p(r_1, r_2)\, dr_1 dr_2.$$

This integration can be performed by expanding $I_o(x)$ as a power series. The result is:

$$R_v(\tau) = v_o^2 + c^2 \sum_{k=1}^{\infty} \psi^{2K}/k^2 \qquad (I.3)$$

where

$$v_o = \langle v(t)\rangle = c\,\{\ln(2\sigma^2) - \gamma\}$$

γ = Euler's constant = .57722.

Note that the average of the envelope values in dB (v_o) is $c\gamma$ (2.507 dB) less than the RMS envelope expressed in dB.

The power spectrum of v(t) is the Fourier transform of $R_v(\tau)$ which for $\psi(\tau) = J_o(2\pi f_m \tau)$ gives

$$S_v(f) = v_o^2\, \delta(f) + \frac{c^2}{\pi^2 f_m} K\left[\sqrt{1-f^2/4f_m^2}\right] + \ldots, \qquad (I.4)$$

K being defined in Equation (4.2).

The first term is the DC component, the second term has a logarithmic infinity at f=0 and is bandlimited to $2f_m$, the third term corresponds to the Fourier transform of

$$\frac{1}{4} c^2 J_o^4 (2\pi f_m \tau)$$

and is finite at f=0 and bandlimited to $4f_m$, and succeeding terms are also finite at f=0. Fig. I.1 shows this power spectrum.

The asymptotic form at high frequencies can be found by considering the behaviour of $R_v(\tau)$ for small τ.

$$R_v'(\tau) = -2c^2 \frac{\psi'(\tau)}{\psi(\tau)} \ln[1-\psi^2(\tau)]$$

$$\sim -8\pi^2 c^2 f_m^2\, \tau \ln|\tau| \quad \text{for small } \tau.$$

Hence the asymptotic spectrum at high frequencies is:

$$S_v(f) \sim c^2 f_m^2 / f^3 \quad \text{for large } f. \qquad (I.5)$$

Figure (I.1) also shows this high frequency asymptote.

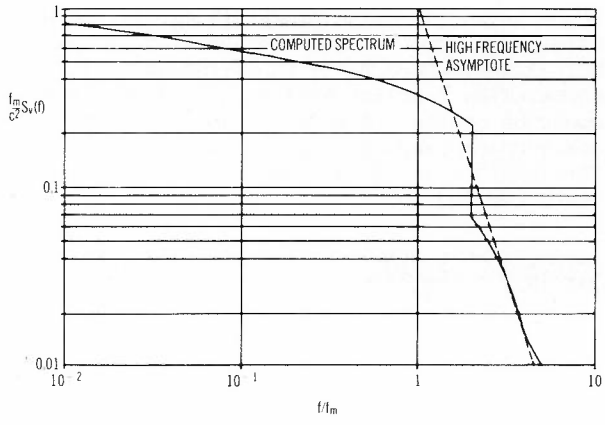


Fig. I.1 - Power Spectrum of log envelope

APPENDIX II

EFFECT OF FILTERING

In several instances we wish to estimate the local mean of the fast-fading. This is done by averaging over a large number of samples.

Let $v(t) = 20 \log_{10} r(t)$ decibels and form an estimate of the mean as

$$\hat{v}_o = \frac{1}{2k+1} \sum_{i=-k}^{k} v(ih) \qquad (II.1)$$

where v(ih) are the samples of v(t) taken at time intervals of h. Actually the samples are taken at spatial intervals d (usually 5 cm) and hence for a mobile vehicle travelling at velocity v we have d = hv.

For the purposes of this analysis we will assume that the sum can be approximated by an integral

$$\hat{v}_o = \frac{1}{T} \int_{-T/2}^{T/2} v(t)\, dt \qquad (II.2)$$

where T is the averaging time and T = (2k+1) h.

Now $\hat{v}_o$ is a random variable and if v(t) is stationary the following are true

$$\langle \hat{v}_o \rangle = v_o \text{ (the true mean)} \qquad (II.3)$$

$$\text{Var}\,[\hat{v}_o] = \frac{2}{T} \int_0^T (1-\tau/T)\, C_v(\tau)\, d\tau \qquad (II.4)$$

where $C_v(\tau)$ is the autocovariance function of v(t).

When $C_v(\tau)$ is integrable this gives a simple result

$$\text{Var}\,[\hat{v}_o] \sim \frac{\text{Var}\,[v(t)]}{2B_n T} \quad \text{for large } T \qquad (II.5)$$

where B_n is the noise bandwidth of v(t). This result was used to estimate Var $[\hat{v}_o]$ by taking $B_n = 1/2\tau_o$, where τ_o is the first zero of the correlation function $\psi(\tau) = J_o(2\pi f_m \tau)$. [This would be correct if v(t) had an ideal low pass spectrum of bandwidth B_n, for which $R_v(\tau) = \text{Var}[v(t)]\ \text{sinc}\ 2B_n\tau$]. This approximate method therefore led to

$$\text{Var}[\hat{v}_o] \sim \frac{5.570^2}{T/\tau_o} \qquad (II.6)$$

However v(t) does not have an ideal low pass spectrum and $C_v(\tau)$ is not integrable.

From Appendix I we have

$$C_v(\tau) = c^2 \sum_{k=1}^{\infty} \frac{1}{k^2} J_o^{2k}(2\pi f_m \tau) \qquad (II.7)$$

Hence

$$\text{Var}[\hat{v}_o] = \frac{2}{T} \int_o^T (1-\tau/T)\, c^2 \sum_{k=1}^{\infty} \frac{1}{k^2} J_o^{2k}(2\pi f_m \tau)\, d\tau$$

$$\sim \frac{2c^2}{X}\left[\frac{1}{\pi}\ln X + A + 0\left(\frac{\ln X}{X}\right)\right] \qquad (II.8)$$

where

$$X = 2\pi f_m T \quad \text{and} \quad A \approx 0.90.$$

For d = 0.05 m and λ = 0.60 m equations (II.6) and (II.8) become

$$\text{Var}[\hat{v}_o] \sim \frac{143}{N} \qquad (II.6a)$$

$$\text{Var}[\hat{v}_o] \sim \frac{50 + 23 \ln N}{N} \qquad (II.8a)$$

These are plotted in Figure (II.1) and it can be seen that the simplistic estimate (II.6a) is only slightly in error for averages involving up to 10^4 samples.

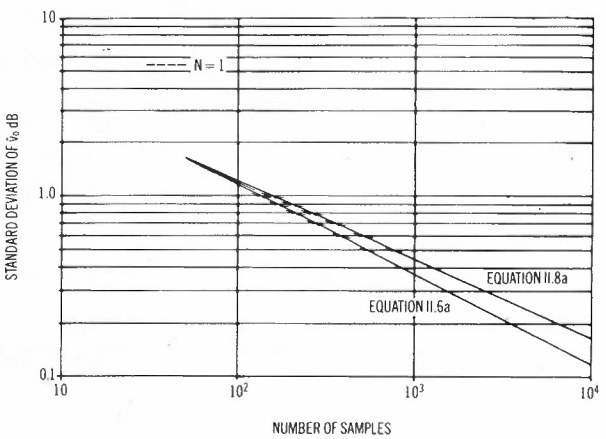


Fig. II.1 - Standard deviation of the estimated mean

APPENDIX III

AN ILLUMINATION MODEL

In many cases the received field may be considered to be composed of the superposition of many scattered components most of which are produced by scatterers in the vicinity of the receiver. Then the power entering this vicinity is a major factor in determining the resultant sum of scattered components. In the absence of shading by major obstacles, of diffraction effects or surface wave phenomena, and of atmospheric effects, the power W_A entering a given ground area A at distance D will depend on D via an inverse-square law, and on the orientation of A to the direction of arrival (Fig. III.1). The incident radiation density W is then

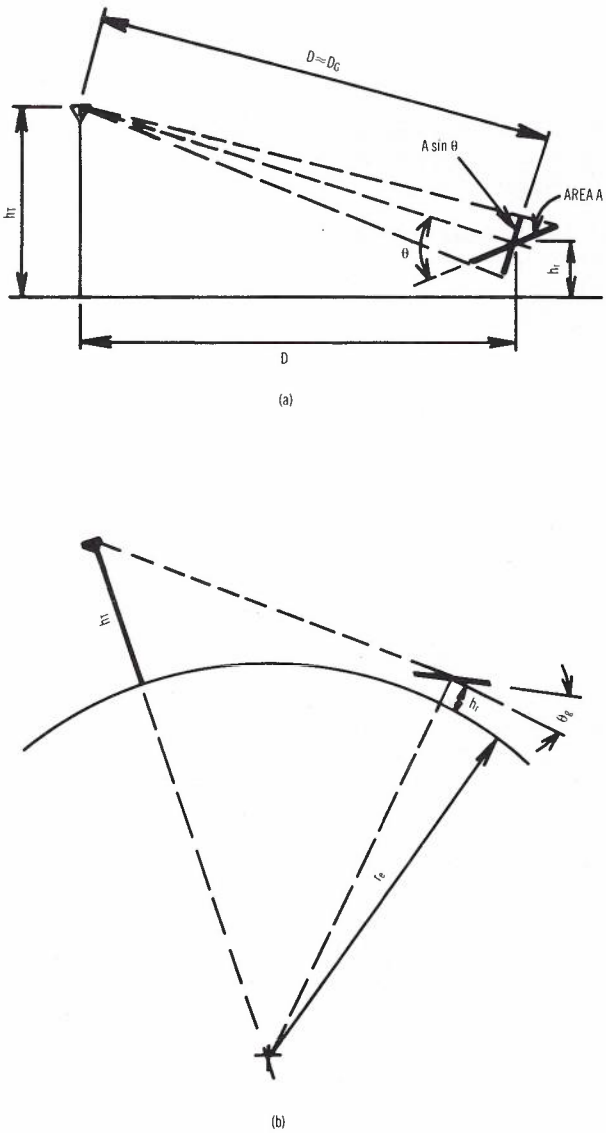


Fig. III.1 - Geometry for Illumination Model
(a) Model with flat earth, defining symbols
(b) Curvature of Earth included

$$W = \frac{P_T}{4\pi D^2} \quad w\ m^{-2}$$

where P_T is the transmitted power, referred to an equivalent isotropic radiation. The projection of the local area A which determines the intercepted power is A sin $\theta \approx A\theta$ so that the intercepted power W_A is

$$W_A = \frac{P_T A}{4\pi} \cdot \frac{\theta}{D^2} \qquad (III.1)$$

Taking into account the transmitter height h_T, the receiver height h_r and the effective radius of the earth, r_e (taken to be 4/3 of the actual radius because of the atmospheric density gradient), we find

$$\theta \doteq \frac{h_T - h_r - D^2/2r_e}{D} + \theta_g$$

where θ_g is the local slope of the ground.

Then from equation III.1 we have

$$W_A = C\left(\frac{h_T - h_r}{D^3} + \frac{\theta_g}{D^2} - \frac{1}{2Dr_e}\right) \qquad (III.2)$$

This expression is applicable only when $W_A > 0$; if $W_A \leq 0$ then the nett slope of the ground is away from the transmitter and the incident ray enters from below A, i.e. the transmitter is not visible from the receiver. The received illumination must then come via indirect paths, diffraction or reflection not being incorporated in the model.

APPENDIX IV

RECOMMENDATIONS FOR DATA GATHERING

(a) Observations with moving receiver

Data from closed paths of small periphery, with repeated traverses and accurate (to about 2 metres if possible) event markers should be collected. These should permit checks of repeatability, and of effects of traffic variations, street orientation, time of day and weather. They are also required to provide sets of data within which slow fading effects would be the same but fast fading be substantially uncorrelated.

Specifically, it is recommended that the following paths should be traversed at the times shown. Where there are successive traverses of the same path they should if possible be driven with lateral displacement of the vehicle by about 1 metre, or in a different lane.

	Path	Schedule
1.	Melbourne: Flinders St, Spring St, Victoria St, King St, Flinders St.	On one dry day: 4 traverses between 11 am and 12 noon 4 traverses between 4 and 5 pm 4 traverses between 11 and 12 pm (if possible) One one wet day: 4 traverses between 11 am and 12 noon
2.	same as 1. but in reverse direction	One dry day, 4 traverses between 11 am and 12 noon
3.	Jordanville (Map 61): Highbury Rd, Huntingdale Rd, Outlook Rd, Quaintance St, Jubilee Rd, Highbury Rd.	One one dry day: 4 traverses in any one hour

(b) Stationary Observations

These are required to study time variable effects in the environment.

4.	Stationary, close to a busy city intersection	In a busy period, for 10 minutes. Observe noise only for a further 5 minutes.
5.	Same as 4. and under similar weather conditions	In a very quiet period, e.g. Sunday morning, for 10 minutes. Observe noise only for a further 5 minutes.

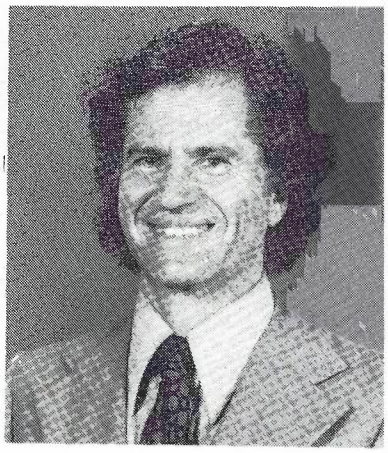
(c) Long-distance Runs

The following data is needed to study the phenomena associated with "slow" variations:

6.	G.P.O. to Tullamarine airport via main route.	Any time. Note on record major features such as cuttings, hills, underpasses, major buildings as well as location via cross roads.
7.	G.P.O. to Dandenong via main route.	As for (6)

8. G.P.O. to Warrandyte P.O. via High St, Doncaster Rd, Williamsons Rd, Warrandyte Rd.	As for (6)	10. Laverton Railway Station, via Princes Highway.	As for (6)
9. A circumferential route at a radius of about 15 to 20 km from the transmitter.	As for (6)		

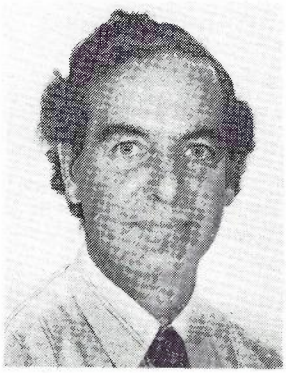
BIOGRAPHIES



ROBERT E. BOGNER was born in Melbourne in 1934, and gained the B.E. (Elec.) and M.E. from the University of Adelaide in 1956 and 1959. He worked with the Australian Postmaster-General's Department (now Telecom Australia) as a research engineer from 1956 through 1961. He was Lecturer and Senior Lecturer in Electrical Engineering at the University of Queensland (1962-1966) and at the Imperial College of Science and Technology, University of London (1967-1973), gaining the Ph.D in 1973 for work on phase processing of angle modulated signals. He is currently Professor of Electrical Engineering at the University of Adelaide.

His major field is communication engineering and he has contributed to this in electroacoustics, speech communication, human factors, microwave modelling, modulation, and signal processing. A subsidiary interest is in unusual electrical machines. Current efforts are in speech signal processing and ultrasonic techniques applied to automated sheep shearing. A number of consultancies with industry, including two periods at Bell Laboratories have been welcome experiences related to these interests.

Professor Bogner is married, with three children.



BRUCE R. DAVIS has been with the University of Adelaide since 1964 and at present is a Senior Lecturer in Electrical Engineering. His research interests are in the field of communication systems. During 1970 he was with Bell Laboratories, Holmdel, New Jersey, studying various aspects of mobile radio communications, and again in 1977 when he was involved in satellite system research. He is a member of the IEEE and holds the following degrees: B.E. 1960, B.Sc. 1963, Ph.D 1969, University of Adelaide.

Performance Models for a Class of Non-Hierarchical Networks

L.T.M. BERRY
A.J. COYLE

Applied Mathematics Department
University of Adelaide

It is shown that the chain flow dimensioning model (Ref. 1), originally developed for hierarchical alternative routing telephone networks, can be conveniently extended to apply also to non-hierarchical networks.

1. INTRODUCTION

At the present time, dimensioning techniques are mainly based on hierarchical structures. With the proposed introduction of new digital switching and transmission systems, however, there has grown an awareness of the potential advantages of incorporating some degree of non-hierarchical routing. The possibility of using a non-hierarchical network structure for the Australian International Digital Network is discussed in Ref. 2. The authors point to a number of possible benefits such as an expected improvement in origin-destination grade of service, an increased robustness and a greater flexibility to accommodate different traffic peaks more economically. Thus there is a need to develop new dimensioning models for non-hierarchical networks.

A number of dimensioning models, suitable for application to hierarchical networks, have been extensively documented in the literature. It is natural to examine the extent to which they may be modified to incorporate non-hierarchical routing. In this paper we examine some simple non-hierarchical networks with the view to determining whether the chain flow dimensioning model has potential, when used in an iterative manner, for accurately estimating individual chain flows.

Throughout the paper, the term "link" is used to denote a group of channels or circuits between two exchanges. An exchange at which a group of subscribers initiate calls is referred to as an origin. These calls are carried on a sequence of links, called a "chain" or route, between the origin and a destination exchange. The mean carried traffic on a link is called a "link flow" whereas the mean number of calls carried on a chain (between a given origin and destination) is called a "chain flow". A network is said to be non-hierarchical if every exchange may function as a transit exchange. We restrict our discussion to the case where one-way channels are provided on links.

A class of non-hierarchical networks is considered in which the alternative routes are prescribed for each O-D pair. That is, it is assumed that data tables exist, giving for each O-D pair an ordered set of chains. In particular, we consider the case when overflow traffic from one O-D pair is permitted to use a direct route for another O-D pair.

In the next sections two problems related to the generalization of the chain flow model to non-hierarchical networks are discussed and solutions to both of these problems are found. These new methods are subsequently incorporated in an outline of a performance model suitable for application to larger networks.

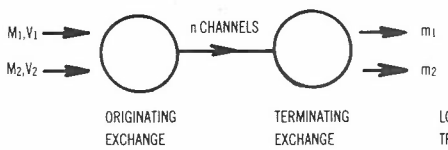


Fig. 1 - A single link offered two overflow streams.

2. DETERMINATION OF CHAIN FLOWS

Consider the lost calls cleared network of Fig. 1. It is a single link of n channels (circuits) offered two independent streams of overflow traffic with means M_1, M_2 and variances V_1, V_2 respectively. According to the chain flow model, the total flow f on the link may be computed by solving the following equation for f.

$$n = f + A\left[\frac{(M-f)^2 + V}{(M-f-1)(M-f) + V} - \frac{M^2 + V}{(M-1)M + V}\right] \quad (1)$$

where

$$M = M_1 + M_2 \quad (2)$$

$$V = V_1 + V_2 \quad (3)$$

$$A = V + 3\frac{V}{M}\left(\frac{V}{M} - 1\right) \qquad (4)$$

and

$$v = \frac{(M-f)}{6}\left[3-M+f+\sqrt{(3-M+f)^2 + 12A\left(1-\frac{f}{M}\right)^{0.1}}\right] \qquad (5)$$

An extremely fast algorithm for computing f from equations (1) to (5) is given in Ref. 3. We shall denote the total flow computed by this algorithm when a link with n channels is offered overflow traffic with mean M and variance V by FLOW (M,V,n).

Now consider the same traffic offered to the network of Fig. 2. If $n_1 \geq n_2 + n_3$ the two streams may be considered independently and their individual chain flows (mean carried traffics) computed from

$$f_1 = \text{FLOW}\ (M_1,V_1,n_2)$$

$$f_2 = \text{FLOW}\ (M_2,V_2,n_3) \quad . \qquad (6)$$

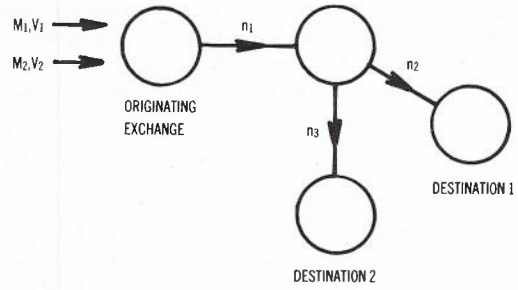


Fig. 2 - Two chains sharing a common link.

When dimensioning non-hierarchical networks it is necessary to determine *individual* chain flows when overflow traffics are offered to chains sharing a common link. (We recall that this problem does not occur when applying the chain flow model to the minimum cost design of hierarchical alternative routing networks, since the model provides a direct mapping *from* a set of chain flows, $\underset{\sim}{h}$, *to* the circuit allocation vector $\underset{\sim}{n}$. In the present application we are considering the inverse mapping problem.) In general, an arbitrary circuit allocation does not allow simple independent computation of the individual chain flows as was possible in the previous example; that is, frequently $n_1 < n_2 + n_3$. To estimate the chain flows for the network illustrated in Fig. 2, we propose the following method. We denote by CHANNEL (M,V,f) the number of channels, computed by equation (1), required to carry a flow f when a single link is offered overflow traffic with mean M and variance V where this overflow traffic stream may be the sum of a number of independent pure chance or overflow traffic streams.

1. Compute $f_1 = \text{FLOW}\ (M_1,V_1,n_2)$

$$f_2 = \text{FLOW}\ (M_2,V_2,n_3)$$

2. Set $f = f_1 + f_2$

3. Compute $\hat{n}_1 = \text{CHANNEL}\ (M_1+M_2, V_1+V_2, f)$

4. If $\hat{n}_1 \leq n_1$ the individual chain flows are: $h_1 = f_1$, $h_2 = f_2$. If $\hat{n}_1 > n_1$, the flows f_1, f_2 are decreased, *in such a way that the ratio* $f_1 : f_2$ *is kept constant*, until the computed value $\hat{n}_1$ is equal to n_1. This is achieved by simply computing the zero, x^*, of the function CHANNEL $(M_1+M_2,\ V_1+V_2,\ f_1 - x + f_2 - f_2x/f_1) - n_1$, in the interval (0, min (f_1,f_2)). In this case the chain flows are

$$h_1 = f_1 - x^* \ , \quad h_2 = f_2 - \frac{f_2}{f_1}\ x^* \ .$$

In order to test the accuracy of the above method exact solutions were computed directly from the Kolmogorov equations for the service system illustrated in Fig. 3.

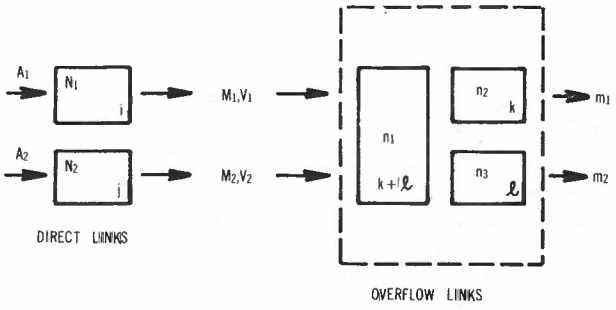


Fig. 3 - Service system representing a network with two direct circuits and two overflow routes sharing a common link.

The variables m_1, m_2 are the mean lost traffics for the two streams, and the chain flows h_1, h_2 are related by $h_i = M_i - m_i$, $i = 1,2$.

If the state (i,j,k,ℓ) represents i and j calls on the direct links of N_1 and N_2 channels and k and ℓ calls on the two overflow routes respectively, then the equilibrium probabilities $p(i,j,k,\ell)$ are found by solving

$$\sum_{i,j,k,\ell} p(i,j,k,\ell) = 1$$

and all but one of the following equations:

$$\begin{aligned}
0 = &- p(i,j,k,\ell)(A_1+A_2+i+j+k+\ell) \\
&+ p(i,j,k,\ell)A_1 \quad \{i = N_1 \text{ and either } k = n_2 \text{ or } k+\ell = n_1\} \\
&+ p(i,j,k,\ell)A_2 \quad \{j = N_2 \text{ and either } \ell = n_3 \text{ or } k+\ell = n_1\} \\
&+ p(i+1,j,k,\ell)(i+1) \quad \{i \neq N_1\} \\
&+ p(i,j+1,k,\ell)(j+1) \quad \{j \neq N_2\} \\
&+ p(i,j,k+1,\ell)(k+1) \quad \{k \neq n_2 \text{ and } k+\ell \neq n_1\} \\
&+ p(i,j,k,\ell+1)(\ell+1) \quad \{\ell \neq n_3 \text{ and } k+\ell \neq n_1\} \\
&+ p(i-1,j,k,\ell)A_1 \quad \{i \neq 0\} \\
&+ p(i,j-1,k,\ell)A_2 \quad \{j \neq 0\} \\
&+ p(i,j,k-1,\ell)A_1 \quad \{i = N_1 \text{ and } k \neq 0\} \\
&+ p(i,j,k,\ell-1)A_2 \quad \{j = N_2 \text{ and } \ell \neq 0\} \ , \qquad (7)
\end{aligned}$$

for integer values $i \in [0,N_1]$, $j \in [0,N_2]$, $k \in [0,n_2]$, $\ell \in [0, \min(n_3,n_1-k)]$ and $n_2 \leq n_1$, $n_3 \leq n_1$, $n_2 + n_3 \geq n_1$. The conditions in brackets are required to be satisfied for the inclusion of the adjacent terms.

The mean lost traffic can be computed from

$$m_1 = A_1 \sum_j \sum_k \sum_\ell p(N_1,j,k,\ell)\delta_1$$

where

$$\delta_1 = \begin{cases} 1 & \text{if } k = n_2 \text{ or } k+\ell = n_1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$m_2 = A_2 \sum_i \sum_k \sum_\ell p(i,N_2,k,\ell)\delta_2 \qquad (8)$$

where

$$\delta_2 = \begin{cases} 1 & \text{if } \ell = n_3 \text{ or } k+\ell = n_1 \\ 0 & \text{otherwise.} \end{cases}$$

As the computational time, and computer storage requirements, for solving equations (7) are $O(N^2)$ where

$$N = (N_1+1)(N_2+1)\left[(n_1+1)(n_2+1) - \frac{n_2(n_2+1)}{2} - \frac{(n_1-n_3)(n_1-n_3+1)}{2}\right]$$

we are restricted, at the present time, to cases where N is not much larger than about 300. (The central processor time on the VAX11/780 for N = 288 was 464 seconds). Nevertheless, this allows examination of the region where the model is likely to be least accurate, viz. when N_1, N_2, n_1, n_2, n_3, A_1, A_2 are small.

A total of 45 sets of values N_1, N_2, n_1, n_2, n_3, A_1, A_2 were chosen, to cover a wide range of overall grade of service, and values of m_1, m_2 then computed by both the exact equations and the model. In contrast to the exact method, the computational time for the model is independent of the size of N_1, N_2, n_1, n_2, n_3 and central processor time is less than 0.1 sec. The mean absolute error for m_1, m_2 was 0.040 erlangs and the variance of this error was 0.0036. Some typical results are shown in Table 1.

TABLE 1 - A Comparison of Results

N_1	N_2	n_1	n_2	n_3	A_1	A_2	h_1	h_2	m_1	m_2	
1	1	1	1	1	2.0	2.0	0.3511	0.3511	0.9821	0.9821	(1)
M_i		V_i					0.3371	0.3371	0.9962	0.9962	(2)
1.300		1.500							0.014	0.014	(3)
1.300		1.500					4.0	4.0	1.4	1.4	(4)
3	1	3	2	2	5.0	5.0	0.9755	1.3440	1.6728	2.7227	
							1.0268	1.3628	1.6215	2.8038	
2.648		3.668							0.051	0.019	
4.167		4.663					5.3	1.4	3.1	0.7	
2	1	2	1	2	6.0	4.0	0.6306	0.9857	3.6894	2.2143	
							0.6118	1.0814	3.7082	2.1186	
4.320		5.294							0.019	0.096	
3.200		3.626					3.0	9.7	0.5	4.3	
3	2	4	3	2	2.0	2.0	0.3900	0.6058	0.0310	0.1942	
							0.3968	0.5487	0.0242	0.2513	
0.421		0.592							0.007	0.057	
0.800		1.049					1.7	9.4	22.0	29.0	
4	1	3	2	1	3.75	3.75	0.6997	0.7225	0.3728	2.2381	
							0.6465	0.7051	0.4260	2.2556	
1.073		1.654							0.053	0.018	
2.961		3.367					7.6	2.4	14.0	0.8	
4	3	1	1	1	1.25	1.25	0.0255	0.0825	0.0113	0.0388	
							0.0244	0.0800	0.0124	0.0412	
0.037		0.048							0.001	0.002	
0.121		0.159					4.3	3.0	9.7	6.0	
4	3	2	2	1	6.0	4.0	0.9928	0.4268	1.8246	1.3760	
							0.9946	0.4435	1.8228	1.3593	
2.817		4.181							0.002	0.017	
1.803		2.553					0.2	3.9	0.1	1.2	

(1) Exact Values
(2) Computed from the model
(3) Absolute Error
(4) Percentage Error

It can be concluded that the model for determining chain flows is sufficiently accurate for practical purposes. The extension to more than two streams is clearly straightforward.

3. AN ITERATIVE APPROACH

With non-hierarchical routing it is possible for overflow traffic to be offered to direct links between exchanges *en route* to its destination. In this section we examine an iterative method for determining chain flows in such a network. In order that we may compare the results from the method with *exact* analytical results we are restricted to consideration of small networks of the following type.

Consider the network illustrated in Fig. 4. A Poisson stream of a_1 erlangs is offered to its direct route $O_1 - D_1$, the overflow traffic is then offered to route $O_1 - D_2 - D_1$. Overflow from this route is offered to route $O_1-O_2-D_2-D_1$. Similarly, the Poisson stream with mean a_2 is is offered to chains O_2-D_2, $O_2-D_1-D_2$ and $O_2-O_1-D_1-D_2$ respectively.

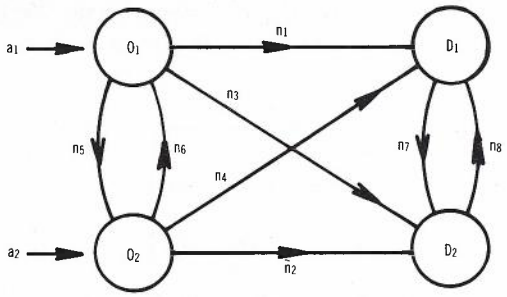


Fig. 4 - A non-hierarchical network.

We shall assume, in the following, that $n_5 \geq n_2$, $n_6 \geq n_1$, $n_7 \geq n_1+n_4$ and $n_8 \geq n_2+n_3$. Thus, the state of the network can be described by the vector (i,j,k,ℓ) where i is the number of calls on link O_1-D_1, j the number of calls on link O_2-D_2, k the number of calls on link O_1-D_2 and ℓ the number of calls on link O_2-D_1. The steady state probabilities for the network satisfy:

$$
\begin{aligned}
0 = &- p(i,j,k,\ell)(a_1+a_2+i+j+k+\ell) \\
&+ p(i,j,k,\ell)a_1 \qquad \{i = n_1 \text{ and } j = n_2 \text{ and } k = n_3\} \\
&+ p(i,j,k,\ell)a_2 \qquad \{i = n_1 \text{ and } j = n_2 \text{ and } \ell = n_4\} \\
&+ p(i+1,j,k,\ell)(i+1) \quad \{i \neq n_1\} \\
&+ p(i,j+1,k,\ell)(j+1) \quad \{j \neq n_2\} \\
&+ p(i,j,k+1,\ell)(k+1) \quad \{k \neq n_3\} \\
&+ p(i,j,k,\ell+1)(\ell+1) \quad \{\ell \neq n_4\} \\
&+ p(i-1,j,k,\ell)a_1 \qquad \{i \neq 0\} \\
&+ p(i-1,j,k,\ell)a_2 \qquad \{i \neq 0 \text{ and } j = n_2 \text{ and } \ell = n_4\} \\
&+ p(i,j-1,k,\ell)a_2 \qquad \{j \neq 0\} \\
&+ p(i,j-1,k,\ell)a_1 \qquad \{i = n_1 \text{ and } j \neq 0 \text{ and } k = n_3\} \\
&+ p(i,j,k-1,\ell)a_1 \qquad \{i = n_1 \text{ and } k \neq 0\} \\
&+ p(i,j,k,\ell-1)a_2 \qquad \{j = n_2 \text{ and } \ell \neq 0\} ,
\end{aligned}
$$

where n_1, n_2, n_3, n_4 are any non-negative integers and i,j,k,ℓ are integers such that $i \in [0,n_1]$, $j \in [0,n_2]$, $k \in [0,n_3]$, $\ell \in [0,n_4]$. The mean lost traffic can be computed from

$$m_1 = a_1 \sum_{\ell} p(n_1,n_2,n_3,\ell)$$

and

$$m_2 = a_2 \sum_{k} p(n_1,n_2,k,n_4). \qquad (10)$$

Exact values computed from the above equations were used to test the following *iterative* model.

Initially, the traffic offered to the two direct links is assumed to be just the Poisson streams with means a_1,a_2 respectively. When losses are determined from the second choice chains, these are then offered to the directs together with the Poisson traffic. New losses are computed from the direct channels and second choice routes and the procedure repeated until the numerical values so determined converge to fixed values. To describe the method in more precise detail, we denote by M_j^k, V_j^k the mean and variance of traffic *offered* to the jth choice route for O-D pair k. Thus, $M_1^1 = V_1^1 = a_1$, $M_1^2 = V_1^2 = a_2$.

1. Apply the Erlang loss formula and the Riordan-Giltay formula

$$M_3^1 = a_1 E_{n_1+n_3}(a_1), \quad M_3^2 = a_2 E_{n_2+n_4}(a_2)$$

$$V_3^1 = a_1\left(1 - M_3^1 + \frac{a_1}{n_1+n_3+M_3^1-a_1+1}\right)$$

$$V_3^2 = a_2\left(1 - M_3^2 + \frac{a_2}{n_2+n_4+M_3^2-a_2+1}\right)$$

where $E_n(a)$ denotes the Erlang loss formula.

2. The well-known equivalent random method is used to compute the *total* means and variances of the lost traffic when

(i) the combined stream with mean $a_1 + M_3^2$ and variance $a_1 + V_3^2$ is offered to the n_1 direct channels,

(ii) the combined stream with mean $a_2 + M_3^1$ and variance $a_2 + V_3^1$ is offered to the n_2 direct channels.

3. The individual mean losses from the directs (M_2^1, m_2 and M_2^2, m_1) are determined from Olsson's splitting formula (for example, the loss for stream 1 (m_1) is proportional to $V_3^1 + (M_3^1)^2/V_3^1$), and the individual overflow variances are found from the variance splitting formula due to Harris and Helm (Ref. 4). The loss from the n_2 channels, for stream 1, gives the first estimate of the loss, m_1, for O-D pair 1. Similarly the loss from the n_1 channels, for stream 2, gives the first estimate of the loss, m_2, for O-D pair 2.

4. From the values M_2^1, V_2^1, and M_2^2, V_2^2, obtained from step 3, new estimates of M_3^1, V_3^1 and M_3^2, V_3^2 are calculated by the equivalent random method

when these overflow streams are offered to the groups of n_3 and n_4 channels respectively.

5. Steps 2,3,4 are repeated until the loss estimates m_1, m_2 converge to fixed values.

Forty sets of values $n_1, n_2, n_3, n_4, a_1, a_2$ were chosen, to cover a wide range of overall grade of service, and values of m_1, m_2 computed by both the exact equations and the model. The mean absolute error for m_1, m_2 was found to be 0.0072 erlangs and the variance of this error was 0.000026. Some typical results are shown in Table 2.

Since convergence to 4 decimal place accuracy was always obtained in either 3 or 4 iterations, it may be concluded that the iterative procedure is not only very accurate but also rapid and numerically stable.

It is interesting to note that the mean percentage decrease in total lost traffic, by permitting overflow streams from second choice routes to use the additional direct routes was 38%. In some cases the decrease in lost traffic was as much as 70%.

TABLE 2 - A Comparison of Results

n_1	n_2	n_3	n_4	a_1	a_2	m_1	m_2	
1	1	1	1	2.200	2.200	0.8474	0.8474	(1)
						0.8499	0.8499	(2)
						0.0025	0.0025	(3)
						0.029	0.029	(4)
3	1	2	1	5.500	2.200	1.7073	0.7306	
						1.6973	0.7422	
						0.0100	0.0116	
						0.58	1.59	
3	2	1	1	2.053	1.400	0.1111	0.0669	
						0.1017	0.0607	
						0.0094	0.0062	
						8.6	9.3	
3	3	2	1	5.500	4.400	1.4560	1.3808	
						1.4550	1.3688	
						0.0010	0.0120	
						0.07	0.87	
4	1	3	1	7.000	1.129	1.4854	0.2131	
						1.4742	0.2310	
						0.0112	0.0179	
						0.8	8.4	
4	3	1	1	5.500	4.400	1.5619	1.2138	
						1.5488	1.1978	
						0.0131	0.016	
						0.9	1.3	
4	4	2	3	6.600	7.700	1.8135	1.8769	
						1.7989	1.8818	
						0.0146	0.0049	
						0.8	0.3	

(1) Exact Values
(2) Computed from the model
(3) Absolute Error
(4) Percentage Error

4. AN APPROACH FOR LARGER NETWORKS

We now, very briefly, consider the problem of how to combine the ideas of the previous two sections to give a practical model for determining chain flows in larger non-hierarchical networks. We restrict discussion to the class of networks for which data tables exist, giving for each origin-destination pair an ordered set of permissible chains. The following comment is included to give an indication of the possible direction of future research.

Given the routing strategy for a particular network and a circuit allocation vector $\underset{\sim}{n}$, part of the dimensioning problem of interest during a network optimisation is to determine the chain flow vector $\underset{\sim}{h}$ corresponding to $\underset{\sim}{n}$. The following outline, based on the methods of the previous sections, answers this question *in general terms*:

1. Start by determining M_2^k, V_2^k, for each O-D pair k, when the initial offered traffic alone is offered to its direct route.

2. Consider second choice routes, for each O-D pair k, and determine the flows h_2^k on the link with *least* number of channels. This calculation is performed as though the overflow streams (M_2^k, V_2^k) were offered independently to the network.

3. Re-adjust the chain flows h_2^k by considering common links belonging to different chains; the ratios of the chain flows on the common links is kept constant and the flows are decreased as necessary to meet the channel feasibility condition viz. $\underset{\sim}{h} \rightarrow \hat{\underset{\sim}{n}}$ where $\hat{n}_i \leq n_i$. $\hat{\underset{\sim}{n}}$ is computed by the chain flow model and n_i is the actual channel allocation to link i.

4. Repeat the previous step for 3rd and subsequent choice routes.

5. Re-compute the losses from direct routes allowing for computed overflow traffic which is also offered to these links.

6. Repeat steps 2 to 5 until the computed feasible chain flow pattern $\underset{\sim}{h}$ converges to fixed values.

5. CONCLUSIONS

Two models related to the determination of chain flows in sections of non-hierarchical telephone networks have been proposed and their accuracy validated by comparison with exact calculations. By incorporating the ideas in these models, the chain flow dimensioning model can be conveniently modified for application to the class of non-hierarchical networks defined in section 1.

Finally, an outline was given for a model, incorporating the methods of sections 1 and 2, suitable for estimating chain flows in larger non-hierarchical networks. A subsequent paper will report on the detail, accuracy and extensions of such a model.

6. REFERENCES

1. Berry, L.T.M., "A mathematical model for optimizing telephone networks", Ph.D thesis, University of Adelaide, Dec. 1971.

2. Jessop, C.W.A., McLeod, N.W., "Thoughts on a new telephone network structure for Australia", Telecom. Journal of Aust., Vol. 33, No. 1, 1983, pp. 67-71.

3. Berry, L.T.M., "Optimal dimensioning of circuit switched digital networks", 10th ITC Montreal, 1983.

4. Harris, R.J., Helm, A.V., "Distribution of overflow traffics", Traffic Engineering News Sheet, No. 12, Telecom, 1976.

BIOGRAPHIES

LES BERRY received the degree of B.Sc. from the University of Adelaide in 1958. For 10 years from 1959-68 he taught in secondary schools and completed a Dip.Sec.Ed. He joined the South Australian Institute of Technology in 1969 as a tutor and was a senior tutor in 1970 and lecturer in 1971. During 1970 he began work for a higher degree under the supervision of Professor R.B. Potts and in 1972 received a Ph.D. for a thesis entitled "A mathematical model for optimizing telephone networks". In 1972 he joined the Department of Applied Mathematics at the University of Adelaide where he is currently a senior lecturer.

His research interests include teletraffic modelling and fundamental traffic theory.

ANDREW COYLE received his Honours Bachelor of Science degree from the University of Adelaide in 1983. He is now working towards a Ph.D degree at the University of Adelaide studying teletraffic theory under the supervision of Les Berry.

Comments on a Property of the Equivalent Random Method

L.T.M. BERRY
University of Adelaide

K.I.M. McKINNON
University of Edinburgh

It was shown in Ref. 1 that, for a certain network, application of the equivalent random method results in an apparent "paradox", namely, the addition of further channels can result in an increase in the total traffic lost due to congestion. This paper shows that this phenomenon cannot occur in reality for the previously considered network and discusses a related class of networks which do display the paradoxical phenomenon.

1. INTRODUCTION

The equivalent random method (Ref. 2) is widely used in practical teletraffic engineering. For example, it is commonly applied to the problem of analysing gradings and is also the basis of many network dimensioning models. When used to determine the performance of a single channel which is being offered two independent traffic streams (see Fig. 1), under the usual assumptions of lost calls cleared and negative exponential holding times, the equivalent random method gives a perhaps surprising conclusion. If the total offered traffic is M $(= M_1 + M_2)$ and the combined variance of the offered traffic is V $(= V_1 + V_2)$ then the total loss m is given by (see Ref. 1)

$$m = M\left[1 - \frac{1}{M + V/M}\right]. \qquad (1)$$

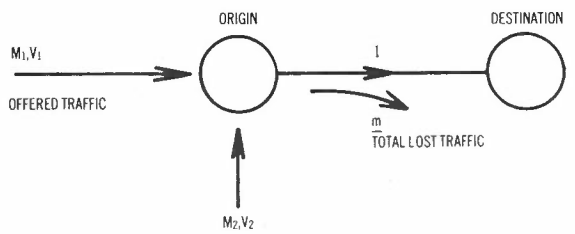


Fig. 1 - A single channel offered two independent streams.

Now if stream 2 is Poisson $(V_2 = M_2)$, the total loss when calls from stream 1 are *prevented from using the single channel* is equal to $M_1 + M_2^2/(1 + M_2)$. Thus for V_1/M_1 sufficiently large, it is possible to satisfy the inequality

$$M\left[1 - \frac{1}{M + V/M}\right] > \frac{M_2^2}{1 + M_2} + M_1. \qquad (2)$$

Therefore, the total loss can (according to (2)) be *increased* by allowing stream 1 to use the common link, even though the calls from both streams have the same holding time distribution! Numerical examples illustrating this phenomenon are given in Ref. 1.

In Section 2 it is shown that the above prediction, based on the equivalent random model, is in fact incorrect. Regardless of the distributions of the independent streams of calls offered to the *single channel*, the best policy (assuming that the aim is to minimize the total loss) is to *never reject an offered call*. Whilst this policy is best for a single link, it is shown in section 3 that when sequences of links (chains) are involved the total traffic lost *can* be decreased by rejecting offered calls. For the class of networks considered, a criterion is given for deciding when to reject offered calls. As a corollary of the previous deductions, it is shown that under some circumstances the addition of further links to a network can lead to an *increase* in the total traffic lost due to congestion.

Section 4 generalizes the results of section 3 and some comments on the practical significance of these results are given in section 5.

2. THE TOTAL LOSS FROM A SINGLE CHANNEL

Consider a stream of calls with an arbitrary inter-arrival time distribution offered to a single channel.

Theorem: Under the following assumptions

(i) the holding time distribution is the same for all calls

(ii) lost calls are cleared from the system

the total lost traffic is minimized by never rejecting a call when the channel is free.

Proof: We first examine the problem of minimizing the expected number of calls lost before time T. Consider a realization of the stochastic arrival process in which call i arrives at time t_i, $-\infty < i \leq N$. We shall now show for this realization that the expected number of lost calls (this being dependent on the common holding time distribution and the rejection policy) is minimized by never rejecting a call when the channel is free.

We make the following definitions:

p_{ik} = the probability that a call ends on or after time t_{k-1} and before time t_k, for $i < k \leq N$;

p_{iN+1} = the probability that a call ends on or after time t_N;

$P_{ik} = \sum_{j=i+1}^{k} p_{ij}$, for $i < k \leq N+1$;

$P_{ii} = 0$;

L_i = the expected number of lost calls from call i to call N inclusive, conditional on the channel being free immediately prior to t_i and assuming that optimal decisions are made on accepting or rejecting calls i to N;

$L_{N+1} = 0$.

For $i \leq k \leq N$, P_{ik} is therefore the probability that a call which starts at time t_i ends before time t_k. Also $P_{iN+1} = 1$. Since all calls have the same holding time distribution and the probability of termination of a call is a non-decreasing function of time

$$P_{ik} \geq P_{i+1\,k}, \quad \text{for } i < k \leq N+1. \tag{3}$$

The L_i satisfy the stochastic dynamic programming equations

$$L_i = \text{Min}\left[1 + L_{i+1},\ \sum_{k=i+1}^{N+1} p_{ik}(L_k + k - i - 1)\right] \quad \text{for } i \leq N, \tag{4}$$

where

$R_i \triangleq 1 + L_{i+1}$

is the expected number of lost calls if call i is rejected, and

$A_i \triangleq \sum_{k=i+1}^{N+1} p_{ik}(L_k + k - i - 1)$

is the expected number of lost calls if call i is accepted.

From (4) it follows that

$$L_i \leq 1 + L_{i+1}, \quad \text{for } i \leq N. \tag{5}$$

We shall now prove by induction starting at $i = N$ that

$R_i \geq A_i$, for $i \leq N$, i.e. that

$$L_i = \sum_{k=i+1}^{N+1} p_{ik}(L_k + k - i - 1). \tag{6}$$

Since $R_N = 1 + L_{N+1} = 1$ and

$$A_N = p_{NN+1}(L_{N+1} + N + 1 - N - 1) = 0,$$

(6) holds for $i = N$.

Now assume (6) holds for $i + 1$.

$$\begin{aligned} A_i &= \sum_{k=i+1}^{N+1} p_{ik}(L_k + k - i - 1) \\ &= \sum_{k=i+1}^{N+1} (P_{ik} - P_{ik-1})(L_k + k - i - 1) \\ &= -P_{ii}(L_{i+1} + i + 1 - i - 1) \\ &\quad + \sum_{k=i+1}^{N} P_{ik}\left[(L_k + k - i - 1) - (L_{k+1} + k + 1 - i - 1)\right] \\ &\quad + P_{iN+1}(L_{N+1} + N + 1 - i - 1) \\ &= \sum_{k=i+1}^{N} P_{ik}(L_k - L_{k+1} - 1) + N - i \end{aligned}$$

(since $P_{ii} = 0$, $P_{iN+1} = 1$ and $L_{N+1} = 0$).

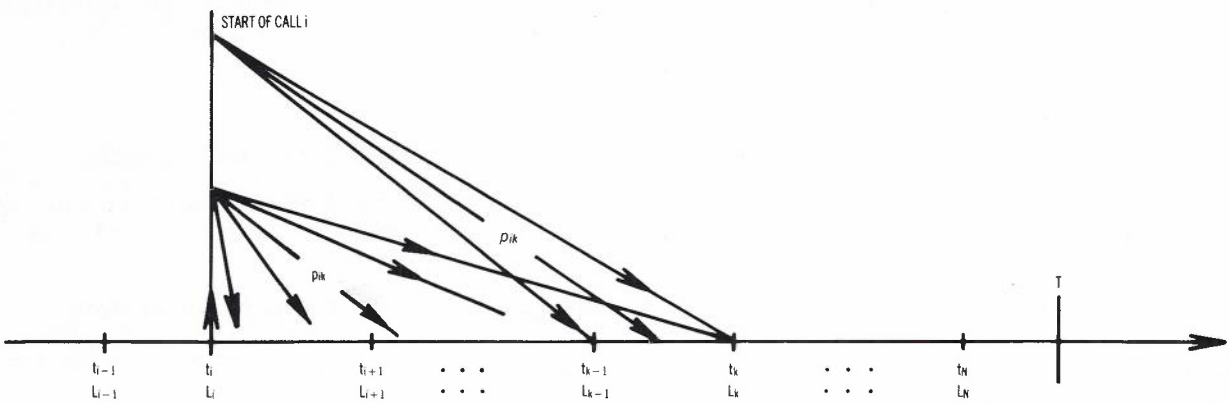


Fig. 2 - A realization of an arrival process.

$$R_i = 1 + L_{i+1} = 1 + \sum_{k=i+2}^{N+1} p_{i+1k}(L_{k+k-i-2})$$

(from inductive assumption)

$$= \sum_{k=i+2}^{N} P_{i+1k}(L_k - L_{k+1} - 1) + N - i$$

(using previous argument)

$$\text{Hence } R_i - A_i = \sum_{k=i+2}^{N} (P_{i+1k} - P_{ik})(L_k - L_{k+1} - 1)$$

$$- P_{ii+1}(L_{i+1} - L_{i+2} - 1).$$

From (3) and (5) and the fact that $P_{ii+1} \geq 0$, it follows that $R_i \geq A_i$, completing the induction.

Since this result holds for an arbitrary arrival time realization, it holds also for arbitrary arrival time distributions (not necessarily stationary). We have now shown that to minimise the expected number of lost calls up to an arbitrary time T the policy of never rejecting a call when a channel is free is optimal. That this policy is optimal over the infinite time horizon follows from the fact that a non-optimal policy over the infinite time horizon must be non-optimal over some finite time. Thus the lost traffic (which for a finite interval is proportional to the number of lost calls) is minimised by a policy of never rejecting an arrival when the channel is free.

Q.E.D.

Consider now two streams of calls, each with arbitrary inter-arrival time distributions but the same holding time distribution, offered to a common link consisting of one channel (Fig. 1). The combined streams viewed as a single stream satisfy the conditions of the theorem and so we can deduce that there is never an advantage in rejecting one of the streams.

3. THE TOTAL LOSS ON A CHAIN WITH SHARED LINKS

In the following section we assume a lost calls cleared regime and negative exponential holding times.

Suppose that a Poisson stream (stream 1), with mean M_1, is offered to a chain of n links with each link consisting of one channel. Each link (i, i+1), i = 1, ... , n, is also offered a Poisson stream with mean M_{i+1} (see Fig. 3).

The loss from stream 1, offered to the entire chain, is denoted by m_1. Losses from the n other streams, each offered only to a single link of the chain, are denoted by $m_2, \ldots, m_{n+1}$.

Firstly, we consider the case n = 2. Let p_{ijk} denote the steady state probability that the chain (1,2), (2,3) is occupied by i calls from stream 1, link (1,2) is occupied by j calls from stream 2 and link (2,3) is occupied by k calls from stream 3. Clearly, there are 5 states [i,j,k] in which the network may be, viz. [0,0,0], [1,0,0], [0,1,0], [0,0,1] and [0,1,1]. If we choose the mean holding time as our unit of time the equilibrium equations for the system are

$$(M_1 + M_2 + M_3)\, P_{000} = P_{100} + P_{010} + P_{001}$$

$$P_{100} = M_1\, P_{000}$$

$$(1 + M_3)\, P_{010} = M_2\, P_{000} + P_{011}$$

$$(1 + M_2)\, P_{001} = M_3\, P_{000} + P_{011}$$

$$2\, P_{011} = M_3\, P_{010} + M_2\, P_{001}. \qquad (7)$$

To remove the redundancy of equations (7) any one equation is replaced by the standardization equation

$$\Sigma\, p_{ijk} = 1\ , \qquad (8)$$

where the summation is over the 5 states. It can be readily seen that the solution to the resultant system is

$$P_{100} = M_1\, P_{000}$$

$$P_{010} = M_2\, P_{000}$$

$$P_{001} = M_3\, P_{000}$$

$$P_{011} = M_2 M_3\, P_{000}$$

$$P_{000} = \frac{1}{1 + M_1 + M_2 + M_3 + M_2 M_3}\ . \qquad (9)$$

It can be shown that the result (9) generalizes in a straight-forward way for n (> 2) links. If $p_{j_1 j_2 \cdots j_{n+1}}$ is the equilibrium probability that j_1 calls from stream 1 occupy the entire chain and j_i calls from stream i (i = 2, ... , n+1) occupy links (1,2), ... , (n,n+1) respectively then

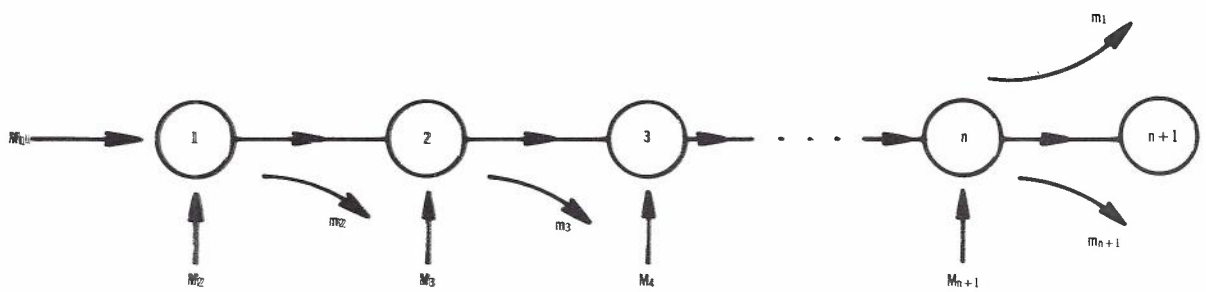


Fig. 3 - A chain offered n+1 Poisson streams.

$$p_{10\ldots0} = M_1 p_{00\ldots0}$$

$$p_{0j_2j_3\ldots j_{n+1}} = M_2^{j_2} M_3^{j_3} \ldots M_{n+1}^{j_{n+1}} p_{00\ldots0}, \tag{10}$$

and $p_{00\ldots0}$ is readily found from the standardization equation.

Suppose now that all calls from stream 1 may be prevented from using the chain. We now investigate the condition under which this rejection results in the total loss *decreasing*. Consider firstly the case n = 2. Clearly, the total loss when calls from stream 1 are rejected is given by

$$M_1 + \frac{(M_2)^2}{1 + M_2} + \frac{(M_3)^2}{1 + M_3}, \tag{11}$$

since the losses for the two independent Poisson streams satisfy the Erlang loss formula. As the total loss from the chain when stream 1 is permitted to use the chain is given by

$$M_1(p_{100} + p_{010} + p_{001} + p_{011}) + M_2(p_{100} + p_{010} + p_{011}) + M_3(p_{100} + p_{001} + p_{011}), \tag{12}$$

application of (9) and some algebraic simplification yields the desired condition

$$M_2 M_3 > 1 \tag{13}$$

or equivalently

$$\frac{M_2}{1 + M_2} + \frac{M_3}{1 + M_3} > 1. \tag{14}$$

A straightforward, but tedious, analysis for the general case reveals that, in order to minimize the total loss, it is always best to prevent stream 1 from using the chain whenever

$$\sum_{j=2}^{n+1} \frac{M_j}{1+M_j} > 1 \tag{15}$$

and best to accept it when the inequality is reversed.

It is interesting to note that the inequality does not involve M_1. Therefore, whenever

$$\sum_{j=2}^{n+1} \frac{M_j}{1+M_j}$$

exceeds 1 the total loss is increased by allowing stream 1 to use the chain, no matter how small M_1 may be. That is, for this situation the optimal policy is to always reject calls from stream 1. We also observe that, as the left-hand side of (15) is the sum of non-negative terms, the longer the chain the more likely it is that the performance (measured by the total traffic carried) deteriorates when stream 1 calls are switched onto the chain. If

$$M = \min_{i>1} M_i,$$

we should reject stream 1 whenever $nM/(1+M) > 1$, that is whenever $n > 1 + 1/M$. Clearly, in the limit as $n \to \infty$ it is always optimal to reject stream 1 calls.

The inequality (15) which was obtained by viewing the arrival/service process as a birth and death process has an intuitive interpretation. The right-hand side is the mean carried traffic over the set of time intervals during which stream 1 occupies the chain whereas the left-hand side is the mean traffic that would have been carried for these same time intervals if stream 1 had been switched off.

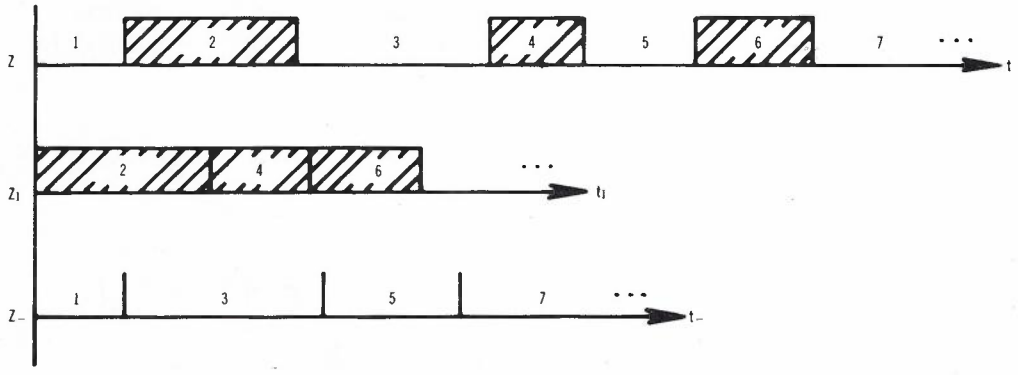


Fig. 4 - The division of a realization into sub-realizations. The shaded areas represent the times when calls from stream 1 occupy the chain.

In the next section we use this observation to provide weaker conditions under which (15) holds.

4. A GENERALIZATION OF A PREVIOUS RESULT

In this section we re-examine the network analysed in section 3 and once again assume a lost calls cleared regime. Initially we consider arbitrary arrival and holding time distributions for all streams. Let H_i be the mean of the holding time distribution for stream i, $1 \leq i \leq n+1$.

We shall refer to the stochastic process corresponding to operating the network with stream 1 permanently rejected as process $S\_$. Let N_i be the average number of calls accepted per unit time from stream i, $2 \leq i \leq n+1$ in process $S\_$. Consider now a realization Z of the operation of the network when some calls from stream 1 are accepted. Fig. 4 depicts such a realization and the shaded areas indicate the times when a call from stream 1 occupies the chain. At all other times the links are available for calls from streams 2 to n+1 and may be free part of that time. Conceptually we group together all those times when stream 1 occupies the chain (we refer to the resulting sub-realization as Z_1) and those times when it does not (we refer to the resulting sub-realization as $Z\_$), and measure time separately within each realization. We refer to time measured within a realization as *internal* time. It is helpful to think of calls from stream 1 switching off and on the channels for the remaining streams.

Let q_1 be the proportion of time in Z when calls from stream 1 occupy the chain and let $q\_ = 1 - q_1$. Let N_i' be the average number of calls accepted from stream i per unit internal time in $Z\_$, for $2 \leq i \leq n+1$. The average number of calls accepted from stream 1 per unit internal time in Z_1 is $1/H_1$. The average total number of calls accepted per unit time in Z is therefore

$$A = \frac{q_1}{H_1} + q\_ \sum_{i=2}^{n+1} N_i' .$$

The average total number of calls accepted per unit time in process $S\_$ is

$$R = \sum_{i=2}^{n+1} N_i = q_1 \sum_{i=2}^{n+1} N_i + q\_ \sum_{i=2}^{n+1} N_i$$

Hence

$$R-A = q_1 \left[\sum_{i=2}^{n+1} N_i - \frac{1}{H_1} \right] + q\_ \sum_{i=2}^{n+1} (N_i - N_i') , \quad (16)$$

and calls from stream 1 should be rejected or accepted depending on whether this is positive or negative. The second term arises from $Z\_$ and measures the effect on streams 2 to n+1 of the interruptions due to stream 1.

If a stochastic process is interrupted at certain times and then is always restarted from the same state as it was in when interrupted, the resulting stochastic process viewed through its internal time is indistinguishable from the uninterrupted process. In particular the expected value of any quantity measured by internal time will be the same.

When the arrival processes of streams 2 to n+1 are Poisson, the state of the system serving these n streams is defined by

(i) a marker for stream i, for $2 \leq i \leq n+1$ which indicates whether or not a call is in progress.

(ii) for those streams with a call in progress, the time since the arrival of the call.

((ii) is unnecessary for exponential holding time distributions.)

A call from stream 1 can only be accepted when the remaining system is in the state with no calls in progress, and when the call from stream 1 ends the system is again in this state. Hence the interruptions do not affect the properties of the process viewed through internal time. It therefore follows that, with probability 1, $N_i = N_i'$ for $2 \leq i \leq n+1$. Also, since the Erlang loss formula applies to general holding time distributions,

$$N_i = \frac{M_i}{(1 + M_i) H_i} .$$

From (16) we can now deduce that calls from stream 1 should all be rejected if

$$\sum_{i=2}^{n+1} \frac{M_i}{(1 + M_i) H_i} > \frac{1}{H_1} , \quad (17)$$

and accepted if the inequality is reversed. When all average holding times are 1 (i.e. $H_i = 1$), (17) reduces to (15).

Note that this argument depends on streams 2 to n+1 having Poisson arrivals. If a stream is not Poisson then the state descriptor for that stream needs to be augmented and there is no longer any guarantee that the state will match before and after the interruption from stream 1. For example, if the arrivals are described by a general renewal process, the state description needs (ii) whether or not a call is in progress.

When $n = 1$ and $H_1 = H_2$ and stream 2 has Poisson arrivals, (17) shows that stream 1 should never be rejected. This is a special case of the result in section 2.

In the case when $N_i = N_i'$ for $2 \leq i \leq n+1$, (16) leads to the following condition:

Calls from stream 1 should be rejected if

$$\sum_{i=2}^{n+1} N_i > \frac{1}{H_1} \text{ and accepted if } \sum_{i=2}^{n+1} N_i < \frac{1}{H_1} . \quad (18)$$

This result has the following intuitive interpretation:

> If the number of calls accepted from stream 1 during those time periods when stream 1 occupies the chain is *strictly less than* the number which would be accepted from the other streams within these same time periods with stream 1 rejected, then stream 1 should be rejected. If the inequality is reversed, then stream 1 should be accepted. (18')

We conclude this section with an example which shows that without Poisson arrivals for streams 2 to n+1, conditions (18) and (18') need not hold.

Consider a 2 link chain whose 3 streams have the following properties: Stream 1 has an arbitrary arrival distribution and a fixed holding time of 1 unit. Hence $H_1 = 1$. Stream 2 has a fixed inter-arrival time of 2 units and a fixed holding time of 1 unit. Hence $N_2 = 1/2$ and $H_2 = 1$. Stream 3 has arbitrary arrival and holding time distributions.

Every call from stream 1 which is accepted prevents exactly one call from stream 2 being accepted and also 'switches off' the channel for stream 3. Section 2 shows that rejecting calls from stream 3 cannot improve the performance of this channel. Hence whatever are the properties of stream 3, it is always optimal to reject stream 1. However

$$\sum_{i=2}^{3} N_i - \frac{1}{H_1} = \frac{1}{2} + N_3 - 1 = N_3 - \frac{1}{2}$$

and this is negative if $N_3 < 1/2$, showing that conditions (18) and (18') do not hold for this system.

5. CONCLUSIONS

Consider the network shown in Fig. 5. Stream 2 is Poisson traffic offered to destination 2, whereas stream 1 is rough overflow traffic.

By choosing n = 1 or 0 we accept or reject calls offered to D_1 via the common link O-T. According to the equivalent random model, there are values of M_1 and V_1 for which the total traffic lost is *increased* by allowing stream 1 to use the single common channel. We have proved in section 2 that this prediction is incorrect and therefore due solely to the inaccuracies of the equivalent random model. The addition of a further circuit to the network (i.e. changing n from 0 to 1) results in a decrease of the total lost traffic.

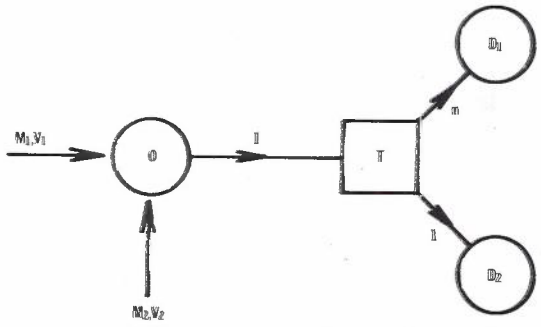


Fig. 5 - Single origin two destination network.

A further class of networks has been examined to illustrate that it is indeed possible for the addition of circuits to a network to cause an *increase* in the total lost traffic. (Consider the addition of an extra link after node n+1 in Fig. 3; this link controls the acceptance or rejection of traffic from stream 1).

Finally, we note an important practical consequence of inequality (17). As the effect of offering a traffic stream to a "long" chain in a telephone network is likely to result in an increase in the total lost traffic, caution is needed when contemplating allowing long origin-destination routes in non-hierarchical networks.

6. REFERENCES

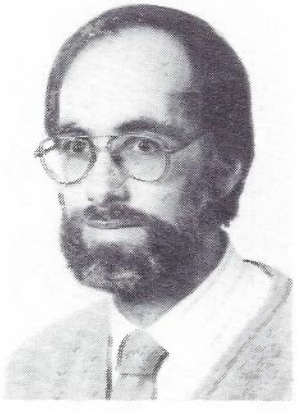
1. Berry, L.T.M., "Telecommunications Paradoxes", A.T.R., Vol. 17, No. 2, 1983, pp. 5-9.

2. Wilkinson, R.I., "Theories for Toll Traffic Engineering in the U.S.A.", B.S.T.J., Vol. 35, No. 2, March 1956, pp. 421-514.

BIOGRAPHIES

LES BERRY received the degree of B.Sc. from the University of Adelaide in 1958. For 10 years from 1959-68 he taught in secondary schools and completed a Dip.Sec.Ed. He joined the South Australian Institute of Technology in 1969 as a tutor and was a senior tutor in 1970 and lecturer in 1971. During 1970 he began work for a higher degree under the supervision of Professor R.B. Potts and in 1972 received a Ph.D. for a thesis entitled "A mathematical model for optimizing telephone networks". In 1972 he joined the Department of Applied Mathematics at the University of Adelaide where he is currently a senior lecturer.

His research interests include teletraffic modelling and fundamental traffic theory.



KENNETH IAIN MUIR McKINNON received a 1st class Honours degree in Mathematics and Physics at the University of Glasgow in 1969 and a Ph.D. at the University of Cambridge in 1972 on the topic "Stochastic Optimisation".

Since 1972 he has been a lecturer in Mathematics at the University of Edinburgh. As a member of the Edinburgh Management Science Partnership he assisted in the production of the integer programming package IP200 for ICL-Dataskil and has also participated in other projects, mainly related to linear programming.

His main interests are in Large Scale Mathematical Programming Problems and techniques for their solution.

Maximum Likelihood Estimation of the Frequencies of Multiple Tones in Noise

K.S. ENGLISH

Telecom Australia Research Laboratories

The maximum likelihood approach provides an accurate method for estimating the frequencies and amplitudes of closely spaced complex-valued tones in additive noise. However, there is a major implementation obstacle in that the frequency estimation problem leads to minimisation of a non-convex function of several variables. This function exhibits many non-global local minima, and the minimisation problem cannot be reliably solved by conventional gradient-based algorithms. A heuristic minimisation algorithm is described which is computationally efficient and almost always converges to the global minimum. Some alternative methods are briefly reviewed in order to place the proposed method in its proper context. The algorithm has been successfully applied to the analysis of data from a multipath radio propagation experiment.

1. INTRODUCTION

The problem to be studied is that of estimating the parameter vector $\underset{\sim}{\alpha} = \{a_1, a_2, \ldots a_K, \omega_1, \omega_2, \ldots \omega_K\}$ from a single observation of the sequence of complex samples $Z = \{z_0, z_1, \ldots z_{N-1}\}$, where the unknown parameters enter through the expression:

$$z_m = \sum_{k=1}^{K} a_k \, e^{j\omega_k (m - \frac{N-1}{2})} + n_m, \quad m=0, 1, \ldots N-1 \qquad (1)$$

The sequence $\{n_m\}$ is from a complex-valued white noise process. The in-phase and quadrature components of $\{n_m\}$ are independent zero-mean white noise processes. The parameters a_k are complex, ω_k are real variables in the range $-\pi < \omega_k \leq \pi$ and N and K are assumed to be known integers. No harmonic relationship between the ω_k is assumed. The important companion problem of estimating the parameter K is not treated in the present paper. The exponential function of an imaginary argument in (1) will be referred to as the "cisoidal" function.

The close relationship between this parameter estimation problem and the wider class of spectral estimation problems is immediately evident when we consider the Fourier transform of $\{z_m\}$. If computed over the full index range of m (not restricted to the window 0 to N-1), the spectrum consists simply of an impulse of weight a_k at each of the frequencies ω_k, $k = 1,2, \ldots K$, and a term due to the noise component. This suggests that the unknowns can be estimated from the positions and heights (or areas) of the peaks in a spectral estimate of the observed sequence.

Implementation of this basic approach is very effective in situations where the components are well separated in frequency (Ref. 1). With availability of the fast Fourier transform (FFT) and other efficient algorithms (Ref. 2) for computing the discrete Fourier transform (DFT), the resulting procedure is very favourable from a computational viewpoint. The number of prominent peaks in the spectral estimate also provides a good indication of the number of components present.

If the separation in normalised frequencies of the components is in the order of $2\pi/N$ (dependent on the relative phase) then the spectral peaks overlap sufficiently to disturb the peak estimation procedure, even in the absence of noise. For still closer frequency separations the two peaks merge into a single peak hiding any evidence of multiple components. The situation is exacerbated by noise which tends to distort the spectral peaks making parameter estimation still more error-prone. It should be noted that data windowing (multiplying the observed sequence by a smoothly tapered window function (Ref. 3) before computing the DFT) only exaggerates the problem. Windowing is designed to suppress the side-lobes in the spectral estimate (secondary peaks not centred on the main-lobe) at the expense of widening the main lobe. Hence, in the context of spectral resolution, the rectangular window is optimum.

Short-comings of the simple DFT method of processing data containing closely spaced components have led to development of a number of higher resolution techniques. Broadly, these can be classified into two types:

. spectral estimation: An improved spectral estimate is derived from which the unknown parameters are again extracted from the positions and amplitudes of the spectral peaks.

. parameter estimation: Bypassing the frequency domain representation of the observed sequence, parameter estimation techniques are directly applied to estimate the unknown parameters.

Recent developments in high resolution spectral estimation (Ref. 4) have tended to be dominated by the problem of estimating the power spectrum of a stationary random process. Spin-offs from this work have contributed powerful techniques applicable to the tone estimation problem considered here. However, it is important to appreciate that the underlying processes generating the observed sequence is different in each case. There is ample evidence in the literature on spectral analysis (for example, see Ref. 5) to suggest that grossly incorrect inferences can be drawn by applying a good spectral analysis procedure in the wrong situation.

Noting that the frequency separation resolution limit is inversely related to the length of the observed sequence, we see that the resolution enhancement problem is essentially equivalent to that of extrapolating the sequence beyond the observation limits. This requires some additional constraint on the data expressed in a computationally convenient form. The two signal models which have received attention are in modelling the observed process as an auto-regressive process (the output of an all-pole filter driven by white noise) or by restricting the observed process bandwidth to a sub-interval of the normalised range $(-\pi,\pi)$. Neither model is generally applicable to the tone estimation problem. In the tone estimation problem the white noise enters as an additive sequence corrupting the "signal" component, rather than as a driving process of an all-pole filter. Thus the auto-regressive process model is not representative unless an observation noise component is added to the output of the all-pole filter. This modification invalidates most of the analytic approaches to auto-regressive modelling. In the bandlimited case, even if it is possible to restrict the component frequencies to given ranges, it is seldom feasible to similarly restrict the noise spectrum. Unless corrections are made for these effects, or the results interpreted with appropriate caution, it is very easy to be led to incorrect estimates of the unknown parameters.

The parameter estimation approach avoids the abovementioned difficulties by incorporating the underlying model in a very direct manner. Here, the major hurdle is instead one of computational efficiency. Even for $K = 2$ the parameter estimation formulation can lead to minimisation of a non-convex function of 6 real variables for which no analytic solution is possible. Hence, the goal is to select an estimation algorithm which is both physically representative and able to be implemented economically.

The algorithm described in this paper is an implementation of a parameter estimation technique based on the maximum likelihood (ML) criterion. Assuming a Gaussian distribution for the noise samples $\{n_m\}$ the ML formulation leads to a least squares approximation problem. For a fixed set of component frequencies minimisation of the mean square error with respect to the amplitude variables can be done analytically. However, minimisation with respect to the frequency variables leads to a difficult K-variable non-convex function minimisation problem. Gradient-based minimisation algorithms are unreliable unless started close to the global minimum. Globally convergent algorithms of the "space-filling" type (Ref. 13) are capable of finding the global minimum but are very expensive on computer time. The paper proposes an alternative minimisation algorithm which converged to the global minimum on all test problems but requires much less computation than a full space-filling algorithm. With this modification the ML approach becomes a practical alternative to algorithms based on auto-regressive spectral analysis. The algorithm has been extensively tested on simulated data and is now being applied to the analysis of data generated by a multipath radio propagation experiment.

The modified ML algorithm described here has previously been reported in Refs. 6 and 7. The ML algorithm itself (without the sub-optimum minimisation algorithm) has been studied over a number of years by several authors (Refs. 8-12). In particular, Rife and Boorstyn (Ref. 8) first presented the multi-component Cramer-Rao bounds which indicate limits on the achievable spectral resolution enhancement set by the background noise level.

2. ML ESTIMATION ALGORITHM

2.1 ML Formulation

Assuming a complex Gaussian distribution (Ref. 19) of zero-mean for the noise samples, the ML formulation of the parameter estimation problem can be stated as follows.* Given a sample vector $\underline{z} = [z_0, z_1, \ldots z_{N-1}]^T$, compute values for the elements of the parameter vector $\underline{\alpha} = [a_1, a_2, \ldots a_K, \omega_1, \omega_2, \ldots \omega_K]$ so as to maximise $p(\underline{z}|\underline{\alpha})$ given by:

$$p(\underline{z}|\underline{\alpha}) = \frac{1}{\pi^N \sigma_m^{2N}} \exp\left\{-\frac{1}{\sigma_m^2}\sum_{m=0}^{N-1} |z_m - s_m|^2\right\}, \tag{2}$$

where

$$s_m = \sum_{k=1}^{K} a_k e^{j\omega_k\left(m-\frac{N-1}{2}\right)} \tag{3}$$

and

$$\sigma_m^2 = E[|n_m|^2].$$

Taking the customary step of minimising instead the log-likelihood function we obtain:

* Throughout the paper the circumflex is used to denote a parameter estimate. The superscript T denotes transpose and superscript H denotes conjugate transpose.

$$L(\underset{\sim}{z}|\underset{\sim}{\alpha}) = \ln[p(\underset{\sim}{z}|\underset{\sim}{\alpha})]$$

$$= - N \ln(\pi\sigma_n^2) - \frac{N}{\sigma_n^2} J_K(\underset{\sim}{\alpha}) \quad (4)$$

where

$$J_K(\underset{\sim}{\alpha}) = \frac{1}{N} \sum_{m=0}^{N-1} |z_m - s_m|^2 . \quad (5)$$

If the noise variance σ_n^2 is known then $L(\underset{\sim}{z}|\underset{\sim}{\alpha})$ is maximised by minimising $J_K(\underset{\sim}{\alpha})$ with respect to the unknown parameters in $\underset{\sim}{\alpha}$. If σ_n^2 is unknown, then fixing $\underset{\sim}{\alpha}$, σ_n^2 can be estimated by maximising $L(\underset{\sim}{z}|\underset{\sim}{\alpha})$ with respect to σ_n^2 resulting in the estimate

$$\hat{\sigma}_n^2(\underset{\sim}{\alpha}) = J_K(\underset{\sim}{\alpha}). \quad (6)$$

Substituting for σ_n^2, $L(\underset{\sim}{z}|\underset{\sim}{\alpha})$ is again maximised by minimising $J_K(\underset{\sim}{\alpha})$. This establishes the well-known result that the ML estimate of the parameters of a signal in additive Gaussian white-noise is obtained by performing a least squares fit of the signal model to the observed sequence.

2.2 Estimation of Amplitudes

Expression (5) can be written in a convenient matrix form as:

$$J_K(\underset{\sim}{\alpha}) = \frac{1}{N} \underset{\sim}{e}^H \underset{\sim}{e} \quad (7)$$

where:

$$\underset{\sim}{e} = \underset{\sim}{z} - E\underset{\sim}{a} , \quad (8)$$

$$\underset{\sim}{z} = [z_0, z_1, \ldots z_{N-1}]^T , \quad (9)$$

$$\underset{\sim}{a} = [a_1, a_2, \ldots a_K]^T \quad (10)$$

and E is an (N,K) matrix with generic element

$$[E]_{m,k} = e^{j\omega_k (m - \frac{N-1}{2})} ,$$

$$m=0, 1, \ldots N-1, k=1, 2, \ldots K. \quad (11)$$

Consider firstly the minimisation of $J_K(\underset{\sim}{\alpha})$ with respect to $\underset{\sim}{a}$. Setting the gradient $\nabla_{\underset{\sim}{a}} J$ to zero obtains a necessary condition:

$$(E^H E)\, \underset{\sim}{a} = E^H \underset{\sim}{z} . \quad (12)$$

For notational convenience we define the (K,K) matrix $D(\underset{\sim}{\omega})$ and K-vector $\underset{\sim}{A}(\underset{\sim}{\omega})$ by:

$$D(\underset{\sim}{\omega}) = \frac{1}{N} E^H E , \quad (13)$$

$$\underset{\sim}{A}(\underset{\sim}{\omega}) = \frac{1}{N} E^H \underset{\sim}{z} , \quad (14)$$

where:*

$$\underset{\sim}{\omega} = [\omega^1, \omega^2, \ldots \omega^K]^T . \quad (15)$$

Equation (12) can then be written as

$$D(\underset{\sim}{\omega})\, \underset{\sim}{a} = \underset{\sim}{A}(\underset{\sim}{\omega}) . \quad (16)$$

This equation can be solved uniquely for $\underset{\sim}{a}$ if the matrix $D(\underset{\sim}{\omega})$ is non-singular. From (13), for all K-vectors $\underset{\sim}{x}$:

$$\underset{\sim}{x}^H D(\underset{\sim}{\omega})\, \underset{\sim}{x} = \frac{1}{N} \|E\underset{\sim}{x}\|^2 \geq 0$$

Thus $D(\underset{\sim}{\omega})$ is non-negative definite. It is strictly positive definite, and hence non-singular, if the matrix equation $E\,\underset{\sim}{x} = \underset{\sim}{0}$ has no solution for $\underset{\sim}{x} \neq \underset{\sim}{0}$. This is true if $K \leq N$ and no two elements of $\underset{\sim}{\omega}$ are identical modulo 2π. Therefore, subject to these conditions, (15) can be solved to give the unique optimum amplitudes for a given set of frequency estimates:

$$\hat{\underset{\sim}{a}}(\underset{\sim}{\omega}) = D(\underset{\sim}{\omega})^{-1}\, \underset{\sim}{A}(\underset{\sim}{\omega}) . \quad (17)$$

The elements of vector $\underset{\sim}{A}(\underset{\sim}{\omega})$ are given by:

$$[\underset{\sim}{A}(\underset{\sim}{\omega})]_k = A(\omega^k) , \; k=1, 2, \ldots K , \quad (18)$$

where the scalar function $A(\omega)$ is:

$$A(\omega) = \frac{1}{N} \sum_{m=0}^{N-1} z_m\, e^{-j\omega(m - \frac{N-1}{2})} . \quad (19)$$

The elements of matrix $D(\underset{\sim}{\omega})$ are:

$$[D(\underset{\sim}{\omega})]_{i,k} = D(\omega^i - \omega^k), \; i, k=1, 2, \ldots K , \quad (20)$$

where $D(\omega)$ is the (scalar) Dirichlet kernel function:

$$D(\omega) = \frac{1}{N} \sum_{m=0}^{N-1} e^{j\omega(m - \frac{N-1}{2})} = \frac{\sin \frac{N\omega}{2}}{N \sin \omega/2} \quad (21)$$

---

* The superscript notation ω^i denotes the i-th element of vector $\underset{\sim}{\omega}$ as distinct from the candidate values ω_i of any element of $\underset{\sim}{\omega}$, as used later.

Substituting the solution for $\hat{\underset{\sim}{a}}(\underset{\sim}{\omega})$ back into (7) gives an expression for the least square error $J_K(\underset{\sim}{\omega})$ for an arbitrary frequency estimate $\underset{\sim}{\omega}$ using the optimum amplitude estimate:

$$J_K(\underset{\sim}{\omega}) = J_0 - \underset{\sim}{A}(\underset{\sim}{\omega})^H D(\underset{\sim}{\omega})^{-1} \underset{\sim}{A}(\underset{\sim}{\omega}) \qquad (22)$$

where:

$$J_0 = \frac{1}{N} \underset{\sim}{z}^H \underset{\sim}{z} = \frac{1}{N} \sum_{m=0}^{N-1} |z_m|^2 . \qquad (23)$$

2.3 Estimation of Frequencies

The remaining task is to compute the frequency estimate $\hat{\underset{\sim}{\omega}}$ at which $J_K(\underset{\sim}{\omega})$ is minimized. This minimization problem has not been amenable to analytic solution. Furthermore, the function $J_K(\underset{\sim}{\omega})$ is found to possess many non-global local minima and hence gradient-based procedures yield non-global solutions unless started from a point very close to the global minimum. These observations are illustrated in Figs. 1 and 2.

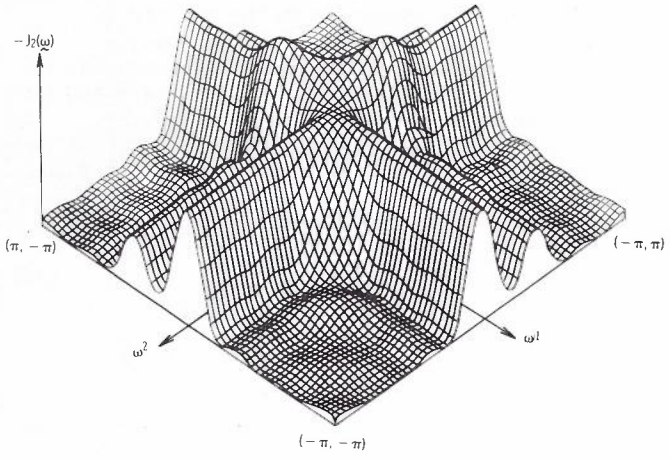


Fig. 1 - Isometric representation of typical function $J_2(\underset{\sim}{\omega})$.

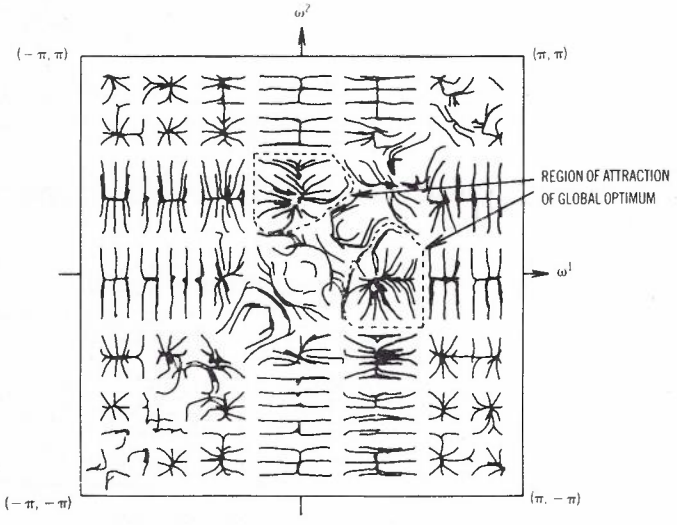


Fig. 2 - Trajectories of simple gradient search of $J_2(\underset{\sim}{\omega})$.

Fig. 1 shows an isometric representation of $-J_2(\underset{\sim}{\omega})$ for an observed sequence consisting of components at $\omega=0$ and $\omega=\pi/3$, with N=8 and no additive noise. Fig. 2 shows the trajectories of a simple gradient-based search of this function starting from points distributed throughout the domain of $J_2(\underset{\sim}{\omega})$. Notice that the searches converge to more than 60 separate local minima. The region in which starting points result in convergence to a particular local minimum is commonly called the region of attraction of that local minimum (Ref. 13). We see from Fig. 2 that the region of attraction surrounding the global minimum is only slightly larger than around other minima. Therefore the common practice of performing gradient-based searches from a number of randomly selected starting points, then selecting the best result, is unlikely to give any major improvement in this case.

The only general class of non-convex function minimisation algorithm with guaranteed convergence to the global minimum* is the class of "space-filling" search algorithms (Ref. 13). In essence, these involve evaluating the objective function ($J_K(\underset{\sim}{\omega})$ in this case) at a set of sample points distributed over the full function domain then selecting the point at which the function is minimised. By utilising prior knowledge of the function smoothness (typically bounds on its derivatives) it is possible to construct a finite set of sample points which confine the global minimum to an arbitrarily small region of the function domain. For functions of one or two variables the technique is quite feasible, although often expensive in computer time. However, the computational cost of a full space-filling search is usually prohibitive for functions of more than two variables.

To overcome these difficulties a search strategy is reported here which performs in practical terms just as well as the space-filling searches. Although global convergence cannot be rigorously established, as for the space-filling algorithm, it has been found to converge to the global minimum in all test cases. The special structure of the objective function is utilised to achieve a significant reduction in the number of function evaluations required to locate the global function minimum.

The approach has been to replace the full global optimality criterion C1,

$$\text{C1:} \quad J_K(\hat{\underset{\sim}{\omega}}) \leq J_K(\underset{\sim}{\omega}) \quad \text{for all } \underset{\sim}{\omega} , \qquad (24)$$

by a more easily tested sub-optimality criterion C2,

* One often sees reference in the literature of function minimisation to "global-convergence". In contemporary usage this does not imply convergence to the global minimum, but merely convergence to a local minimum in the presence of other local minima. This is in itself an important property of an algorithm which cannot be taken for granted.

C2: $J_K(\hat{\underset{\sim}{\omega}}) \leq J_K[\hat{\underset{\sim}{\omega}}_k(\omega)]$, $-\pi < \omega \leq \pi$

$1 \leq k \leq K.$ (25)

The K-vector $\hat{\underset{\sim}{\omega}}_k(\omega)$ is obtained from $\hat{\underset{\sim}{\omega}}$ by replacing its kth element by the value of the argument ω. Stated differently, a solution vector $\hat{\underset{\sim}{\omega}}$ will be said to satisfy sub-optimality criterion C2 if it cannot be improved by fixing all but one of the component frequencies in and then varying the remaining frequency variable. The underlying premise of the search strategy to be described is that a point $\hat{\underset{\sim}{\omega}}$ which satisfies C2 will also satisfy C1. There is no theoretical proof of this equivalence and in fact it would be surprising if counter-examples could not be constructed. However, there is fairly compelling evidence to suggest that in all situations of interest the equivalence can be assumed. Some of the arguments are briefly discussed below:

(i) Any point in the function domain satisfying C1 will also satisfy C2. This follows trivially from the observation that C2 involves a search over a restricted part of the domain searched in C1. Hence we have not excluded the global minimum by testing for C2 instead of C1.

(ii) A point $\hat{\underset{\sim}{\omega}}$ satisfying C2 is necessarily a local minimum of the objective function (subject to some regularity conditions met by $J_K(\underset{\sim}{\omega})$). This follows from the point $\hat{\underset{\sim}{\omega}}$ being a local minimum in each of the K orthogonal search directions. Hence, we are limiting the set of possible solutions to being a sub-set of the function local minima.

(iii) The equivalence of C1 and C2 follows intuitively from representations such as Fig. 1. This shows an objective function resulting from two components at $\omega=0$ and $\pi/3$. Notice the presence of ridges in $-J_2(\underset{\sim}{\omega})$) (equivalent to troughs in $J_2(\underset{\sim}{\omega})$) at $\omega=0$ and $\pi/3$ for each function argument, intersecting at the points $(0,0)$, $(0, \pi/3)$, $(\pi/3, 0)$ and $(\pi/3, \pi/3)$. Criterion C2 is violated at all points other than these ridge intersection points because a better solution can always be found at a point where the region tested for C2 crosses a ridge. Of the four ridge intersection points only the two equivalent global maxima $(0, \pi/3)$ and $(\pi/3, 0)$ satisfy C2 demonstrating the equivalence of C1 and C2 in this particular case. This same ridge structure is also expected in situations of more than two components, very close components and signals corrupted by noise, although the ridge positions may deviate slightly from the true component frequencies.

(iv) The equivalence of C1 and C2 was verified by extensive computer simulation. During testing on several hundred sequences generated from a multipath propagation experiment (Ref. 14) not a single input sequence (within the class considered) was shown to violate the equivalence*.

Verification of the optimality of $\hat{\underset{\sim}{\omega}}$ by C2 is much simpler than by C1, since K one-dimensional searches are required instead of a search over a K-dimensional space.

2.4 Very Close Components

The limiting case as two elements of $\underset{\sim}{\omega}$ approach arbitrarily closely is of special interest. As two (or more) elements of $\underset{\sim}{\omega}$ become equal the matrix $D(\underset{\sim}{\omega})$ becomes singular and equation (22) cannot be applied directly. But by considering also the effect of the limiting process on $\underset{\sim}{A}(\underset{\sim}{\omega})$, a numerical result can still be obtained. For example, working through the calculation of $J_2(\underset{\sim}{\omega})$ as both components approach the value ω, one obtains (Ref. 6):

$$J_2(\underset{\sim}{\omega}) = J_0 - |A(\omega)|^2 - \frac{12}{N^2-1} |A'(\omega)|^2 , \quad (26)$$

where

$$A'(\omega) = \frac{dA(\omega)}{d\omega}$$

$$= \frac{-j}{N} \sum_{m=0}^{N-1} (m - \frac{N-1}{2}) z_m e^{-j\omega(m-\frac{N-1}{2})} . \quad (27)$$

The first two terms in (26) give the mean square error in fitting a single cisoidal component of frequency ω. The third term gives the additional benefit of fitting two components of infinitesimally different frequency. On computing the reconstructed sequence $\{s_m\}$ using (3) we find an expression of the form:

$$s_m = [a+b(m-\frac{N-1}{2})] e^{j\omega(m-\frac{N-1}{2})}$$

$$= \lim_{\delta \to 0} 1/2 \left[(a + \frac{jb}{\delta}) e^{j(\omega+\delta)(m-\frac{N-1}{2})} + (a - \frac{jb}{\delta}) e^{j(\omega-\delta)(m-\frac{N-1}{2})} \right] , \quad (28)$$

where:

$$a = A(\omega) , \quad b = \frac{6jA'(\omega)}{N^2-1} .$$

That is, the two components of infinitesimal separation have an equivalent interpretation as a single cisoidal component weighted by a linear function of m. Notice also that as $\delta \to 0$ the individual component amplitudes become infinitely large but of opposite polarity and hence almost cancel in the reconstructed sequence.

* Testing of global optimality was not done exhaustively. Instead a number of local minima in the vicinity of the anticipated solution were compared with the solution satisfying C2. In all cases only a single solution satisfying C2 was detected. This solution coincided with the best local minimum found.

The issue raised here is extremely important in properly interpreting the results of the estimation procedure. In practice one is not dealing with an observed sequence which precisely fits the assumed model. Instead, the model of equation (1) is likely to be the result of an idealised analysis or simply a computationally convenient simplification of the known true form. Situations can arise in which the data exhibits a small, almost imperceptible, slope across the observed sequence. The ML procedure unknowingly models this perturbation as the effect of two very close components and fits the component frequencies and amplitudes accordingly.

From the results above, this leads to estimation of components of large amplitude but opposite polarity, such that their net contribution is much smaller. If we adopt a strategy of ranking the importance of components in terms of their estimated amplitude, then the two close components will quite invalidly be assigned an undue importance. This highlights an important interpretation guideline that in the case of component separations less than about $2\pi/N$ the estimated amplitude is not in itself an accurate measure of the true component amplitude.

One means of avoiding this difficulty is to simply discard solution vectors $\underset{\sim}{\omega}$ whose component separation is less than some arbitrary limit. An alternative approach (Ref. 6) is to add a penalty term $\lambda \underset{\sim}{a}^H \underset{\sim}{a}$ to $J_K(\underset{\sim}{\omega})$ in (7) so as to penalise solutions with large amplitudes. A term λ is then added to each diagonal element of $D(\omega)$ in (22) before matrix inversion. It forces the procedure to reject any solution with unrepresentatively close component frequencies. Selection of λ is largely a matter of trial and error.

The preferable solution is to extend the signal model (3) to include all polynomial weighted components of the form described. The ML estimation procedure can be extended in a straight-forward manner to estimate all weighting coefficients analytically, and to estimate the unknown frequencies by a function search procedure. Alternatively, the estimation can proceed on the basis of the signal model (3) but special attention be given during interpretation of the results to the equivalence expressed in (28). That is, wherever two (or more) components of large amplitude appear close together, consideration should be given to treating the whole group as a single cisoidal component with slowly varying amplitude. This approach is numerically stable as long as the minimum frequency separation is set in relationship with the computer word length to avoid problems of matrix singularity.

As a final comment on this aspect of the problem, it is equally valid in many physical situations to consider a model with cisoidal components of slowly varying frequency as well as amplitude. The computational problem is less tractible here because the additional parameters enter in a non-linear manner, thus preventing their removal by simple analytic techniques.

2.5 Cramer Rao Bounds

The Cramer Rao (CR) parameter estimation bound (Ref. 17) provides a lower bound on the variance of any unbiased estimator, and under rather general conditions provides information on the asymptotic performance of the ML estimator. If any estimator attains the CR bound then the result is numerically equal to the ML estimate of the parameter.

The CR bounds for the multiple tone estimation problem were first derived and numerically evaluated by Rife and Boorstyn (Ref. 8); the analytical derivation is carried a little further in Ref. 6. The frequency estimation CR bound for two components and N=8 is plotted against the component separation in Fig. 3. The vertical axis is normalised to the CR bound for one isolated component (Ref. 1) and the horizontal axis is normalised to the Rayleigh resolution limit of $2\pi/N$. Normalised in this manner the CR bound is almost independent of N, particularly for components closer than the Rayleigh resolution limit. There is a strong dependence of the CR bound on the relative phase $\Delta\phi$ of the two components (where the phase angle of each component is measured at a point mid-way across the observation window). For "in-phase" components the CR bound is considerably larger than for "in-quadrature" components. Another interesting feature (discussed in Ref. 8) is the dependence of the CR bound on the relative frequency and phase but not the relative amplitude of the two components. The component amplitude only enters through the signal/noise ratio of each component separately.

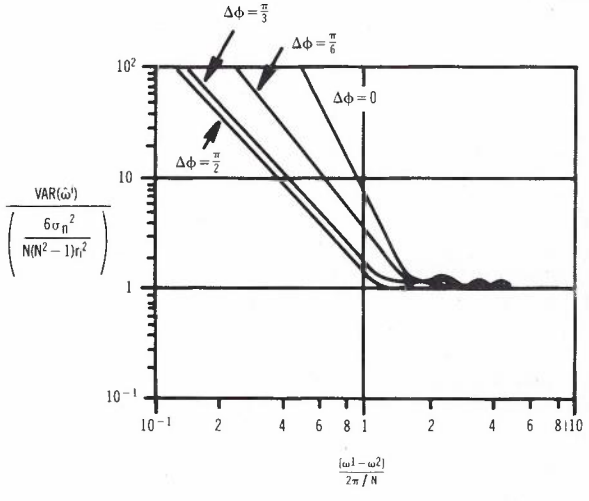


Fig. 3 - Cramer-Rao bound on frequency estimation (N=8, K=2, $r_i=|a_i|$).

The most important feature of the CR bound is the sharp knee evident between component separations of $2\pi/N$ and $4\pi/N$. For larger separations there is only a small penalty in estimating each component frequency due to the anticipated presence of the other component. For smaller separations the CR bound rises sharply above the single component bound becoming asymptotically large as the separation goes to zero. This quantifies the empirical result that resolution enhancement beyond the Rayleigh resolution limit is only effective in situations of high signal/noise ratio.

The ML estimator only approaches the CR bound for situations of high signal/noise ratio. This is because the CR bound only effectively models the disturbance caused by noise to the shape of the main lobe in the spectral estimate which results in a small perturbation to the position of the peak (Ref. 6). For values of σ_n comparable to the component amplitude there is also a significant probability that the global spectral peak may occur anywhere across the spectrum independent of the true component frequency (Ref. 1). This causes a rapid degradation of mean square error with increasing σ_n. Furthermore, for large values of σ_n there is less confidence that the solution minimising C2 will also minimise C1. Despite these limitations the CR bounds provide a very useful first approximation to the actual achievable performance.

3. IMPLEMENTATION ASPECTS

3.1 Basic Algorithm Structure

Summarising the results of the last section, our goal is to compute a solution vector for which the criterion C2 defined in (25) is satisfied. The optimum amplitude estimates can then be obtained from (16).

The function minimisation is implemented as a search procedure; a number of sample points are selected, $J_k(\underset{\sim}{\omega})$ is evaluated at each sample point, then the results are used to select the next sample points. The overall computation efficiency is related both to the number of function evaluations required and to the computational load of each function evaluation.

To reduce the computation associated with evaluating $J_k(\underset{\sim}{\omega})$, the allowable values of any element of $\underset{\sim}{\omega}$ are restricted to M points uniformly distributed over the expected solution range:

$$\omega_i = \omega_0 + \Delta\omega\, i \ , \ 0 \leq i \leq M-1 \ ,$$

$$-\pi < \omega_0 < \omega_{M-1} \leq \pi \ . \qquad (29)$$

This permits the full set of possible values for the elements of $D(\underset{\sim}{\omega})$ and $A(\underset{\sim}{\omega})$ to be pre-computed and stored to save repetitive calculation later on.

With this discrete sampling of the parameter space of $\underset{\sim}{\omega}$, any frequency estimate is uniquely specified by a set of K integers in the range 0 to (M-1) arranged in strictly ascending order. The strict ordering of the elements of $\underset{\sim}{\omega}$ has several desirable effects:

(i) By avoiding the possibility of two elements of $\underset{\sim}{\omega}$ being identical, modulo 2π, we ensure that the corresponding matrix $D(\underset{\sim}{\omega})$ is non-singular.

(ii) If all elements of $\underset{\sim}{\omega}$ were allowed to range over the complete interval $[\omega_0, \omega_{M-1}]$, then there would be at least K! equivalent global minima obtained by permutation of the elements of $\underset{\sim}{\omega}$. By imposing the ordering constraint the solution vector becomes unique (unless there is in fact more than one genuine global minimum) and avoids possible convergence problems caused by multiple global minima.

(iii) The ordering constraint on $\underset{\sim}{\omega}$ causes the matrix $D(\underset{\sim}{\omega})$ to be dominated by its elements close to the leading diagonal. This ensures that $D(\underset{\sim}{\omega})$ is well-conditioned for inversion using diagonal pivots.

The K component frequency estimate is obtained by first computing the 1, 2, ... K-1 component estimates. In fact, this approach was dictated more by the requirements of the application (relating to order determination) than to requirements of the minimisation algorithm. In cases where only the K component estimate is of interest the lower order estimates could be bypassed by choosing any convenient starting point without invalidating the algorithm.

The search algorithm is structured as a sequence of searches along one dimensional line segments and local-minima seeking pattern searches (Ref. 15), confined throughout to the lattice of candidate solution points defined by (29). The essential features of the algorithm are summarised in Fig. 4. At each stage the solution vector $\hat{\underset{\sim}{\omega}}$ is set equal to the point minimising $J_K(\underset{\sim}{\omega})$ over all previously evaluated points. The line searches are cycled through the K variables of $J_K(\underset{\sim}{\omega})$ and terminated when K successive searches fail to improve the solution. The starting point for the K-component search is obtained by setting K-1 frequency variables equal to the (K-1) component estimate then searching over the remaining variable for the minimum function value.

The purpose of interleaving pattern searches with the one-dimensional searches is to speed convergence close to the optimum. Without this feature the successive line searches in orthogonal directions approach the same result in small steps, but at the cost of a large number of redundant function evaluations. Details of the pattern search algorithm are given in Ref. 16.

The parameter $\Delta\omega$ in (29) must be chosen sufficiently small to resolve the fine structure of the function $J_K(\underset{\sim}{\omega})$; this requires a value not larger than about $\pi/3N$. Assuming a residual error in the frequency estimate uniformly distributed over the interval $(-\Delta\omega/2, \Delta\omega/2)$ this results in a lower bound on the frequency estimate mean square error of $(\Delta\omega)^2/12$. For typical component separations this bound greatly exceeds the CR bound described in section 2.5. If higher accuracy results are required then a solution close to the CR bound can be obtained by starting a gradient descent search from the final result of the initial search algorithm. In most cases a frequency resolution of about $\pi/3N$ will be acceptable and the gradient-based search can be omitted.

```
procedure global_search;
begin
   i := 0;
   for K = 1 to KMAX do
      repeat
      begin
         ω' := ω;
         line_search (i,ω);
         i := (i+1) mod K;
         if ω' ≠ ω then pattern_search (ω)
      end
      until ω is unchanged by K successive
                                 line searches
end
procedure line_search (i, ω);
begin
   for j = 0 to M-1 do
   begin
      ω' := ω_{i+1} (ω_j);
      if J_K(ω') < J_K(ω) then ω := ω'
   end
end
procedure pattern_search (ω);
begin
   while a suitable direction vector δ can be
                                       found do
   begin
      choose some δ for which J_K(ω+δ) < J_K(ω);
      repeat
         ω := ω+δ
      until J_K(ω+δ) ≥ J_K(ω)
   end
end
```

Fig. 4 - Pseudo code for function minimisation algorithm

3.2 Modifications of the Basic Algorithm

Introducing the sub-optimality criterion C2 provides a big saving in computation over C1. However, the most time-consuming part of the search algorithm is still the one-dimensional searches. By paying close attention to the relationship between K-1 and K component estimates, it was found possible to omit some of these searches without affecting the solution vector. The penalty is some loss of confidence in the algorithm having reached a solution close to the optimum, but for routine data processing applications this may be worthwhile for the saving in computation. The approach is illustrated for the case of K=2.

Suppose that two components ω_a and ω_b are present. Then the search of $J_1(\omega)$ (equivalent to finding the peaks of $|A(\omega)|^2$) will yield one of two results:

(i) For two well-separated components: $|A(\omega)|^2$ exhibits peaks near both ω_a and ω_b and also at other low-level sidelobes. The minimum of $J_2(\underset{\sim}{\omega})$ can be reached by starting a local search from the vector obtained by combining solutions for the two dominant peaks of $|A(\omega)|^2$.

(ii) For two close components: $|A(\omega)|^2$ shows a single dominant peak somewhere between ω_a and ω_b, say at ω_c. Initialising a local search from the point $[\omega_c-\delta, \omega_c+\delta]^T$, for some small δ, will successfully locate the global minimum of $J_2(\underset{\sim}{\omega})$.

Thus in almost all cases the global minimum of $J_2(\underset{\sim}{\omega})$ can be located by following both these strategies then selecting the best result. During 150 representative trials this approach yielded the global optimum in all but one case.

3.3 Evaluation of A(ω)

Prior to the search algorithm, function A(ω) is evaluated and stored for each candidate value of ω:

$$A_i = A(\omega_i)$$

$$= \frac{1}{N} \sum_{m=0}^{N-1} z_m \, e^{-j(\omega_0 + i\,\Delta\omega)(m - \frac{N-1}{2})} ,$$

$$i=0, 1, \ldots M-1 . \quad (30)$$

The function A(ω) is just the unwindowed Fourier transform of the finite sequence $\{z_m\}$; the sequence $\{A_i\}$ can be efficiently computed by the chirp - Z transform (CZT) algorithm (Ref. 2).

3.4 Evaluation of $J_K(\underset{\sim}{\omega})$

An efficient method of evaluating $J_K(\underset{\sim}{\omega})$ contributes significantly to the overall computational efficiency. Restating (22), $J_K(\underset{\sim}{\omega})$ is given by the expression:

$$J_K(\underset{\sim}{\omega}) = J_0 - \underset{\sim}{A}(\underset{\sim}{\omega})^H \, D(\underset{\sim}{\omega})^{-1} \, \underset{\sim}{A}(\underset{\sim}{\omega}) . \quad (31)$$

One efficient means of evaluation is to perform a Cholesky decomposition of D into P^TP then compute $J_K(\underset{\sim}{\omega})$ according to:

$$J_K(\underset{\sim}{\omega}) = J_0 - \| P^{-T} A(\underset{\sim}{\omega}) \|^2 \quad (32)$$

Another still more efficient algorithm is described below. The matrix $D(\underset{\sim}{\omega})^{-1}$ is represented in the form:

$$D(\underset{\sim}{\omega})^{-1} = \sum_{i=1}^{K} f_i \, \underset{\sim}{g}_i \, \underset{\sim}{g}_i^T \quad (33)$$

for suitable scalars f_i and K-vectors $\underset{\sim}{g}_i$. Substituting back into (31), $J_K(\underset{\sim}{\omega})$ can then be computed from:

$$J_K = J_0 - \sum_{i=1}^{K} f_i |p_i|^2 , \qquad (34)$$

where

$$p_i = \underset{\sim}{A}(\underset{\sim}{\omega})^T \underset{\sim}{g}_i . \qquad (35)$$

For K=2 the eigenvector expansion of $D(\underset{\sim}{\omega})$ is used to obtain:

$$f_1 = \frac{1}{2(1+D_{1,2})} , \quad \underset{\sim}{g}_1 = [1,1]^T , \qquad (36)$$

$$f_2 = \frac{1}{2(1-D_{1,2})} , \quad \underset{\sim}{g}_2 = [1,-1]^T ,$$

with $D_{1,2} = D(\omega^1 - \omega^2)$, function $D(\omega)$ being given by (21).

For $K \geq 3$ the eigenvector expansion is too complex to be a useful computational approach, but the following algorithm can be used instead.

Define $D(\underset{\sim}{\omega})$ by $D(\underset{\sim}{\omega}) = D_K$, where the matrices D_i are defined recursively by

$$D_i = \begin{bmatrix} D_{i-1} & \underset{\sim}{d}_i \\ \underset{\sim}{d}_i^T & 1 \end{bmatrix} , \quad \underset{\sim}{d}_i = \begin{bmatrix} D(\omega^1 - \omega^i) \\ \vdots \\ D(\omega^{i-1} - \omega^i) \end{bmatrix} , \quad D_o = 1. \qquad (37)$$

Recursively applying the following matrix inversion formula:

$$D_i^{-1} = \begin{bmatrix} D_{i-1}^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0}^T & 0 \end{bmatrix} + f_i \begin{bmatrix} \underset{\sim}{y}_i \\ -1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{y}_i^T , -1 \end{bmatrix} \qquad (38)$$

where:

$$f_i = (1 - \underset{\sim}{d}_i^T D_{i-1}^{-1} \underset{\sim}{d}_i)^{-1} \qquad (39)$$

$$\underset{\sim}{y}_i = D_{i-1}^{-1} \underset{\sim}{d}_i , \qquad (40)$$

we obtain the representation for D_K^{-1} in (33) with $\underset{\sim}{g}_i^T = [\underset{\sim}{y}_i^T, -1, \underset{\sim}{0}^T]$. Vector $\underset{\sim}{y}_i$ has dimension (i-1) and $\underset{\sim}{0}$ adjusts the dimension of $\underset{\sim}{g}_i$ to K. After some further re-arrangement the final algorithm for computing $J_K(\underset{\sim}{\omega})$ becomes, for $D_{i,k} = D(\omega^i - \omega^k)$ and $h_{k,i} = \underset{\sim}{d}_k^T \underset{\sim}{g}_i$, that given in Fig. 5.

```
procedure evaluate_J_K(ω);
begin
   f_1 := 1/[2(1+D_{1,2})];
   f_2 := 1/[2(1-D_{1,2})];
   p_1 := A(ω^1) + A(ω^2);
   p_2 := A(ω^1) - A(ω^2);
   J_2 := J_0 - f_1|p_1|^2 - f_2|p_2|^2;
   for k = 3 to K do
   begin
      h_{k,1} := D_{1,k} + D_{2,k};
      h_{k,2} := D_{1,k} - D_{2,k};
      if k ≥ 4 then [for i = 3 to k-1 do
                     h_{k,i} := Σ_{j=1}^{i-1} f_j h_{k,j} h_{i,j}];
      f_k := (1 - Σ_{i=1}^{k-1} h^2_{k,i} f_i)^{-1};
      p_k := Σ_{i=1}^{k-1} f_i h_{k,i} p_i - A(ω^k);
      J_k := J_{k-1} - f_k|p_k|^2
   end
end
```

Fig. 5 - Calculation algorithm for $J_K(\underset{\sim}{\omega})$

TABLE 1 - Comparison of Function Evaluation Algorithms

K=	Cholesky Method			Proposed Method	
	ADD/SUB	MULT/DIV	SQRT	ADD/SUB	MULT/DIV
2	7	15	2	10	10
3	17	32	3	20	21
4	34	58	4	34	39
5	60	95	5	53	66

A count of real arithmetic operations for this algorithm is compared with the Cholesky decomposition approach in Table 1. The number of additions/subtractions are similar but there is a worthwhile saving in multiplications/divisions with the proposed method. Also, a square root operation is not required.

4. DISCUSSION AND CONCLUSION

4.1 Selecting the Best Algorithm

With the wide range of spectral estimation and parameter estimation algorithms available, selecting the algorithm appropriate to a particular situation can be quite difficult. To assist in this choice an attempt will be made to place the proposed algorithm in context with other available techniques.

Unless high resolution is specifically required, the appropriate algorithm is almost certainly the method of locating the peaks of the windowed Fourier transform. There are a number of very good window functions to choose from (Ref. 3), all giving similar performance on well separated components. The method is quite efficient in terms of computation and performs robustly in the presence of noise.

High resolution techniques should only be invoked when component separations are known to be possibly less than about $4\pi/N$ (normalised to a total bandwidth of 2π). Also, the background noise must be sufficiently low (see section 2.5) not to mask the fine structure of the spectrum. If both these conditions hold then the choice is principally between methods based on the ML criterion (amounting to least squares fitting of the assumed signal model) and methods based on modelling the observed sequence as an auto-regressive process.

The auto-regressive techniques have been investigated extensively by many authors, and applied both to the power spectral density estimation problem and to the tone estimation problem. Refs. 4 and 18 review much of this work. In summary, these methods perform very well in situations of low noise and provide spectral resolution well beyond the limit of the windowed DFT method. However, the performance degrades rapidly with increasing noise levels and is generally inferior to the optimum ML algorithm.

Tufts and Kumaresan (Ref. 12) have recently described an auto-regressive technique for which performance close to that of ML estimation is claimed. The preliminary results presented in Ref. 12 look interesting but verification of their claim is made difficult by the fact that the algorithm is not derived directly from the ML estimator. Therefore, the relative performance can only be investigated in detail by an extensive programme of simulation. No such detailed comparative results are available. Also, their technique obtains the frequency estimates by first locating the roots of a polynomial of order $L=N-K/2$ (in our notation), rejecting $L-K$ "extraneous roots", then estimating the unknown frequencies from the angular position of the remaining roots. In practical situations there appears to be no simple and reliable procedure for determining which roots are "extraneous".

The ML estimator implementation described in this paper is appropriate in situations where high resolution is required and the anticipated noise level precludes the marginally less computationally demanding auto-regressive techniques. A detailed comparison of the relative amounts of computation has not been made, but the promise of performance as close as is achievable to the CR bound was sufficient justification to adopt the algorithm described in the application which motivated this work.

4.2 On Interpreting the Results

Before applying high resolution spectral analysis techniques one should question very carefully whether the model of very close spectral components is truly representative of the physical process being studied. As discussed in section 2.4 there is a good case for recognising the virtual equivalence of several models of the process: multiple close cisoidal components, a single cisoidal component of slowly varying amplitude and phase or frequency and combinations of these basic models (e.g. several close components with slowly varying amplitude). Once one model is optimally fitted to the observed sequence it can be transformed to one of the other representations without significantly changing the reconstructed sequence and least squares fit (or other criteria). Unless the observations are almost noise-free we cannot expect to be able to discriminate between these models of the process just on the basis of the observed sequence. Additional information on the physical process is required. The details are clearly very application-specific.

4.3 Conclusion

A method of estimating the frequencies and amplitudes of closely spaced complex-valued tones in additive noise has been described. The proposed method performs in most cases identically to the optimum ML estimator, and should give performance close to the CR bound over a wide range of signal/noise ratios.

In addition to the estimation algorithm itself an introduction is given to the problem of correctly interpreting the results in situations of multiple components with very close frequencies. In some cases this may be representative of the physical situation whereas in others it may be more meaningful to treat this collection of components as representing a single component of slowly varying amplitude. Recognising this equivalence it is sometimes possible to avoid the pitfalls of high resolution spectral analysis by accordingly modifying the assumed model of the process.

5. ACKNOWLEDGEMENT

The author is indebted to Mr John Murphy, project leader for the multipath propagation experiment, for his guidance and encouragement throughout the course of this work.

6. REFERENCES

1. Rife, D.C. and Boorstyn, R.R., "Single Tone Parameter Estimation from Discrete-Time Observations", IEEE Trans. Inform. Theory, IT-20, No. 5, Sept. 1974, pp.591-598.

2. Digital Signal Processing Committee, IEEE ASSP Society, (Eds), "Programs for Digital Signal Processing", IEEE Press, New York, 1979.

3. Harris, F.J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", Proc. IEEE, Vol. 66, No. 1, Jan. 1978, pp.51-83.

4. Childers, D.G. (ed), "Modern Spectral Analysis", IEEE Press, New York, 1978.

5. Nitzberg, R., "Spectral Estimation : An Impossibility?", Proc. IEEE, Vol. 67, No. 3, March 1974, pp.437-438.

6. English, K.S., "High Resolution Estimation of Line Spectra", M.Eng.Sci Thesis, Monash University, Melbourne 1979.

7. English, K.S., "Super Resolution Techniques for Processing Multipath Propagation Data", IEEE Workshop - Digital Signal Processing, L'Aquila, Italy, Sept. 1980.

8. Rife, R.C. and Boorstyn, R.R., "Multiple Tone Parameter Estimation From Discrete-Time Observations", Bell System Tech. J., Vol. 55, No. 9, Nov. 1976, pp.1389-1410.

9. Palmer, L.C., "Coarse Frequency Estimation Using the Discrete Fourier Transform", IEEE Trans. Inform. Theory, IT-20, No. 1, Jan. 1974, pp.104-109.

10. Tufts, D.W. and Kumaresan, R., "Improved Spectral Resolution", Proc. IEEE, Vol. 68, No. 3, March 1980, pp.419-420.

11. Tufts, D.W. and Kumaresan, R., "Improved Spectral Resolution II", Proc. ICASSP, April 1980, pp.592-597.

12. Tufts, D.W. and Kumaresan, R., "Estimation of Frequencies of Multiple Sinusoids : Making Linear Prediction Perform Like Maximum Likelihood", Proc. IEEE, Vol. 70, No. 9, Sept. 1982, pp.975-989.

13. Dixon, L.C. and Szego, G.P., "Towards Global Optimisation", North-Holland, 1975.

14. Burton, J. and Murphy, J., "Tropospheric Multipath Propagation Analysis Using a Microwave Holographic Array", Telecom Australia Research Laboratories Report 7038, September 1975.

15. Hooke, R. and Jeeves, T.A., "Direct Search Solution of Numerical and Statistical Problems", Jour. Assoc. Comput. Mach., Vol. 8, No. 2, April 1961, pp.212-229.

16. English, K.S., "A Least Squares Approach to Multipath Propagation Analysis", Telecom Australia Research Laboratories Report 7567.

17. Zacks, S., "The Theory of Statistical Inference", Wiley, 1971.

18. Kay, S.M. and Marple, Jr., S.L., "Spectrum Analysis - A Modern Perspective", Proc. IEEE, Vol. 69, No. 11, Nov. 1981, pp.1380-1419.

19. Miller, K.S., "Complex Stochastic Processes", Addison-Wesley, Massachusetts, 1974.

BIOGRAPHY

KEVIN ENGLISH received the B.Eng.(Hons) degree in 1974 and M.Eng.Sc. degree in 1982 from Monash University. He has been with Telecom Australia Research Laboratories since 1974, involved initially in studies of adaptive echo cancellation, mathematical modelling and analysis of experimental data. In 1979 he was awarded a GEC Overseas Fellowship and undertook two years of industrial training with GEC Telecommunications Ltd in Coventry, U.K. and GEC Hirst Research Centre, Wembley, U.K. Since returning to Australia in 1981 he has been studying switching and signalling aspects of the future integrated services digital network (ISDN).

Administration of Traffic Forecasting for Sparse Rural Telephone Networks

J.P. FARR
E.H.J. WALDRON

Telecom Australia, Perth, Western Australia

This paper addresses the problems associated with the forecasting of traffics for a large number of small exchanges in a sparse rural network. A technique for directing forecasting effort is described which has been shown to enhance the efficiency, accuracy and timeliness of producing traffic and trunk circuit forecasts for terminal exchanges.

1. INTRODUCTION

"Are not two sparrows sold for a penny? And not one of them will fall to the ground without your Father's will." (Matthew Ch. 10, v. 29). Planning of rural areas is a challenging task because of its special characteristics: low subscriber density; subscribers irregularly distributed and separated by distances ranging from some hundred metres to more than 20 km; traffic dispersion mainly oriented towards the commercial centre serving the area; considerable variability in traffic characteristics for different exchanges; and continuing evolution of facilities offered to rural telephone subscribers (Refs. 1,2).

A report (Ref. 3) by the U.S. Committee on Government Operations has commented that "providing telephone services to rural areas is a costly enterprise as without the benefit of a high population density and with fewer business subscribers, rural telephone enterprises suffer from high operating costs."

The sparse nature of some regions of the Australian network may be demonstrated by the case of the State of Western Australia in which, excluding the metropolitan concentration of customers, 100,000 country (rural) customers are distributed over an area of 2.5 million square kilometres. The network catering for these customers comprises numerous small and very small exchanges separated by large distances. The situation is aggravated by the traffic limitations of the exchange equipment and transmission equipment capacity considerations. These small terminal exchanges switch local traffic internally and, in most cases, all outgoing traffic is trunked to the nearest Minor Switching Centre, usually via a bothway route. Although the revenue per customer in the areas may be quite high relative to more densely populated areas, the total revenue does not justify the staff numbers which would be necessary for highly detailed planning work to be undertaken.

Before computer aids may be utilised to carry out network design, exchange and route dimensioning, acceptable forecasts of traffic are needed. This paper outlines a computer-assisted scheme to produce traffic forecasts simultaneously for very large numbers of rural exchanges and dimension their trunk routes. The purpose is to illustrate how traffic forecasting for sparse networks can be administered to support co-ordinated network planning and also to focus the attention of the planning engineers on the more critical areas. Techniques are presented which realise both these goals and require staff levels in keeping with the economic considerations applicable to such networks. Minimising the requirement for expertise in this manner is attractive to all telecommunications administrations (Ref. 4) and may be of particular value to administrations serving developing countries (Ref. 5).

2. SOME PROBLEMS FACING TRAFFIC FORECASTERS

As the number, location and calling habits of telephone users change with time in response to demographic, social and economic factors, the traffics offered to the various links of a telecommunications network are never static. It follows that if the network is to continue to meet prescribed standards of congestion loss, the quantities of traffic-carrying devices must change fairly frequently. Final routes, for example, should be augmented no later than when the traffic has reached a nominal value corresponding to the Grade-of-Service objective previously fixed for each route (Ref. 6).

Typical planning and design tasks which require traffic forecasts are:

- Selection of the best type of switching equipment to suit the traffic
- Dimensioning of switching equipment
- Dimensioning of junction and trunk circuit groups.

Engineers responsible for rural network planning thus require timely updates of the forecasts of traffic for many exchanges in the network. This leads to difficulties in resourcing the forecasting activities since, for small exchanges, mechanised methods for projecting traffic usage rates are not widely used and remain the subject of further study. Some of the major problems arise from the following factors:

(a) Volume considerations. In Australia, there are approximately 3000 small country automatic exchanges (capacity up to 300 services) and each year a significant proportion require new design work, re-design or design checking. In the State of Western Australia, there are 566 exchanges with less than 300 customers, this being 83% of all exchanges in the State. Providing frequent forecasts of data items such as number of customers, traffic calling and terminating rates, traffic volumes and possibly traffic dispersion for a network with such a large number of relatively small exchanges requires considerable use of skills which may be scarce. Also, the small exchanges generate a disproportionate volume of planning workload compared to the revenue produced by these exchanges.

(b) Variations between exchanges. An essential parameter used for dimensioning telephone equipment is the traffic volume to be handled, which is calculated from estimates of the number of subscribers and the "mean traffic per subscriber" (also called "traffic usage rate per service"). This elementary, but vital, concept conceals a complex reality. Within a given subscriber group, some subscribers use the telephone a great deal and others seldom. Moreover, "subscriber-telephone traffic" histograms which show the uneven distribution of traffic also have the property of changing their shape as the period of observation increases (Ref. 7).

Fig. 1, which is built up from traffic data for all small rural exchanges in the State of South Australia, provides evidence of the extreme variability of average trunk traffic rates per service for different exchanges (Ref. 9). Traffic per service data for small exchanges in other States of Australia have been found to exhibit a similar degree of variance.

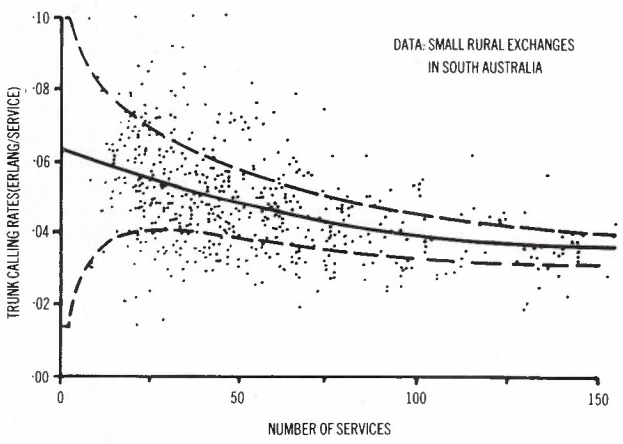


Fig. 1 - Scatter Diagram of Calling Rates

Fig. 1 also indicates a degree of inverse correlation between the number of services and the trunk/junction calling rates.

Each data point on the plot is obtained from a measured trunk traffic level to which an Engset correction has been applied. The central curve is a best fit quadratic, ($y = a - bx + cx^2$). This can be considered as an average trunk usage rate function. The upper and lower curves show the 90% confidence limits for measurements which actually lie on the average usage rate function. Measurement sampling error can account for much of the scatter but by no means all of it. As can be seen, the distance between the average curve and the curves for its confidence limits of measurement bears an inverse relationship to the number of services.

Fig. 2 is intended to illustrate the great variability in exchange growth rates. For this purpose, total subscribers growth was calculated for the 5 year period June 1977 to June 1982 for a sample of 100 rural telephone exchanges in Western Australia, and the frequency distribution of these growth factors plotted.

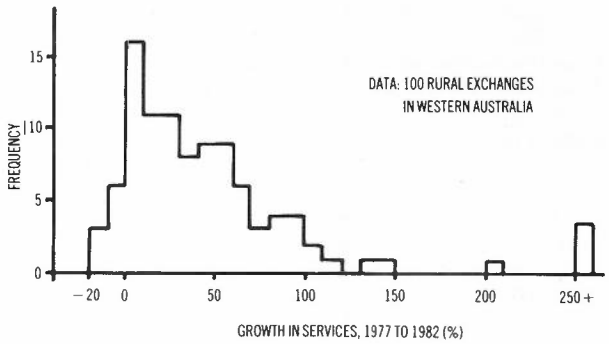


Fig. 2 - Frequency Distribution of Exchange Growths

The variability depicted in Figs. 1 and 2 emphasises the requirement for skilled forecasting personnel in the Traffic Engineering and Market Planning Branches if traffic and circuit forecasts are to be credible.

(c) Changes in Environment. Inherent problems in traffic forecasting for rural networks can be aggravated by other influences, e.g.:

- suppressed demand which causes the true level of potential traffic to be masked,
- when significant tariff changes are introduced,
- in situations where the network is still evolving e.g. conversion from manual to automatic, introduction of subscriber trunk dialling,
- the influence of economic growth on the balance between local and trunk traffic, and
- rising expectations of business and residential customers requiring improved Grades of Service.

3. SOLUTIONS CONSIDERED

Despite the absence of reliable automated procedures for projecting calling and terminating rates etc., the forecasting administrator must consider ways of improving effectiveness and efficiency. One approach to the task of estimating future traffic usage rates per service is the characterisation of customers. Heterogeneous overall markets are disaggregated into more homogeneous sub-markets: e.g. (i) PABX (ii) WATS and INWATS (iii) Coin Telephones (iv) Business sector subdivided into economic branches and (v) Residential sector sub-divided into occupational groups of the heads of households. Estimates of average traffic per subscriber are assigned to each sub group. Work in this area is likely to be fruitful for larger exchanges (Ref. 8). However, for small exchanges, the number of customers in the groups would be too small to warrant this approach at present.

The data shown in Fig. 1 suggest that, given a forecast of the number of services, the trunk traffic usage rates for small rural exchanges at a future date can be estimated from the curve of best fit. This method has been applied successfully using data applicable to specific geographic regions in South Australia (Ref. 9) but some difficulties were experienced in attempts to apply the technique to Western Australian rural exchanges, apparently due to a greater variance of traffic usage rates for small and very small exchanges.

3.1 Strategy Adopted

Our principal objective was to produce a system which would enable a mass-production approach to the task of producing up-to-date forecasts of traffic parameters and circuit requirements for large numbers of rural telephone exchanges, taking due account of both historical and new data for each exchange. Success has been achieved by approaching the forecasting task in the following way:

(a) Variations in customers forecasts which could imply significant changes in the traffic usage per service characteristics of an exchange are distinguished from variations which do not imply a change in character. Highlighted exchanges can be examined in detail, thus avoiding the need to attempt characterisation of traffic per service.

(b) The latest available traffic measurement data for traffic usage per service are compared with the previous most recent forecast for this point of time. Significant variations are detected.

(c) The system highlights those exchanges where significant variations occur. The reports focus the detailed traffic forecasting effort onto those specific parts of the large, sparse network where it is most needed.

(d) Where no appreciable variation was observed, the new customer forecasts and the previous traffic usage rate forecasts are amalgamated and re-published for the use of planning staff. This is done as a volume process on a regular basis, e.g. 700 exchanges can be updated annually in one bulk process. Thus new forecasts (for 12 future points in time) of Calling Rate, Terminating Rate, Trunk/Junction Usage Rate, Trunk/Junction Traffic and Circuits Required become available simultaneously for all exchanges which pass the test criteria.

Publishing all the forecasts after bulk testing leads to the concept of accreditation. It is usual to have planning staff regularly requesting new traffic forecasts. We suggest that an efficient way of satisfying these requirements is by making use of a formal accreditation procedure. We have found that staff carry out their detailed planning responsibilities with more efficiency and confidence when they have a recent set of figures which have been endorsed as those they

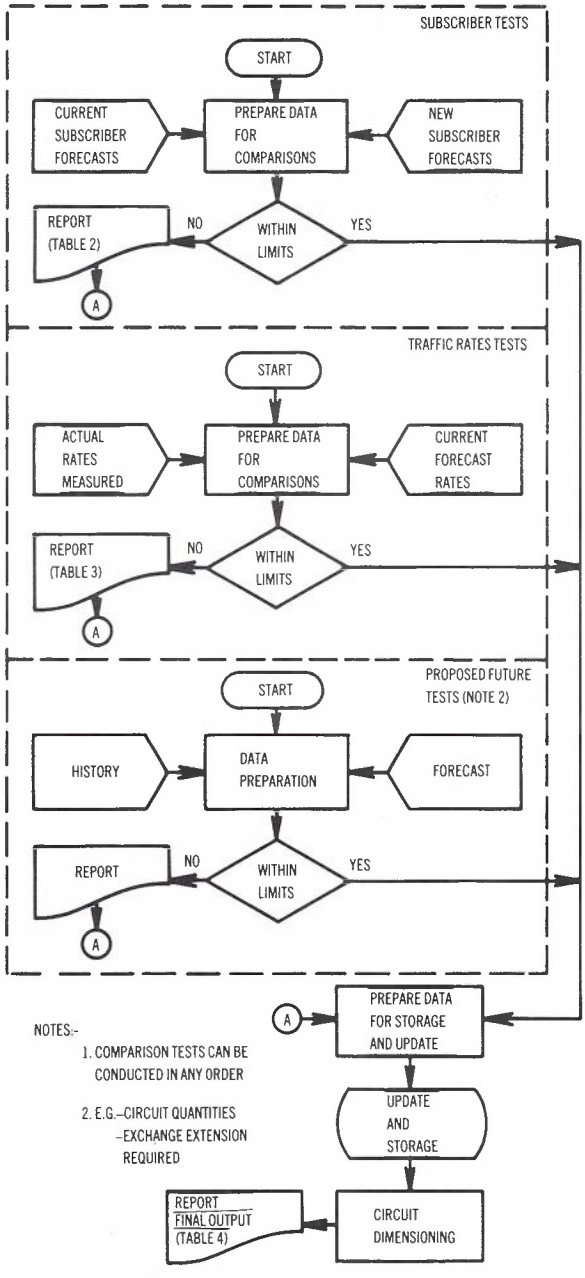


Fig. 3 - Procedural Flow Chart for FERRET System

should use. Economies of forecasting effort are also obtained since, for most exchanges, the accreditation does not require the preparation of new forecasts of traffic usage per service.

Our acronym for the system is FERRET, derived from the title "Finding Exchanges Requiring Revision of Earlier Trends".

The system has been used to provide three productions of volume accreditation of detailed exchange traffic forecasts for the Western Australian country network. This was carried out by using a set of file manipulation computer programs. These programs are simple in concept. No attempt is made at automated projection of usage rate per service or automatic design of exchanges. An outline of the program functions may be seen in Fig. 3.

4. EXPERIENCE TO DATE

Our experience has shown that although a significant number of changes in subscriber forecasts occur when comparing annual issues of forecasts, the number of exchanges for which the variations can be expected to have a significant impact on traffic rates is small.

As examples, we will consider the Albany and Margaret River Minor Switching Areas in Western Australia. These are rural/tourist areas over 300 km from the capital city containing 22 and 12 terminal exchanges respectively. An analysis of the changes that have occurred in the subscribers forecast since 1978 for these exchanges demonstrates that although about half of the forecast figures varied each year by up to 10% compared to the forecast made the year before, the number of exchanges for which variations from one year's forecast to the next exceeded 30% was very small. The history of revision of exchange traffic usage rate forecasts by Traffic Engineering staff is set out in Table 1.

TABLE 1 - Traffic Rate Forecasting History

Year Forecast Made	Albany		Margaret River	
	Exchanges	Reforecasts	Exchanges	Reforecasts
1978	25	21	12	12
1979	25	20	12	6
1981	24	4	12	0
1982	22	6	12	4

This shows the predictable outcome that with the introduction in 1981 of a discriminatory system i.e. FERRET, the number of exchanges actually reforecasted decreased. Prior to FERRET, forecasters tended to completely review the forecast for each exchange examined although the new customer forecast data and new traffic usage rate data may only have been marginally different from earlier data. This human factor is curbed by means of the FERRET system.

4.1 Overall Performance of the System

In the 1981 annual production, the testing limits chosen resulted in approximately 12% of the exchanges being flagged as requiring special attention. A similar proportion (15%) of exchanges were flagged in the 1982 annual production. When the exchanges flagged by the system were examined, about half actually required re-forecasting of traffic usage rates per service. The performance of the system in highlighting exchanges requiring revision is set out below.

	1981	1982
Exchanges tested	706	704
No. failed 'Subscriber' test criteria	59	25
No. failed 'Traffic Rates' test criteria	20	82
No. failed both criteria	6	3
No. requiring re-forecast of traffic rates	44	59

With further experience in operating the system it is expected that the limits that the program uses can be adjusted so that less exchanges will be flagged where a review is unnecessary. However, this change will not be made quickly since exchanges not requiring review form a safety margin. Naturally, the setting of the limits is one of the most crucial aspects of management of this system.

The overhead costs associated with this system are small and are more than recovered by the reduction in forecaster clerical effort and the improved design of the network. The labour required using the system and completing the subsequent detailed forecast reviews in 1982 was equivalent to an average of about 1 manhour per exchange. The computer costs associated with running the system were about $A1 per exchange per annum for 700 country exchanges. These costs include the establishment and maintenance of the base data files, which in fact would be common to any computerised system. The machine used is a CYBER 73.

The FERRET system is supported by adequate documentation. This consists of detailed descriptive material, instructions for preparing data, and operating procedures for running the various computer program modules.

6. DESCRIPTION OF THE PROCEDURES

The procedural flow chart (see Fig. 3) indicates the overall process.

Stage 1. Subscribers Forecasts Tests

In the first stage, analysis is performed of the forecasts of number of services for each exchange. The forecasts, produced by the Market Planning Branch, are available for each of Telecom Australia's 'standard years' - i.e. annually for the next five years, and at five yearly intervals for up to 30 years ahead.

The analysis involves these three processes for each exchange:-

(a) the forecast annual growth for each of five years is calculated and these become the short term growth factors;

(b) the forecast growth between each of the five year periods is calculated - these are the long term growth factors.

Prior to this, the traffic engineer specifies limits which will be effective in isolating those exchanges which the Market Planning Branch is forecasting to have relatively high rates of growth for telephone services. A report is produced by the program showing exchanges for which the short term growth in at least one year is expected to exceed a pre-specified limit (e.g. 30%). The long term growth factors are also printed for interest. A similar report is produced for exchanges for which the growth in at least one five year period is expected to exceed a specified test limit.

The exception reports draw the traffic forecaster's attention to the exchanges with the highest anticipated growth since this factor might be expected to have significant consequences on future average traffic rate per service.

(c) The next test is performed to enable the traffic engineering forecasting team to assess whether or not the projected size of the exchange, particularly in the long term, is still expected to be similar to that used previously. In this test, the latest forecast of the number of customers is compared with the previous most recent forecast for each of the standard years. If the percentage variation in any year exceeds the specified lower and upper limits, the test is said to fail and this exchange is placed on an exception report (see Table 2). Otherwise, the exchange is deemed to pass this test, the implication being that the variation between the earlier and present forecast of the number of services is not in itself likely to be substantial enough to affect estimates of the average traffic per service for future years. On the other hand, closer examination of forecast data is indicated as desirable for exchanges which fail the test.

TABLE 2 - Subscribers Forecast Test Exceptions

TEST TOLERANCES : ±30%						
YEAR	1982	1983	1984	1985	1990	1995
EXCHANGE : CAPEL						
Previous Forecast	403	438	475	518	610	740
Current Forecast	385	400	410	420	435	460
Difference(%)	- 5	- 9	-14	-19	-29	-38
EXCHANGE : BEDFORD HARBOUR						
Previous Forecast	27	27	27	27	30	32
Current Forecast	30	32	34	35	45	54
Difference(%)	+11	+18	+26	+30	+50	+68

This examination could either indicate that the new subscribers forecast data are in error and need to be modified or, alternatively, verify the appropriateness of the new data. The latter alternative is more likely. As indicated in Fig. 1, a change in the forecasted number of customers will be expected to affect the average traffic per customer to a greater or lesser extent, depending on the size of the exchange.

In addition to the exception reports, the system also produces a supervisor-oriented report which lists all the exchanges which have any test values outside of the limits, i.e. for any of the three subscriber forecast tests. The report specifies which of the tests each exchange failed.

Stage 2. Traffic Rates Tests

As shown in Fig. 3, in the second stage the latest available traffic measurement data (seasonally corrected) for originating and terminating traffic per service, and originating, terminating and bothway trunk (or junction) traffic per service are compared with the previous most recent forecast for this point of time. A report is produced for each exchange for which new measurement data have been supplied. If the comparison result is outside the limits set it is flagged on this report. A typical report is illustrated in Table 3. The arbitrary limits are varied inversely with the number of services, consistent with the distance between 90% confidence limits in Fig. 1.

TABLE 3 - Traffic Rates Comparison Test Exceptions

EXCHANGE :	CAPEL			
READING DATE :	April, 1982			
CURRENT FORECAST FOR :	1982 made in March 1981			
TEST TOLERANCES :	±10%			
TEST	MEASURED VALUE	CURRENT FORE-CAST	ACTUAL DEVIA-TION	TEST RESULT
Orig Rate(E/Line)	.0307	.038	+ 24%	FAIL
Term Rate(E/Line)	.0243	.031	+ 27%	FAIL
Trunk Orig Rate	.0240	.025	+ 4%	PASS
Trunk Term Rate	.0175	.020	+ 14%	FAIL
Trunk Usage Rate	.0414	.044	+ 6%	PASS
Local Traffic(%)	22.9	20.0	- 13%	FAIL
Subscribers	375	385	+ 3%	PASS

Stage 3. Trunk Circuit Calculations

The program now computes approximate trunk (or junction) circuit requirements for present or future years from knowledge of the expected numbers of services and forecast trunk traffic rates per service. This process is facilitated by the straight-forward trunking used for most of the small terminal exchanges in rural areas of Australia - viz, unidirectional or bothway traffic routes to their parent Minor Switching Centre.

It is intended to augment the system in the future so that a comparison is made of the new forecasts and existing circuits or previous most recent planned circuit requirements, as appropriate. Trunk routes with size variations exceeding pre-specified limits will be reported.

Stage 4. Summary Report of Exceptions

In the next step, a supervisor-oriented report is generated summarising all exchanges for which any test in Stages 1, 2 or 3 produced results outside the set limits.

Stage 5. Iterations

The above procedural steps generally need to be repeated several times, refining data and/or using different limits for the tests in the three stages. During these iterations, the exception and supervisor-oriented reports are invaluable.

Stage 6. Official Forecasts Report

The final version of the forecasts of services, traffic rates and circuit requirements is then produced, and may now rightfully be termed the 'new' official forecast (see Table 4). It includes all exchanges except those requiring further review by the traffic forecasting engineer. For the exceptions, the name of the exchange is printed together with the annotation that the traffic forecast is to be reviewed. (The previous forecast should not, therefore, be used.)

The output illustrated in Table 4 is made available to the various planning and operational staff who may have need of this information as the 'Bulk Accreditation' report. Planning work can proceed using the 'accredited' data in the report, without further reference to the traffic forecasting staff. In the case of the Western Australian rural network, the report contains forecasting data for about 640 telephone exchanges spread over 28 Minor Switching Areas.

Stage 7. Review

Review of exchanges requiring traffic re-forecasts is done later on an individual basis. Complete data for these exchanges are made available to users in the same format as the 'Bulk Accreditation' output report (Table 4).

Stage 8. Reports on Individual Exchanges

In addition to the comprehensive reports mentioned above, reports can readily be obtained for individual exchanges, showing the results for any of the tests or the final output information. This is a valuable aid for specific planning tasks such as the investigation of the impact on trunk circuit requirements which would be caused by the amalgamation of two exchanges.

Stage 9. Macro Forecasting

Since the system is designed to work in the area of detailed micro-forecasting for terminal exchanges, it is complemented by a number of other techniques which provide the forecasting staff with an overall view of the development of traffic usage rates. From time to time, macro trends in data, telecommunication tariffs, overall socio-economic, business and geographic factors are reviewed, and long term models or forecasts are prepared.

Forecasts of traffic between Minor Switching Centres are prepared by manipulation of traffic dispersion matrices.

6. PROPOSED FUTURE ENHANCEMENTS

Comprehensive and extensive comparisons are carried out by the system: 8 tests are performed independently and give rise to many combinations for possible failure. Each of these is defined in the operating manual. This number of error flags constitutes an unnecessary overhead in operating the system. This was deliberately permitted to occur in the early stages to expedite development of the system. One of the enhancements proposed for the system is the simplification and streamlining of the flagging process. It is envisaged that this will be done by both simplifying the system, reducing the number of flags operated on and

TABLE 4 - Bulk Accreditation Output Report

EXCHANGE : KONDININ Minor Switching Area is Narrogin

YEAR	SUBS	TRUNKS								LOCAL			TOTAL		
		OUTGOING		INCOMING		BOTHWAY							TUR AT	ORIGINATING	
		JOR (ME)	TFC (E)	JTR (ME)	TFC (E)	JUR (ME)	TFC (E)	CCTS		LOR (ME)	TFC (E)	CCTS	GBH (ME)	GBHOR (ME)	TRAFFIC (E)
1984	165	20.2	3.33	15.6	2.57	33.2	5.48	11		11.8	1.95	5	57.0	31.0	5.12
1985	170	20.4	3.47	15.8	2.69	33.6	5.71	12		12.0	2.04	6	58.0	31.7	5.39
1986	175	20.5	3.59	15.9	2.78	33.9	5.93	12		12.1	2.12	6	58.7	32.1	5.62
1988	185	20.8	3.85	16.2	3.00	34.5	6.38	13		12.4	2.29	6	60.2	32.8	6.07
1990	200	21.2	4.24	16.4	3.28	35.0	7.00	14		12.6	2.52	6	61.4	33.4	6.68
1995	210	21.7	4.56	16.9	3.55	36.0	7.56	14		13.3	2.79	7	63.5	34.5	7.25
2000	220	22.0	4.84	17.3	3.81	36.8	8.10	15		14.0	3.08	7	64.7	35.0	7.70

Key: JOR - Junction or Trunk Originating Rate per service
JTR - Junction or Trunk Terminating Rate per service
LOR - Local Originating Traffic Rate per service
TUR - Total Usage Rate per service
GBHOR - Originating Rate per service at Group Busy Hour
ME - Milli Erlangs

also by providing some automatic sorting into orders of priority. As more reliable traffic reading data become available through new automatic recording equipment such as CENTOC, and the stability of the forecasts improves, it is envisaged that the need for many of the flags will decline.

The format of the system's Bulk Accreditation report has been designed to aid the (manual) dimensioning work carried out by planning staff. Specific formats are used depending on the particular type of equipment which exists at each exchange. This customising of the data lends itself to a future enhancement of providing more precise automated dimensioning of circuits and exchange equipment. Whilst the equipment quantities may need to be reviewed by those charged with this responsibility, it is expected that such an enhancement would produce a very useful improvement in productivity as has been demonstrated by automated dimensioning facilities for urban networks (Ref. 10).

7. CONCLUSION

The system described simplifies and expedites the work involved in forecasting traffics and dimensioning trunk circuits for a large number of small rural telephone exchanges. This is done with minimal expenditure and is carried out using a computerised system called FERRET which provides volume accreditation of traffic forecasts and extracts exchanges requiring individual review by forecasting staff.

The procedure is not dependent on a particular method for forecasting the traffic usage per service and could be used with any type of forecasting system the administration requires, i.e. rigorous and detailed forecasting work can be carried out on the exchanges highlighted, using whatever degree of sophistication and rigour is warranted. It would be possible, for example, to utilise techniques such as market disaggregation of customers into homogeneous sub-markets (Ref. 8) and Extreme Value Engineering (Refs. 11,12).

8. ACKNOWLEDGEMENTS

The authors wish to acknowledge the significant contribution made to the development, programming, documentation and operation of this system by Geoff Sander, Traffic Engineering Section (Western Australia). Jeff Karutz, Traffic Engineering Section (South Australia) kindly supplied Fig. 1.

9. REFERENCES

1. Hansen, I., Knutsen, K.M., Sole Gil J., and Caballero, P.A., "Computer Aided Planning of the Telephone Network of Rural Areas", Eighth International Teletraffic Congress, 1976, Paper 223.

2. Stave, B., Bastoe, E., and Trosby, F., "Optimal Placement of Exchanges in Rural Areas", Proc. Telecommunications Networks Planning Colloquium, Paris, Sept. 1980, pp. 63-70.

3. United States Congressional Report of the House of Representatives. Committee on Government Operations, 43rd Report. "Maintaining the Network - Rural Telephone Service and FCC Actions" (1982).

4. Fetz, B., Hoekstra, T., Kettler, H., O'Reilly, G., and Vitenas, B., "Planning for Rural Modernization", Bell Laboratories Record, Feb. 1980, pp. 55-62.

5. Elldin, A., "Traffic Engineering in Developing Countries. Some Observations from the ESCAP Region", Eighth International Teletraffic Congress, 1976, Paper 531.

6. Farr, J.P., "Telecommunications Traffic Measurement and Processing in Telecom Australia", Telecom. Journal of Aust., Vol. 29, No. 1, 1979, pp. 70-82.

7. Chemarin, A., "The Metamorphosis of Histograms for Subscribers Telephone Traffic", Telecomm. Journal, Vol. 46-111, 1979, pp. 147-154.

8. Böhn, E., Wacker, W. and Thornton, P., "Causal Models for Forecasting Telephone Services by Market Sectors", Ninth International Teletraffic Congress, 1979, Paper 112.

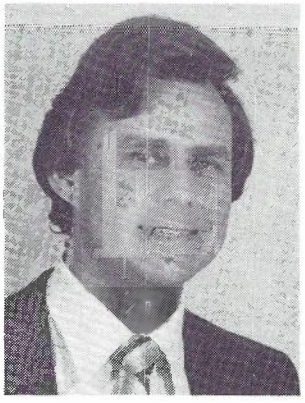
9. Vawser, K.D., "Traffic Usage Estimation for Small Country Exchanges", Telecomm. Journal of Aust., Vol. 25, No. 3, 1975, pp. 209-216.

10. Farr, J.P., "A Computer System for Designing Telephone Networks", Telecomm. Journal of Aust., Vol. 28, No. 1, 1978, pp. 3-8.

11. Barnes, D.H., "Extreme Value Engineering of Small Switching Offices", Eighth International Teletraffic Congress, 1976, Paper 242.

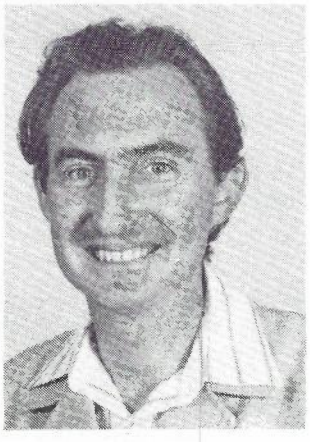
12. Friedman, K.A., "Extreme Value Analysis Techniques", Ninth International Teletraffic Congress, 1979, Paper 313.

BIOGRAPHIES



PETER FARR is District Manager, Central Country of Telecom Australia in Western Australia. A former Executive Aide to the State Manager, he holds degrees in Electrical Engineering, Economics and Applied Science. His special interests in telecommunications include traffic engineering, exchange switching, network performance, data communications, computer systems and global optimisation.

Whilst he has not worked in the Traffic Engineering area since 1972, Mr Farr has maintained a close professional involvement in teletraffic. For instance, he gave a paper on modular engineering of junction groups at ITC 8; represented Australia at a 1977 seminar sponsored by the ITC on SPC exchanges and data networks; presented a paper to ITC 9 at Torremolinos, Spain; and presented the above paper to ITC 10 in Montreal in June, 1983.



ERROLD WALDRON was born in Adelaide and joined the PMG Department as a JPO and subsequently became a Technician-in-Training. He undertook part time studies in Engineering at the Western Australian Institute of Technology and graduated in 1972. After obtaining broad experience in several areas he joined the Traffic Engineering Section in 1974. He has been involved in various aspects of teletraffic including forecasting and culminating in working as the Supervising Engineer of that section. His current responsibilities are the Pier Trunk Switching Centre and Equipment Design Centre in the Trunk Service Section of the Western Australian administration, Telecom Australia.

Information for Authors

ATR invites the submission of technical manuscripts on topics relating to research into telecommunications in Australia. Original work and tutorials of lasting reference value are welcome.

Manuscripts should be written clearly in English. They must be typed using double spacing with each page numbered sequentially in the top right hand corner. The title, not exceeding two lines, should be typed in capital letters at the top of page 1. Name(s) of author(s), in capitals, with affiliation(s) in lower case underneath, should be inserted on the left side of the page below the title. An abstract, not exceeding 150 words and indicating the aim, scope and conclusions of the paper, should follow below the affiliation(s).

ATR permits three orders of headings to be used in the manuscript. First-order headings should be typed in capitals and underlined. Each first-order heading should be prefixed by a number which indicates its sequence in the text, followed by a full stop. Second-order headings should be underlined and typed in lower case letters except for the first letter of each word in the heading, which should be typed as a capital. They should be prefixed by numbers separated by a full stop to indicate their hierarchical dependence on the first-order heading. Third-order headings are typed as for second-order headings, underlined and followed by a full stop. The text should continue on the same line as the heading. Numbering of third-order headings is optional, but when used, should indicate its hierarchical dependence on the second-order heading (i.e. two full stops should be used as separators).

Tables may be included in the manuscript and sequentially numbered in the order in which they are called up in the text. The table heading should appear above the table. Figures may be supplied as clear unambiguous freehand sketches. Figures should be sequentially numbered in the order in which they are called in the text using the form: Fig. 1. A separate list of figure captions is to be provided with the manuscript. Equations are to be numbered consecutively with Arabic numerals in parenthesis, placed at the right hand margin.

A list of references should be given at the end of the manuscript, typed in close spacing with a line between each reference cited. References must be sequentially numbered in the order in which they are called in the text. They should appear in the text as (Ref. 1). The format for references is shown in the following examples.

> Wilkinson, R.I., "Theories for Toll Traffic Engineering in the U.S.A.", BSTJ, Vol. 35, No. 2, March 1956, pp.421-514.
>
> Abramowitz, M. and Stegun, I.A., (Eds), Handbook of Mathematical Functions, Dover, New York, 1965.

Three copies of the paper, together with a biography and clear photograph of each of the authors should be submitted for consideration to the secretary (see inside front cover). All submissions are reviewed by referees who will recommend acceptance, modification or rejection of the material for publication. After acceptance and publication of a manuscript, authors of each paper will receive 50 free reprints of the paper and a complimentary copy of the journal.

Benefits of Authorship for ATR.

- ATR is a rigorously reviewed journal with international distribution and abstracting.
- Contact with workers in your field in the major telecommunication laboratories in Australia can improve interaction.
- Contributions relevant to the Australian context can contribute to the viability and vitality of future Australian research and industry.
- Publication is facilitated because ATR arranges drafting of figures at no cost to the author, and with no need for the author to spend time on learning journal format standards.

Printed by Frank Daniels Pty. Ltd.

ATR AUSTRALIAN TELECOMMUNICATION RESEARCH
ISSN 0001-2777

VOLUME 18, NUMBER 1,
1984

Titles (Abbreviated)