TRANSMISSION SYSTEMS
FOR COMMUNICATIONS
VOLUME I

VOLUME I

TRANSMISSION SYSTEMS FOR COMMUNICATIONS

BELL TELEPHONE LABORATORIES

TRANSMISSION SYSTEMS FOR COMMUNICATIONS

By Members of the Technical Staff
Bell Telephone Laboratories

VOLUME I

Bell Telephone Laboratories, 463 West Street, New York, N. Y.

Foreword

Transmission systems are complex aggregates of electronic and mechanical gear. Their design, manufacture, application and operation involve the coordination of a wide range of scientific and engineering considerations by diverse groups of people. The greatest effectiveness in such an activity will be realized if the people involved have a commonly held understanding of underlying principles.

Bell Telephone Laboratories has two graduate programs in communications technology: the Communication Development Training Program for new employees of the Laboratories; and the Operating Engineers Training Program for engineers from the Bell System Operating Companies. Each of these programs includes a course in transmission systems. While in their future careers these two groups will have different roles to play in dealing with transmission problems, we believe that much can be gained by presenting an integrated view of the subject. We hope this text and courses based on it will help to strengthen and unify the various transmission activities in the Bell System.

M. B. MC DAVITT
Director of
Transmission Development

Preface

Objectives

Transmission is the part of communications engineering concerned with transmitting messages between sources and receivers. The purpose of this text is to provide material on which communication engineers can base a clear philosophy of transmission systems. This will require ability to determine the basic factors in any problem; knowledge of and analytical facility with major concepts; and working familiarity with important methods, details, and vocabulary.

Plan of Text

Transmission, like other fields of engineering, is composed of a central core of method and specialized knowledge surrounded and supported by numerous related scientific and engineering disciplines. Electromagnetic theory, active and passive network theory, various branches of mathematics, and the physics of active devices are a few of the scientific fields that have a large bearing on transmission. Likewise, general methods of analyzing systems and the broad viewpoint supplied by information theory contribute to an understanding of transmission systems. Although this text will touch on many of these topics, the reader must go beyond it for detailed development of the parts these supporting fields play in particular transmission problems.

Presenting a compact and coherent view of transmission requires integrating the parts of the subject, developing a balanced treatment of the important ideas and adapting them to the evolving problems of the communications industry, and establishing basic methods applicable to both old and new transmission techniques. We need precise information and vocabulary with which to think. At the same time we need to generalize and carry ideas from one part of the subject to others. The essence of the method employed in this text is to present selected information, vocabulary, and analysis covering the major problems and techniques of transmission. This material provides the base for delineating general principles and methods that have broad application.

Much of the new thinking about transmission will have its roots in past accomplishments in the field. Furthermore, it is not practical to discuss new situations in terms of details that are still to be evolved. Thus, the material in this text is largely composed of techniques and methods that have been applied to existing transmission problems. It is expected that a thorough grasp of this background will provide the basis for systematically considering the application of new technology to old and new problems.

In making a broad assessment of a transmission problem it will be helpful if the parts played by the following factors are examined:

1. The structure of the network in which the information bearing signals are to be transmitted.

The importance of this factor can be illustrated by examining the requirements on a channel which must be switched automatically every few minutes to form a part of a new chain of channels and comparing them with the requirements on a channel permanently connecting two fixed points.

2. The characteristics of the message to be transmitted.

(As used here and at other points in the text "message" is defined as the original acoustic wave as spoken, a scene as viewed, or a written page) The structure of the message , and the reaction of the user of the message to distortion and interference added during transmission, form the basis for requirements imposed on transmission channels.

3. The characteristics of the signal used to transmit the message.

The signal into which a message is translated will have a form and sensitivity to interference and distortion which is different from the original message. Much of transmission technology deals with matching signal characteristics to physically realizable system techniques so as to achieve a desired quality of message transmission.

4. The performance of the components of transmission systems.

The performance that can be achieved, expressed in such terms as power capacity, noise, bandwidth, and stability is limited by both physical and practical considerations. In addition, a wide variety of adjustments can be made among performance factors and components of the system. These adjustments depend on our ability to evaluate the ways in which specific mechanisms - such as non-linearity in an AM system amplifier, or imperfect equalization in an FM repeater - react on the signals in the system and ultimately on the message.

The way in which these factors are handled in the text can be seen from the summary of chapters which follows this preface. The first chapter deals with the structure of the Bell System telephone and television networks and their impact on the performance required from specific links and components. The second chapter discusses the nature of the telephone message and the way in which requirements on telephone channels are derived. The third chapter deals with voice frequency transmission in terms of how the signal is derived, the performance of transmission facilities and the nature of degradations experienced by the signal.

The problems of transmitting many telephone messages on an amplitude modulated carrier signal are considered next. In Chapter 4 the preparation of the signal is discussed. This is followed, in Chapter 5, by a summary of the principal problems encountered in AM systems. Chapters 6 to 15 give a much more comprehensive treatment of methods used in the design of AM carrier systems. These may be omitted by the reader who is not concerned with detailed design problems.

Consideration of television transmission in Chapter 16 provides an oportunity to examine an entirely different type of message and signal. Similarly, frequency modulated microwave radio systems and pulse modulated systems provide radically different ways of transmitting messages. The signal composition, performance characteristics, and signal degradation mechanisms of these methods of transmission are considered in Chapters 17-28. In the final chapter it is shown how communication theory relates message characteristics, signal characteristics and system performance.

<u>Transmission - An Evolving Field</u>

Transmission is not a static field. Prior to 1930, it was largely concerned with defining bandwidth, loss and interference objectives for telephone and telegraph service and learning to meet them with an economical combination of transmitters, receivers, wire lines and amplifiers. In the following two and a half decades, the bulk of the advanced work in transmission involved designing, manufacturing, and operating carrier systems. The provision of large numbers of economical long and short haul channels, including both message and special service facilities, was the primary objective. From both design and operating viewpoints, the transmission systems required to meet these objectives are well defined entities meeting specified requirements. As a result, the inter-relationships among these toll systems and the connecting local plant were treated as particular and in most instances, not controlling problems.

The growing demand for high quality channels and the advent of new technology is rapidly leading to a new era of transmission. In the coming decade, this will be characterized by the application of new techniques to improve the quality and increase the quantity of both long distant and local facilities, by the increased complexity of plant used in making a call with an attendant large emphasis on engineering for maintenance, and by provision of new services.

Acknowledgements

The present text has evolved over a period of years during the presentation of courses to Bell Telephone Laboratories and Bell System Operating Company groups. Many individuals have provided ideas and written material on which the text is based. Major portions of the text have been written by R. H. Klie, L. H. Morris, R. A. Kelley, E. F. O'Neill, R. W. Hatch, E. E. David, Jr., M. R. Aaron, D. R. Jordan, R. L. Easton and G. R. Leopold. In addition to contributing particular sections, L. H. Morris was responsible for a large part of the editing and rewriting needed to unify and clarify the many individual efforts involved.

C. H. ELMENDORF

SUMMARY OF CHAPTERS

VOLUME 1

Chapter 1

Transmission System Environment

The composition of the Bell System plant is reviewed in terms of the types of transmission facilities used and the ways in which they are interconnected.

Chapter 2

Message Channel Objectives

The Bell System objectives for loss, noise, crosstalk, and echo in message circuits are stated, and the statistical nature and subjective foundation of these objectives are discussed.

Chapter 3

Voice Frequency Transmission

The voice frequency components of the telephone plant - subsets, loops, vf trunks and repeaters - are described. Voice frequency transmission characteristics, noise sources in the local plant, and crosstalk are discussed.

Chapter 4

Amplitude Modulation

Amplitude modulation and demodulation are analyzed, and various forms of AM signals are discussed. Emphasis is placed on the preparation of telephone message signals for transmission over carrier systems.

Chapter 5

Introduction to AM Carrier Systems

The building blocks of AM carrier systems are described. The chapter summarizes many of the important problems encountered in the design and engineering of these systems and serves as an introduction to the material that follows in Chapters 6-15.

Chapter 6

System Layout Terminology

This chapter collects important terminology used in Chapters 6 - 15 and introduces the reader to the problem of detailed system analysis.

Chapter 7

Random Noise

Sources of random or thermal noise in AM systems are discussed; formulae for computing tube noise, and methods of estimating noise figure of repeaters and the addition of noise in a string of repeaters are given.

Chapter 8

Modulation Distortion

Cross modulation between channels arising from non-linearity in an AM system is analyzed. The relation between the power series representation of the non-linear device, and the overall intermodulation performance of an AM multi-repeatered system is developed.

Chapter 9

Load Capacity, Gains and Losses

System load and overload are defined in terms of an equivalent single frequency sinusoid. Equality of repeater section transmission path loss and repeater gain is shown to be an important objective.

Chapter 10

System Layout and Analysis

The material developed in Chapters 6 through 9 is used to illustrate the problems of setting repeater spacing, system levels, and analyzing system performance.

Chapter 11

Misalignment

The problem of systematic misalignment - all repeaters slightly too high or all repeaters slightly too low in gain - is analyzed and the necessarily adverse effect on signal-to-noise ratio studied.

Chapter 12

Overload and Modulation Requirements

Methods of deriving overload and intermodulation requirements for a system from a knowledge of the speech load are studied. It is shown that the peak value of the voltage wave corresponding to a telephone multiplex signal can be expressed in terms of a sine wave having the same peak voltage; this concept is also made use of in FM systems later. Methods of computing modulation noise developed here are similarly adaptable to FM system problems.

Chapter 13

Feedback Repeater Design

The problems of working through a feedback repeater design from its initial conception to its final form, and estimating the repeater performance throughout the design process, are discussed as an example of the interdependence of device development objectives, circuit design and system performance.

Chapter 14

Regulation and Equalization

Requirements on the transmission-frequency characteristic for telephone and television transmission are discussed, and methods for equalizing and regulating to meet these requirements are described. The frequently unexpected impact of the equalization plan on other aspects of system performance illustrates the complex nature of the system problem.

Chapter 15

Shaped Levels, Feedback, Compandors, TASI

The effect of shaped feedback and pre-emphasis of the telephone multiplex load on repeater noise, intermodulation, and overload is discussed. The problems and advantages of compandors are described. The principle of time-sharing of channels is introduced.

VOLUME 2

Chapter 16

Television Transmission

The nature of the television signal, its sensitivity to interference, and the resulting requirements on transmission systems for this signal are discussed.

Chapter 17

Introduction to Microwave Systems

The building blocks of a radio system are described. Some similarities and differences between radio and wire systems are discussed.

Chapter 18

Radio Propagation

Antenna gain and path loss relations are analyzed. Characteristics of typical antennas and the problems of fading and absorption are discussed.

Chapter 19

Properties of the Frequency Modulated Signal

The spectrum of a carrier which is phase or frequency modulated by one or more sinusoidal signals is derived. The spectrum resulting from angle modulation by a band of random noise representing a telephone multiplex signal is given.

Chapter 20

Random Noise in FM and PM Systems

The method of analyzing the noise performance of an FM or PM system is given. The noise advantage of FM over AM systems is derived, and shown to be an example of the principle of trading bandwidth for signal-to-noise ratio.

Chapter 21

Use of the Fourier Transform for Transmission System Analysis

The Fourier Transform is reviewed at this point to serve as a tool for analyzing subsequent FM and PCM material.

Chapter 22

Effect of Transmission Deviations in PM and FM Systems

The methods of analyzing the effects of transmission deviations in an FM or PM system are presented.

Chapter 23

Frequency Allocation

The factors effecting choice of baseband width and the mechanisms of interchannel interference are discussed. Frequency allocations of present radio systems are illustrated.

Chapter 24

Illustrative Radio Systems Design Problem

The material in the previous chapters is summarized by applying it to the analysis of a short haul 100 channel system.

Chapter 25

The Philosophy of Pulse Code Modulation Systems

A general introduction is given to the principles of message sampling, quantizing, coding, decoding, and reconstruction. Time division multiplex and the trading of bandwidth for signal-to-noise ratio are examined for a PCM system, and the results are related to previous discussion of AM and FM systems.

Chapter 26

Preparation and Processing of Signals in PCM

The spectrum of a sampled message is examined to introduce the problem of filter requirements. This is followed by a description of the terminal equipment and a discussion of estimated noise performance of a 24 channel system.

Chapter 27

Pulse Transmission and Reshaping

High-end shaping and transmission deviations are analyzed in terms of error rate. Methods of compensating for the effects of low frequency suppression in transmission systems are discussed.

Chapter 28

Regeneration and Retiming

Ideal vs. partial regeneration and retiming are studied in terms of the system error rate. The advantages of a regenerative system over a conventional AM or FM system are discussed.

Chapter 29

Signal Processing

The nature of speech is discussed, and methods which have been devised to extract and transmit only the information content of the message are examined.

Chapter 1

TRANSMISSION SYSTEM ENVIRONMENT

Introduction

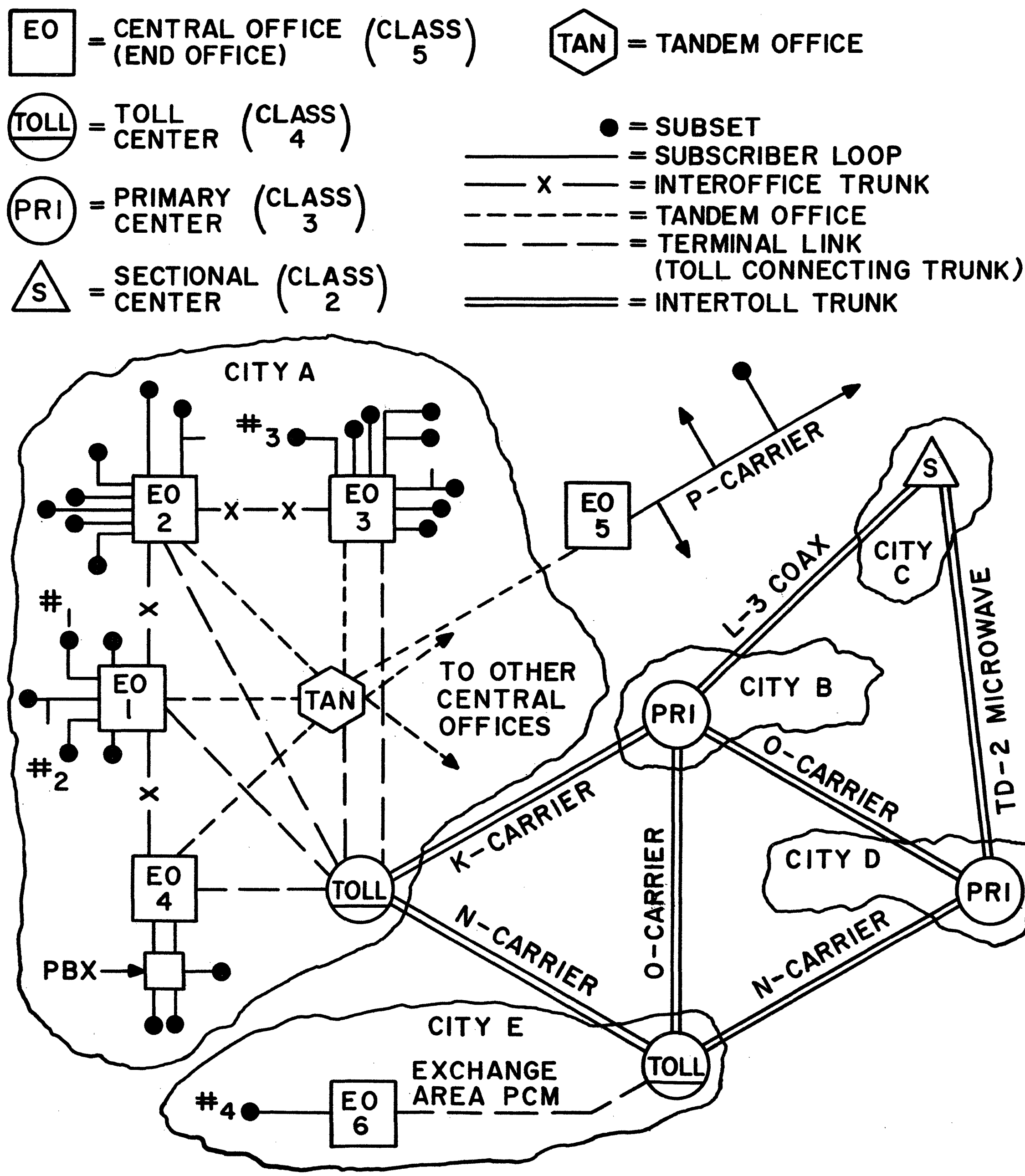
In providing the telephone service which permits people to talk together at a distance, the telephone system must supply the means and facilities for connecting together the particular telephone stations at the beginning of the call and disconnecting them when the call is completed. Both switching and transmission problems are involved. The switching problem is concerned with selecting and connecting together the customers and transmission path, and includes supplying and interpreting the control and supervisory signals needed to perform this operation. The transmission problem, which is the concern of this text, deals with the transmission of these control signals and the customer's message.*

A transmission system in its simplest form is a pair of wires connecting two telephones. More commonly, the term "transmission system" is used to denote a complex aggregate of electronic gear and the associated medium which together provide a multiplicity of channels over which a number of customer's messages and the associated signalling information can be transmitted.

In general, a call between two points, in any but a purely local situation, will be handled by connecting a number of different transmission systems in tandem to form an overall transmission facility between the two points. The way in which these systems are chosen and interconnected has a strong bearing on the characteristics required for each system. Similarly, the way in which television and many other special service facilities are built up reacts on the performance required for each transmission system used to carry these types of services. This is so because each element which forms a link in the overall transmission facility between two points will add some degradation to a message. It follows that the relationship between performance and cost of a transmission system cannot be considered only in terms of the isolated system but must also be viewed with respect to the relation of the system to the building up of a complete facility. Thus, the solution of design, manufacturing, and operations problems for a specific system will more often than not require knowledge of other systems in the telephone plant, both existing and planned.

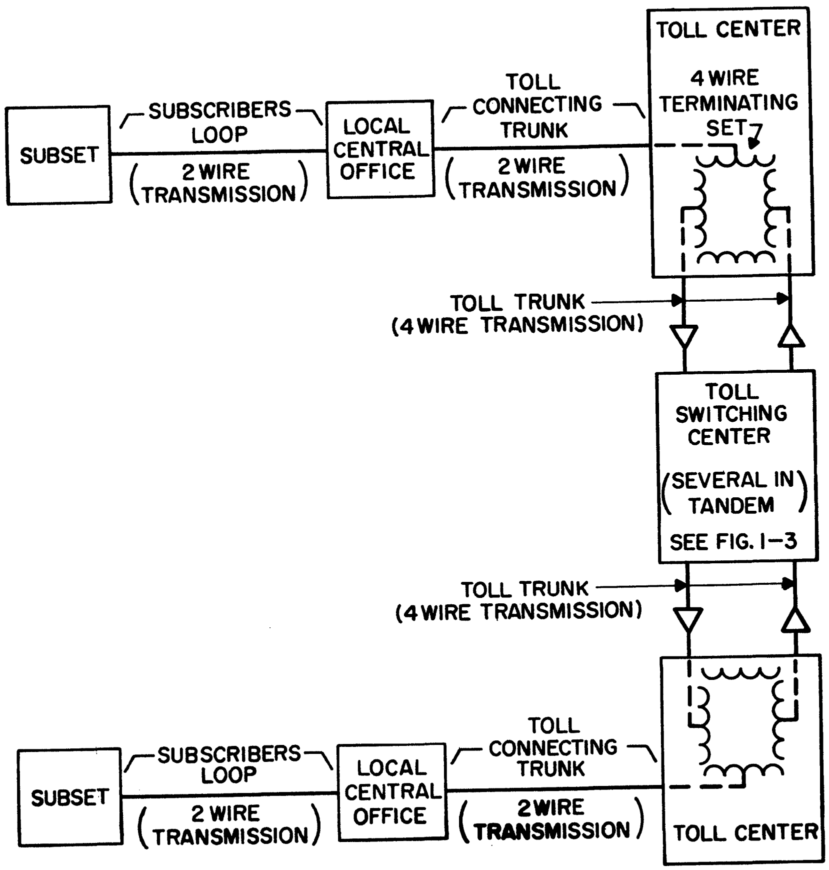
---

*The term message as used in this chapter denotes any information a customer might want to get from one place to another, e.g., speech, TV, data, etc.



A Simplified Telephone System

Figure 1-1



Inter-City Subscriber To Subscriber Connection

Figure 1-2

<u>Telephone Service - Connection Description</u>

An example of the large variety of facilities that might be used in handling a telephone call is illustrated by Figure 1.* A block schematic of a typical long connection is shown on Figure 2. A connection may involve voice frequency transmission between subsets through a single central office or a multiplicity of links including several offices, voice frequency paths and carrier systems.

The subscriber's subset modulates a direct current (usually transmitted from the central office) with the acoustic speech message to be transmitted. The subset also demodulates the received signal and returns it to its acoustic form. In addition, it generates supervisory signals (on-hook and off-hook) and switching signals (dial

*For latest terminology and more detailed discussion see "Notes on Distance Dialing" AT and T circular file.

pulses). There are a variety of subsets in use, each having somewhat different frequency response and transmitter and receiver efficiency. The importance of these factors is discussed in Chapter 2 while the characteristics of the subset are covered briefly in Chapter 3.

The subscriber's loop provides a two way path for the speech signals and the ringing, switching, and supervisory signals. Since the subset and subscriber's loops are permanently associated, their combined transmission properties can be adjusted to meet their share of the message channel objectives discussed in Chapter 2. For example, the greater efficiency of the 500 type subset is used to permit increased loop loss due to longer distances or finer gauge wire. The performance and limitations of loops will be discussed in Chapter 3.

The small percentage of the time (of the order of 2%) that a subscriber's loop is used has led to the consideration of "line concentrators" for introduction between the subscribers and the central office. In effect, the concentrator is a partial central office, and the pair which connects it to the true central office, and which formerly was part of the loop plant, must now be thought of as a trunk. The essential difference between a loop and a trunk is that a loop is permanently associated with a particular subscriber and subset, whereas a trunk is a common usage connection.

Trunks of various types are used to inter-connect central and toll offices. An inter-office trunk connects a local central office to another central office, a tandem trunk connects an end office to a tandem office, while a toll-connecting trunk connects a local office to the toll office.* In toll language, toll connecting trunks are also described as "terminating links".

Up to the point where the signals are connected to inter-toll trunks in the toll office, the message and supervisory signals may be handled on a two-wire voice frequency basis (that is, the same pair of wires is used for both directions of transmission), by a rural carrier (used as a loop), or by short haul N carrier or, in the future, PCM (used as toll connecting trunks). At the toll office, after appropriate switching and routing, the signals are generally connected to the toll trunks by means of a four-wire terminating set, which splits apart the two directions of transmission so that the long haul transmission may be accomplished on a four-wire basis. (Some short intertoll trunks are handled by two-wire transmission.) Through these intertoll trunks, the signals are transmitted to remote toll switching centers (which in turn are connected by intertoll

*In general, local offices connect to toll offices by direct trunks rather than by way of tandem offices.

trunks to other switching centers) and ultimately reach the recipient of the call through another four-wire terminating set and other local switching equipment, toll connecting trunks and a final subscriber's loop.

The types of facilities that might be involved in various connections can be seen by reference to Figure 1. The simplest connection would be a call between subsets 1 and 2, both working out of End Office 1,* in which no trunks would be involved. An inter-office call between subsets 1 and 3 in City "A" would use two trunks, the connection being made via a tandem office. These trunks would normally be voice frequency circuits, possibly equipped with negative impedance repeaters; or a short haul carrier such as N, or, in the future, transmission might be via a pulse code modulation system.

Next, consider a call between Subset 1 in City A and Subset 4 in City E. The path begins at a subscriber loop working into End Office 1. From there it uses a voice frequency toll connecting trunk to the toll center. In the skeletonized system shown here, there are two paths to the toll center in City A (one through the tandem office). Between City A and City E there are a number of routes. If the two cities have a high community of interest, there would be direct trunks between them. Figure 1 shows that in this case the two cities are linked by, for example, "N" Carrier. An alternate route which happens to employ "K" Carrier is also shown, via a primary center. Out of this Primary center (Class 3) there might be direct, high-usage trunks on "O" Carrier to City E. Alternatively, use would be made of "final" trunks to a Sectional center (Class 2) at City C, from which connection might be made to City E through another Sectional center and another Primary center. These latter trunks might be on an L3 coaxial system or a TD-2 microwave radio system.

Toll Switching Plan

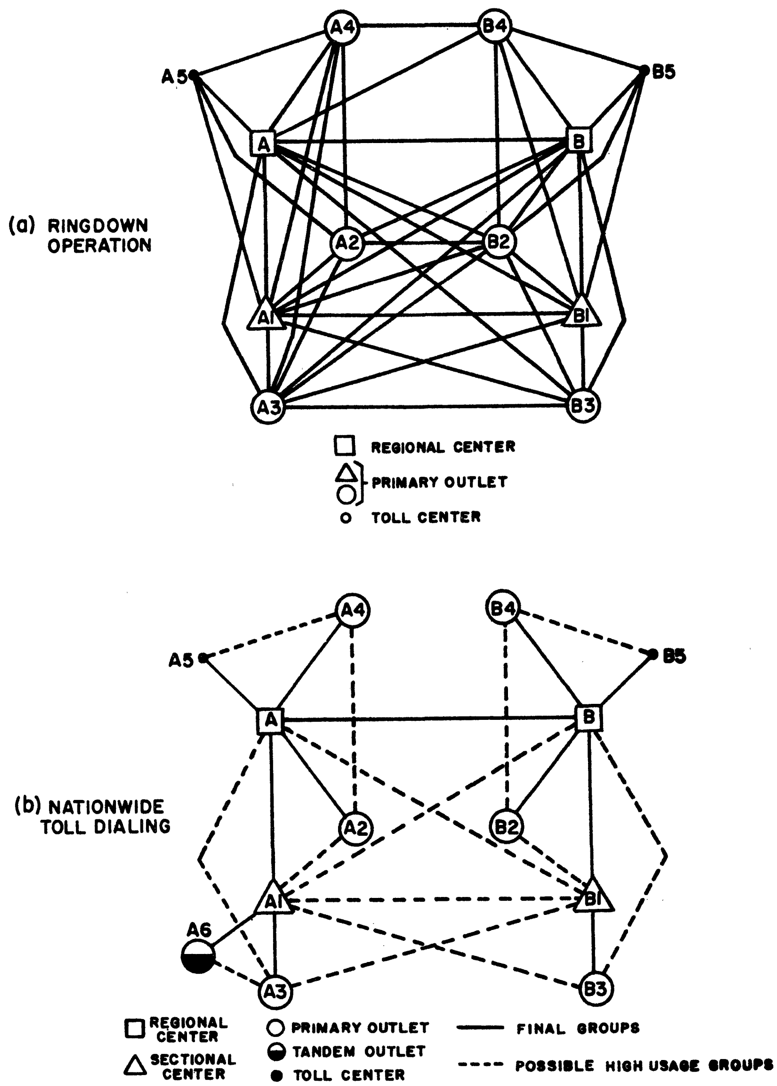
The plan used to connect together toll offices has a large bearing on the performance required of both local and toll transmission systems. In early practice, toll circuits were operated manually by operators on a so-called "ring down" basis. With such arrangements, the number of circuits that could be connected in tandem was severely limited, and comparatively limited use was made of alternate routing. Speed of service was comparatively slow and trunks were inefficiently used in many cases.

The introduction of automatic switching of toll circuits permits calls to be switched so rapidly and cheaply that, by using alternate routes, circuits can be laid out and built up much more economically.

- -

*End office is the designation given to the lowest rank office in the Toll Switching Plan, as discussed in the next section.

An example of the impact of toll dialing on the trunk layout is shown on Figure 3. The upper diagram (a) shows the circuit groups which would be required to handle an assumed flow of traffic on a ring down manual basis. The lower diagram (b) shows the circuit groups that would be required for the same traffic using toll dialing. Direct connection between offices is provided where heavy traffic is expected (the "high usage" groups) but alternative "final" (or "last resort") routes are provided to handle overflow traffic. In (a) there are 42 different circuit groups. In (b) there are 26 circuit groups which are larger and used on a more efficient basis.

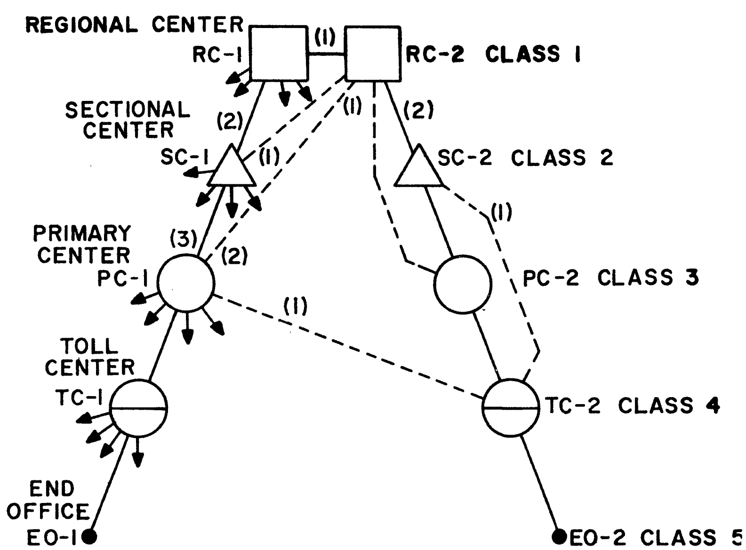


Typical Intertoll Trunk Networks

Figure 1-3

In the present toll switching plan there are five ranks or classes of switching centers. The highest rank is the Regional Center. The lowest rank, called the End Office, is the telephone exchange in which the subscriber's loops terminate. The chain of switching centers and an illustration of how a call might be routed is shown on Figure 4. The order of choice at each control center is indicated on the figure by the numbers in parentheses. In the example there are ten possible routes for the call, only one of which requires the maximum of seven intermediate links.* Note that the first choice route involves two intermediate links. In many cases, a single link exists between the two toll centers which would be the first choice.

The probability that a call will require more than N links in tandem to reach its destination decreases rapidly as N increases from 2 to 7. First, a large majority of toll calls are between end offices associated with the same regional center. The final routes in these cases



NOTES:
1. NUMBERS IN ( ) INDICATE ORDER OF CHOICE OF ROUTE AT EACH CENTER.
2. ARROWS FROM A CENTER INDICATE TRUNK GROUPS TO OTHER LOWER RANK CENTERS THAT HOME ON IT. (OMITTED IN RIGHT-HAND CHAIN)

Illustration Of Choice Of Routes On Assumed Call

Figure 1-4

---

*In this discussion, the number of links refers to the number of toll trunks in tandem and does not include the two "terminating links" (toll connecting trunks) at the ends of the connection.

will not extend as far as the Regional Center, and may not involve even the next lower ranking centers in the chain. The maximum number of toll trunks in these connections is therefore less than seven. Secondly, even a call between telephones associated with different regional centers is routed over the maximum of seven intermediate toll links only when all of the normally available high-usage trunk groups are busy. The probability of this happening in the case illustrated in Figure 4 is only p^5, where p is the probability that all trunks in any one high-usage group are busy. Thirdly, many calls do not originate all the way down the line since each class has all the duties (in its area) of all lower classes except 5.

The following table makes these points more concrete. The middle column of this table shows, for the fictitious system of Figure 4, the probability that the completion of a toll call will require N or more links between toll centers, for values of N from 1 to 7. In computing these probabilities the idealized assumptions are: that the chance that all trunks in any one high-usage group are simultaneously busy is 0.1; that the solid line routes are always available; and that of the available routes, we always select the one with the fewest links. The figures in Table 1 illustrate how increasingly unlikely are the connections requiring more and more links. These numbers are, of course, highly idealized and simplified. Actual figures from a Bell System survey made in 1946 and 1947 are shown in the last column of Table 1. Note that at that time 80 percent of the calls were completed over only one intermediate link (which is not possible in the system shown in Figure 4, which does not show a direct trunk between toll centers) and that as many as 7 intermediate links were required in only 3 out of 100,000 calls.

Table 1-1

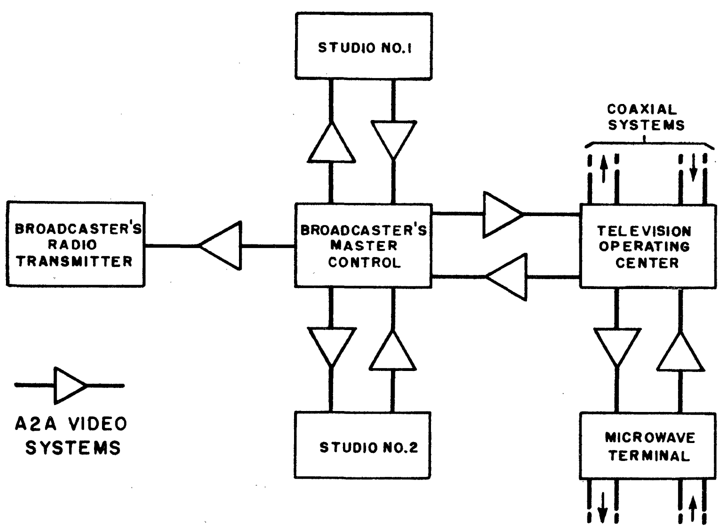
Probability that N or more links will be required to complete a toll call.

No. of Intermediate Links, N	Probability: Fig. 1-3	Probability: Bell System Data
Exactly 1	0	0.8
2 or more	1.0	0.2
Exactly 2	0.9	-
3 or more	0.1	0.03
4 or more	0.1	0.003
5 or more	0.010,9	-
6 or more	0.001,09	-
Exactly 7	0.000,01	0.000,03

It is apparent that the switching pattern that has been described imposes strict transmission requirements on the toll trunks. Up to seven toll trunks may be connected in tandem, and successive calls between the same two telephones may take different routes and encounter different numbers and kinds of circuits. The loss must not be excessive when calls are routed over the maximum number of links, and there should not be too great a variation in the transmission afforded over the different possible routes that a call might take. Loss cannot be permitted to get too low either or echo, singing, crosstalk, and noise can cause excessive transmission impairment. Design and maintenance must both recognize that if unsatisfactory transmission should occur it will not be observed by an operator as in the past, and that the customer's attempt to call unsatisfactory transmission to an operator's attention will disconnect the circuit complained of. Identification of the source of trouble thereupon will be very difficult.

Television Service

Television transmission in the Bell System involves connecting together studios, the broadcaster's master control center, transmitters, and telephone company television operating centers within cities and then inter-connecting cities by means of nationwide television facilities. Figure 5 shows a typical intra-city layout for a large broadcaster. The local television links are usually A2 or A2A video systems. Two-way connections between the master control location and the studio are often required for programming purposes. For example, filmed material from the



Intra-City Television Circuits

Figure 1-5

master control location may be sent to the studio, there to be combined with the live program and returned to the master control room over a second circuit. For local broadcast a third circuit to the radio transmitter is required. For network operation, connecting circuits are required between the master control room and the telephone company's television operating center (TOC) where connection to the inter-city facility is made.

Simpler layouts are found at cities where programs do not normally originate. Such cases require only video systems to carry the programs from the radio or coaxial system terminals to the broadcasting station, with means for selecting the program scheduled for broadcast.

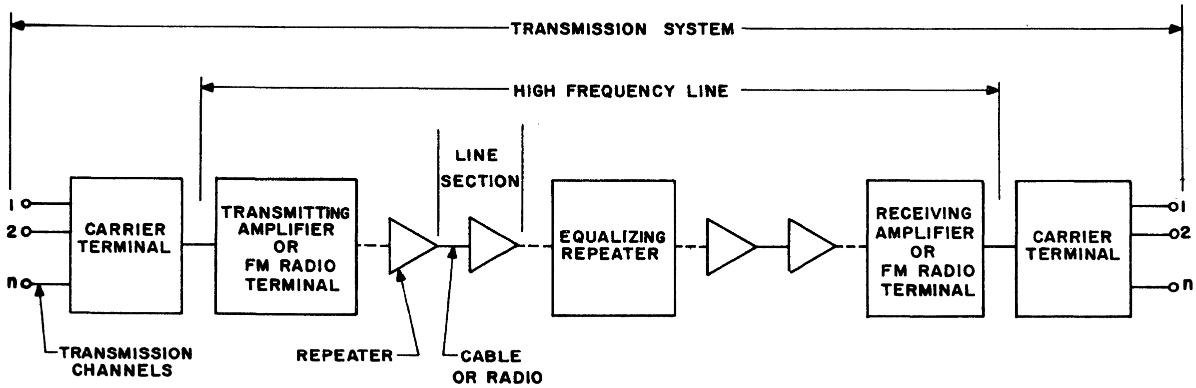
The inter-city channels may be either direct connections between cities or round robin channels with cities fed by circuitous routes. These connecting facilities are formed and re-formed each day depending on the broadcaster's requirements. Thus, links must be connected in tandem in different ways on a day-to-day or hour-to-hour basis. It is, therefore, not usually practicable to line up or equalize on an overall basis. Instead, each link must be capable of a transmission quality such that when all the necessary links are connected in tandem, the signal will have a very small probability of being unduly degraded.

Carrier and Radio Transmission Systems and Their Components

The overall communication channel connecting together two message or television customers is seen to be composed of numerous links, each of which, in general, is itself a complete channel provided by a transmission system. The transmission system may be a pair of wires with or without amplifiers, or a carrier system providing anywhere from a few to many hundreds of telephone channels combined into a single "broadband channel", or it may be a video or carrier system providing a single television channel.

While message systems are normally designed and operated to supply two-way communications, the fact that four wire transmission is used in most of the toll plant (and increasingly in the local plant) makes it possible to analyze such systems in terms of one-way transmission. Similarly, television systems can be analyzed in terms of one-way transmission, even though network operations may at times require a return path over another one-way transmission system.

The components of an AM or FM carrier transmission system to supply the one-way channels that form the links in the overall customer-to-customer telephone channels are shown on Figure 6. The incoming



One-Way Transmission System

Figure 1-6

signals are the voice frequency message signals from the four wire terminating set of Figure 2. The carrier terminals limit the frequency band and signal amplitude and translate the voice frequencies through one or more steps of modulation to the desired frequency in the broadband channel. The carrier terminals typically include modulators, demodulators, amplifiers, pads, and filters. Signal-to-noise penalties, unwanted modulation products, carrier frequency stability, and stability of gains and losses versus time are important factors in such terminals. From these terminals, the broadband channel may then be transmitted directly on to a wire line as in amplitude modulated wire systems or it may frequency-modulate a microwave carrier for transmission on a radio system. The broadband signal is then transmitted through many more or less identical repeater sections composed of lossy wire or radio media and compensating repeaters. After the signals have passed through a number of repeater sections, a more complicated repeater is sometimes required in which the signal can be equalized, i.e., in which accumulated deviations from the desired transmission characteristic can be compensated for by adjustable networks. Finally, a point is reached where at least some of the circuits are to be dropped. At this point the signals must be demodulated back to voice frequency, making use first of an FM terminal if it is a radio system and in all cases by means of appropriate carrier terminals.

If a television channel is being transmitted on an AM carrier system, the carrier terminal translates the video signal to the desired place in the spectrum and then basically the same microwave or coaxial high frequency repeatered line may be used, except that now delay equalizers must also be provided, to correct accumulated delay distortion due to line, repeaters and terminals.

Just as the need to transmit TV set new requirements so the expected growth of data transmission as well as other projected new types of services will place new demands on the plant in terms of the signals it must be capable of transmitting and the connections it must be possible to set up.

Impact of Links and Section Multiplicity

In the preceding sections we have seen that customer-to-customer communications channels can involve a multiplicity of different links connected together in many ways. We have seen that local plant is basic to every connection; its efficiency and uniformity, with respect to loss, noise, and impedance (to mention but a few of the factors that must be considered) affects the entire system. It has also been pointed out that the longer transmission systems used to interconnect central and toll offices, link by link, will often include terminals and sections of line which in turn are composed of numerous more or less identical repeater sections.

This composition of the overall channel gives rise to two problems that have a major bearing on everything that is done in connection with transmission systems. In the first place, the accumulation of performance imperfections from a large number of links and sections leads to severe requirements on individual units and to great concern with the mechanisms causing imperfections and the ways in which imperfections accumulate. In the second place, the variable complement of links forming overall channels makes the problem of allocating tolerable imperfections among links on an economical basis quite complex. Deriving objectives for a channel of fixed length and composition is a problem in customer reactions and economics. However, when the channel objectives must be met for widely varying lengths of channels and composition of links, the problem of deriving objectives becomes an even more complex statistical study involving considerable knowledge of plant layout, operating procedures and the performance of other systems. This will be taken up in the next chapter.

Chapter 2

MESSAGE CHANNEL OBJECTIVES

How well a customer can talk and hear over a channel, and his opinion of the grade of transmission, will depend on:

a) Received speech power - which is a function of the efficiency of the transmitter and receiver and of the electrical loss between them, as well as the acoustical power of the talker.
b) The bandwidth transmitted.
c) The amount and character of the noise introduced.
d) The cross-talk that he hears, expecially cross-talk which either is intelligible or seems nearly so.
e) The magnitude and delay of the echo (of his own voice) that he hears.
f) Most important, perhaps, the customer's expectations. The need for up-grading our objectives to keep pace with the customer's rising standards is an important aspect of our work.

There are other important imperfections which should be considered in a full study of message channel objectives. Tones of various frequencies and character, unintelligible crosstalk of varying degrees of likeness to speech, "clicks" and "bats", for example, also impair message transmission. Similarly, telegraph, program, tele-photograph, and other special services such as data transmission place additional requirements on systems. Examples are limits on "hits", or short duration interruptions, and limits on delay distortion.

A complete discussion of all these sources of impairment would be far too detailed for the purposes of this text. We shall limit ourselves to a discussion of items (a) to (f) above. These are among the most important for telephone message service, and a discussion of them will adequately illustrate the inter-relations which exist between the many types of impairments and their effects on transmission system design, operation and maintenance.

Our discussion will necessarily involve a number of perhaps unfamiliar concepts, measurement techniques and units of measurement.

It may be helpful to follow the rather diverse material to be covered if we begin with an outline:

1- Signals and noise can be measured at various points in a telephone system. The answers will depend upon where in a circuit a measurement is made. The first concept to be discussed is the establishment of a reference point in the system to which all measurements can be tied. It is dealt with in the section on "Level" (Page 2-3).

2- Voltmeters and ammeters are useful in measuring simple sinusoidal signals. In the telephone system, however, we must measure complex signals and noises. Simple instruments are inadequate, particularly since they do not take into account any of the subjective factors which determine the final evaluation of a telephone circuit. The special instruments and unique units of measurements in telephony are discussed in the sections on "Volume" (Page 2-4, which deals with signal magnitudes) and "Noise Measurement" (Page 2-6).

3- Some means for evaluating the overall performance of a telephone circuit is required. Two methods of rating are discussed - the Working Reference System, under "Effective Transmission Loss" (Page 2-11) and the Electro-acoustic Transmission Measuring System (Page 2-16). It is in the section on Effective Transmission Loss that such terms as Noise Transmission Impairment and Distortion Transmission Impairment are introduced and explained. At this point, we pause to take a look at the performance of present day circuits and the customer's opinion of them.

4- Noise objectives are discussed next (Page 2-20), and it is pointed out that such requirements must be stated in statistical terms rather than as a single number (Page 2-23).

5- A brief discussion of crosstalk objectives, in rather broad terms, rounds out this discussion of interferences (Page 2-25). (A more detailed section on crosstalk follows in the next chapter.)

6- Next, we consider talker echo and singing (Page 2-26). We find to our sorrow that loss must be inserted in contemporary telephone circuits to make echoes tolerable and to prevent singing.

7- Finally (Page 2-34) we consider the objectives which have been set for future connections in terms of received volumes and customer satisfaction.

The student is now presumably prepared for the rather varied package of important ideas presented in the following pages of this chapter.

Level

In order to specify the amplitudes of signals or interference, it is convenient to define them at some reference point in the system. The amplitudes at any other physical location can then be related to this reference point if we know the loss or gain (in db) between them. In the local plant, for example, it is customary to make measurements at the jacks of the outgoing trunk test panel, or (if one does not wish to include office effects) at the main frame. For a particular set of measurements, one of these points might be taken as a reference point, and signal or noise magnitudes at some other point in the plant predicted from a knowledge of the gains or losses involved.

In toll telephone practice, it is customary to define the toll transmitting switchboard as the reference point or "zero transmission level" point. To put this in the form of a definition:

> The transmission level at any point in a transmission system is the ratio of the power of a test signal at that point to the test signal power applied at some point in the system chosen as a reference point. This ratio is expressed in decibels. In toll systems, the transmitting toll switchboard is usually taken as the zero level or reference point.

Frequently the specification of transmission level is confused with some absolute measure of power at some point in a system. Let us make this perfectly clear. When we speak of a -9 db transmission level point (often abbreviated "the -9 level"), we simply mean that the signal power at such a point is 9 db below whatever signal power exists at the zero level point. The transmission level does not specify the absolute power in dbm* or in any other such power units. It is relative only. It should also be noted that although the reference power at the transmitting toll switchboard will be at an audio frequency, the corresponding signal power at any given point in a broad band carrier system may be at some

---

* dbm = power in db relative to one milliwatt.

carrier frequency. We can, nevertheless, measure or compute this signal power and specify its transmission level in accordance with the definition we have quoted. The transmission level at some particular point in a carrier system will often be a function of the carrier frequency associated with a particular channel.

Using this concept, the magnitude of a signal, a test tone, or an interference can be specified as having a given power at a designated level point. For example, in the past many long toll systems had 9 db loss from the transmitting to the receiving switchboard; in other words, the receiving switchboard was then commonly at a -9 db transmission level. Since noise measurements on toll telephone systems were usually made at the receiving switchboard, noise objectives were frequently given in terms of allowable noise at a -9 db transmission level. Modern practice calls for keeping loss from transmitting to receiving terminals as low as possible, as part of a general effort to improve message channel quality. As a result, the level at receiving switchboard, which will vary from circuit to circuit, may run as high as -4 or -6 db. Because of this, requirements are most conveniently given in terms of the interference which would be measured at zero level. If we know the transmission level at the receiving switchboard it is easy to translate this requirement into usable terms. If, say, some tone is found to be -20 dbm at zero level and we want to know what it would be at a receiving switchboard at -6 level, the answer is simply -20 -6 = -26 dbm.

Volume

A periodic current or voltage can be characterized by any of three related values: the rms, the peak or the average. The choice depends upon the particular problem for which the information is required. It is more difficult to deal with complex, non-periodic functions like speech in simple numeric terms. The nature of the speech (or program) signal is such that the average, rms and peak values, and the ratio of one to the other, are all irregular functions of time, so that one number cannot easily specify any of them. Regardless of the difficulty of the problem, the magnitude of the telephone signal must be measured and characterized in some fashion which will be useful in designing and operating systems which involve electronic equipment and transmission media of various kinds. We must be able to adjust signal magnitudes to avoid overload and distortion and we must be able to measure gain and loss. If none of the simple characterizations is adequate, a new one must be invented. The characteristic used is called "volume" and is expressed in VU (volume units). It is an empirical kind of measure evolved to meet a practical need and is not definable by any precise mathematical formula. The volume is simply the reading of an audio signal on a carefully specified volume indicator, called the VU-meter, when the meter is read in a carefully specified fashion.

The development of the VU-meter was a joint project of the Bell System and the two large broadcasting networks. They decided that the principal functions required of such a measuring device were:

1. Measuring signal magnitude in a manner which will enable us to avoid overload and distortion.
2. Checking transmission gain and loss for the complex signal.
3. Indicating the relative loudness with which the signal will be heard when converted to sound.

Since one of the principle functions is the detection of overload conditions, one might suspect that a peak reading instrument would be most desirable. A difficulty arises, however, when a peak reading instrument is used to compare signal magnitudes at various points in a long circuit. The effect on wave shape of large values of delay, where delay is a function of frequency, is drastic, particularly on sharp peaks. Thus two readings taken at widely separated points might indicate a loss or gain in the circuit when in fact there was only phase shift. For this reason the VU-meter is an rms-reading instrument, in effect integrating the signal over a short period, but a period long enough so that circuit phase shift will have negligible effect on readings. At the time the VU-meter was proposed, severe subjective tests were made to assure that it was a good indicator of overload, despite its inability to follow the sharpest peaks. The fact that an rms-reading meter can be used satisfactorily is a function of the physiological and psychological factors involved in the ear's appreciation of distortion. Considerable distortion can apparently be tolerated if it occurs in rare, short peaks.

While the rms meter used in the VU-meter is slower than a peak reading device, it does not read long term power. The meter follows a complex signal with a certain amount of sluggishness and, as the results of the overload detection test indicate, is a good indicator of the signal peaks which cause annoying overload. In actual use it is read with a special technique. One observes the peak swings of the meter, averaging the peaks mentally, ignoring occasional rare, high-valued peaks. The resulting "averaged" reading, taken in this unique fashion is the "volume" of the audio signal being observed. The value of the VU-meter as an indicator of relative loudness in the eventual acoustic signal derives from the fact that there is a statistical relationship between peaks, integrated peaks and the longer term average power in speech and program material. As a matter of fact, either a peak reading or an rms reading meter would be satisfactory in this respect. It is true that this statistical relationship depends upon the type of material. In practice, however, it is found that the VU-meter can be used equally well for all speech, whether male or female. There is some difference between music and speech in this respect, and so a different reading technique is used when using the VU-meter for measuring program material.

For convenience, the meter scale is logarithmic, with a 10 log scale. That is, readings bear the same relationship to one another as do decibels. The readings are <u>not</u> in decibels, however. They are in volume units -- VU's. This is a specially invented unit for a specially conceived concept, the invention having been made to meet a practical measurement problem. The VU can only be described as that unit of measurement read on a certain type of meter, built to special specifications and read in a specified manner. There is no other accurate way to describe it. It is true that the meter will give a reading if a continuous sinusoid is imposed upon it. It is also true that one can establish a correlation between the volume of a talker and his long term average power or his peak power. Such correlations are valuable, but the fact that they exist should not be allowed to confuse the real definition of volume and VU's. Putting it as simply as possible -- a -10 VU talker is one whose signal is read on a volume indicator (by someone who knows how) as -10 VU -- period!

One final note. The VU-meter has a flat frequency response over the audible range and is not frequency weighted in any fashion.

Noise Measurement

The measurement of noise, like the measurement of volume, is an effort to characterize complex signal. The noise measurement is further complicated by our interest, not in the absolute magnitude of the noise, but rather in how much it annoys the subscriber. Consider what is required of a meter which will measure the subjective effects of noise:

1 - The meter movement should have a time constant resembling that of the human ear.
2 - The readings should take into consideration the fact that the interfering effect of noise will be a function of frequency spectrum as well as of magnitude.
3 - When different types of noise cause equal interference, as determined in subjective tests, use of the meter should give equal readings.*
4 - When dissimilar noises are present simultaneously the meter should combine them to properly measure the overall interfering effect.
5 - The units of measurement should be meaningful and convenient.

The 2B Noise Meter, the present standard noise measuring device in the Bell System, does not meet all these requirements ideally. The problems of meter movement design and meter reading technique are straightforward and not unlike those of the VU-meter. The establishment of a relationship between a meter reading and a subjective effect however, is not so simple. The first task is to determine the effect of noise on articulation.** This can be done through extensive tests on human observers. Several questions need to be raised first, however. If we are designing the meter for use in the telephone plant, we must remember that, in general, the electro-acoustic transmission between any given measuring point and the subscriber's ear will be a function of frequency. Since we are interested in the effect on the subscriber, the meter must be designed to weight its measurements to take into account the frequency shaping of the external circuit. And if the meter is to be frequency weighted we must also decide where in the telephone plant we want to use it. We must also define the circuit external to the listener's ear, both for the purpose of designing the meter and for using it in the field.

Further, it is known that the ear does not have equal sensitivity to sounds at all frequencies. It is also not necessarily true that noises of equal power but of different types will have equal interfering effects. In making our subjective tests, what kind of noise should we use? If we design a weighting network for use with a particular

---

*Chapter 3 discusses some of the more common types of noise.

**Articulation tests do not give the complete answer, of course. Some types of noise, although annoying, do not affect intelligibility much.

external circuit using one kind of noise, will the measurement be meaningfull for other types of noise?

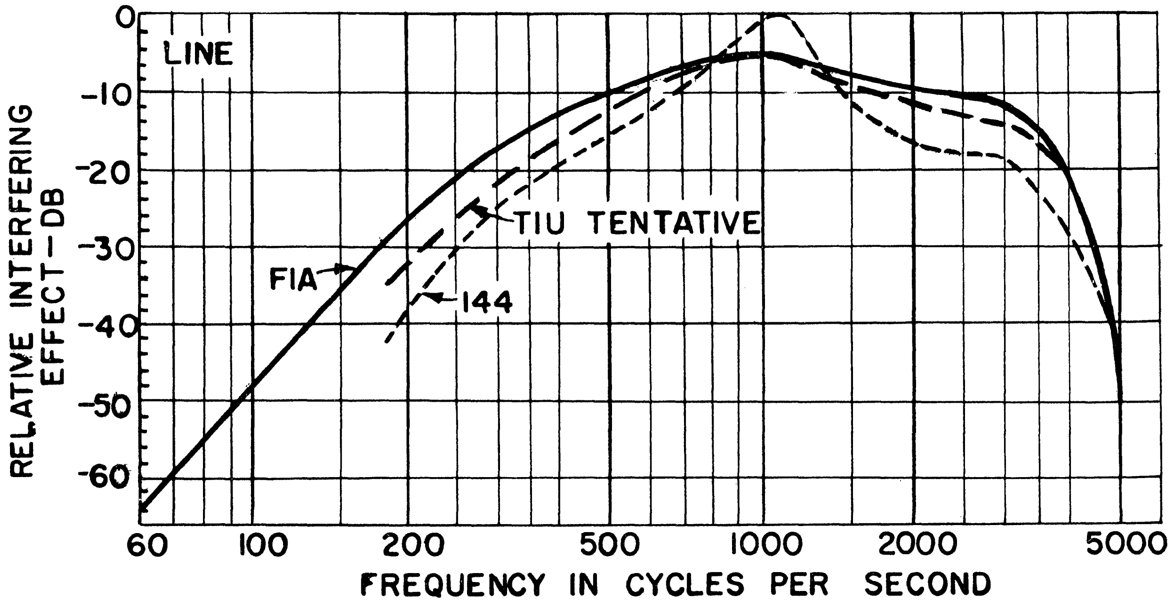
The question of defining external circuits for the subjective tests in answered in terms of convenience of use of the meter. One convenient measurement is made across the receiver terminals in the subset. The frequency weighting associated with this measurement is designated RECEIVER weighting and must, of course, be further associated with a particular design of receiver. The other convenient measurement is made across the line at a central office. That weighting is designated LINE weighting and presumes a specified subscriber loop, central office loss and a specified type of subset across the line, as well as the transmission from receiver to human brain. The weighting networks, in other words, must take into account not only the subjective weighting of noise at different frequencies, but also the weighting associated with the transmission of the test circuit. Since the noise meter is tied by design to specified points in the plant it can only be used at those points - accross the receiver terminals or across the line at the receiving central office. If it is used elsewhere, corrections, assuming they are known, will have to be made.

Once the "standard" test circuits are defined, the subjective tests themselves can be made. In practice what was done was to determine the interfering effects of single frequency test tones on speech. The resulting preliminary weighting was used to make measurements of other types and magnitudes of noise which were subjectively judged to be equally interfering. Ideally, the meter should yield equal answers. It did not, so the weighting curves were slightly modified to minimize the errors (and to take into account limitations in the design of physical weighting networks). The actual LINE weightings are shown in Figure 1.*

---

*For the sake of the student who may be unfamiliar with the various code designation for telephone receivers and transmitters the three types discussed in this chapter are identified as follows:

Subset Type	Designation of LINE Weighting	Receiver Type	Transmitter Type
500	T1U	U-1	T-1
302	F1A	HA-1	F-1
Desk stand type used in Working Reference System	144	144	337



Noise Meter Line Weightings
(Relative Interfering Effects of Telephone Line Voltages or Currents)

Figure 2-1

The 2B Noise Meter does not include LINE or RECEIVER weighting networks for the type 500 subset. The F1A and tentative T1U LINE weighting are sufficiently close, however, to permit LINE measurements on the type 500 set with acceptable error. This is not true of the RECEIVER weightings, however, and the 2B noise meter <u>cannot</u> be used for measurements across the U1 receiver of the 500-type subset.

Now, assuming that we have a meter which takes into consideration the electrical and acoustical transmission of the noise from the measuring point to the subscriber's ear as well as the psychological effects (frequency weighting and time constant), we can consider the problem of adding noises. At the time this measuring technique was developed it was believed that interference effects added according to a power law (10 log). That is, it was thought that if two tones had equal interfering effect when applied individually, then the effect if both were present would be 3 db worse than for each separately. Recent research indicates that the combination of subjective factors in hearing may, in general, follow much more complex laws in which the addition law is a function of the relative frequencies and magnitudes of the signals being combined. Nevertheless, the present meter assumes power addition. This will appear less distressing if it is kept in mind that system objectives are set, not on the basis of arbitrary measurements with a noise meter, but in terms of subscriber reactions.

Finally, we need to consider a rational unit of measurement. The meter reads noise on a decibel scale. Originally these readings were relative to a certain amount of circuit noise, designated "reference noise" and the units known as dbRN. This designation is no longer used.* The new reference, <u>when the meter is used with 144 weightings</u>, is -90 dbm at 1000 cps. The unit has also acquired a new name, dba (decibels, adjusted), for reasons which we will now explain.

With the introduction of the type 302 subset, it became necessary to provide new weighting networks for the 2B Noise Meter. Figure 1 shows the relative shapes of the old and new weightings. Unfortunately, a network design difficulty resulted in the F1A and HA1 networks introducing 7 db of flat loss which has nothing to do with the interfering effects of noise. For this reason it is necessary, when using type 302 weightings, to correct the meter readings by +7 db before the actual measurement is known. Thus, a noise which should measure 0 dba with F1A or HA1 weighting will actually give a meter scale reading of -7 db. Adding the correction of +7 db gives the actual measurement of 0 dba (db adjusted). For all weightings, various other corrections can be made when the noise meter is to be used at points other than the receiving switchboard or the receiver terminals or across non-standard impedances.

Because the HA1 receiver is 5 db less efficient than the 144 type at 1000 cps, it takes -85 dbm of 1000 cps tone to cause the modern subscriber the same annoyance which -90 dbm would have caused a subscriber equipped with a 144-type receiver. Very properly, therefore, a -85 dbm 1000 cps tone will give rise to a 0 dba noise reading with F1A or HA1 weighting.

It also happens, as a matter of fortuitous accident, that white noise - random noise which is flat with frequency - applied at the receiving switchboard across a 3000 cps band will give the same dba reading for either 144 or F1A LINE weighting. In other words, the effect of white noise at the receiving switchboard is the same for both subsets. The student is free to satisfy himself that the arithmetical areas under the F1A and 144 weighting curves of Figure 1 are equal, which is what is

---

*See Appendix I to this chapter.

implied above. The magic number to remember is that 0 dbm of white noise over a 3000 cps band will read +82 dba for either LINE weighting.

The following table summarizes the important numbers to remember in association with the 2B Noise Meter:

Weighting	Noise Source	Noise Power (dbm)	Meter Reading (db)	Correction (db)	Noise Measurement (dba)
144 LINE or RCVR	1000 cps tone	-90	0	0	0
F1A LINE or RCVR	1000 cps tone	-85	-7	+7	0
144 LINE	White Noise (3kc)	0	+82	0	+82
F1A LINE	White Noise (3kc)	0	+75	+7	+82

One precaution should be noted. The fact that -90 dbm of 1000 cps tone gives rise to a 0 dba reading does not mean that 0 dba <u>equals</u> -90 dbm. The noise meter measures the interfering effect of noise, but it cannot tell the observer what kind of noise is causing the interference or how much of a particular kind of noise is present. 0 dba <u>may</u> equal -90 dbm of 1000 cps tone or it <u>may</u> equal -82 dbm of white noise in a 3000 band or it <u>may</u> equal a certain magnitude of tone at any given frequency, but the only thing these three noises will necessarily have in common is that they cause the same amount of interference.*

<u>Effective Transmission Loss</u>

With our definitions of level, volume and dba in mind, we can now begin to consider objectives for the first three of the important categories of imperfections mentioned at the beginning of this chapter -- the impairments caused by loss, noise, and bandwidth restrictions as measured on a subjective basis. We lump these together because of historical reasons -- they have, in the past, all been evaluated in terms of "effective transmission loss".

"Effective transmission loss", often abbreviated to "effective loss", is a composite "figure of merit" for a telephone message channel, which is arrived at by lumping together the impairments

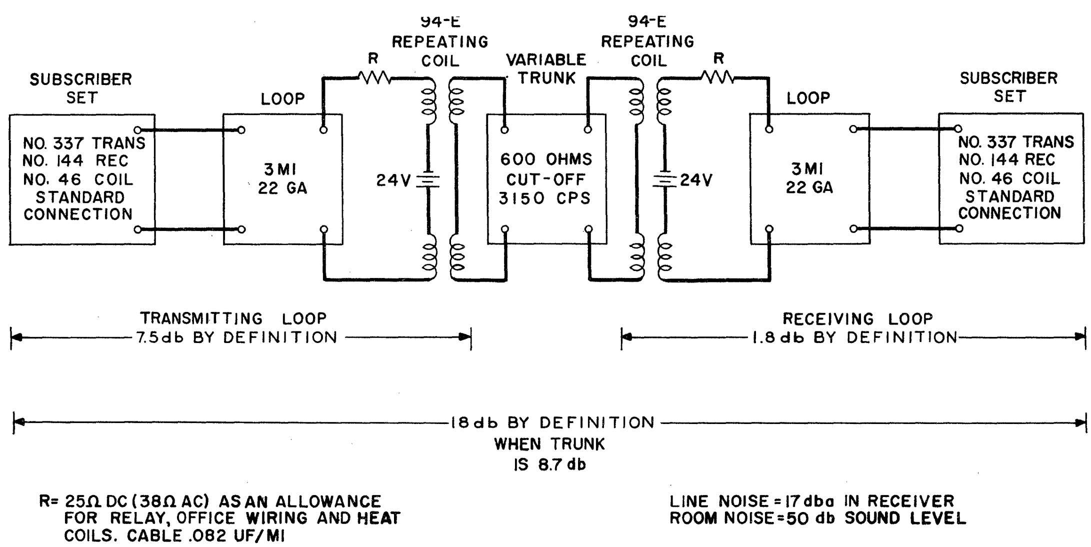
- -

* See Appendix II at the end of this chapter.

caused by loss, noise, and bandwidth restriction.* Noise, for example, can be expressed in terms of loss by setting up two channels of good quality and then degrading both - adding a fixed amount of noise in one, and varying an attenuator in the other. We might find, for example, that the impairment caused by +31 dba of noise, measured across the subscriber's receiver, is (by his subjective judgment) equal to the impairment caused by increasing the attenuation in the comparison circuit by 8 db. Similarly we might find that there was little to choose between limiting the band to 2000 cycles or adding 6.5 db of loss.

The evaluation of transmission impairments is another illustration of the sort of subjective testing which underlies almost all transmission objectives. It might be remarked, in passing, that getting the needed answers from a subjective test calls for great care in setting up the experiment. The values of NTI (Noise Transmission Impairment) obtained, for example, will be affected by room noise, circuit bandwidth, and a number of other factors. To ignore any important effect is, in essence, to ask the wrong question, and there are many examples of subjective tests which have done this.

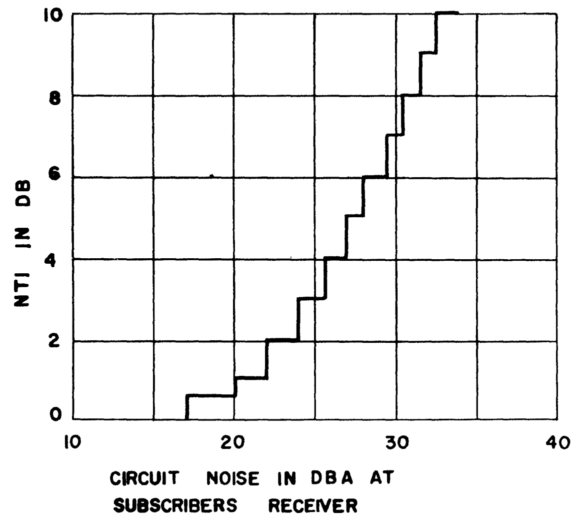
Figure 2 shows the Working Reference System (WRS) which has been used in the past for evaluating subjective transmission impairments. It consists of standard transmitting and receiving loops



Working Reference System for the Rating of Telephone Circuits

Figure 2-2

*Sidetone is also an important factor and will be discussed in the next chapter.



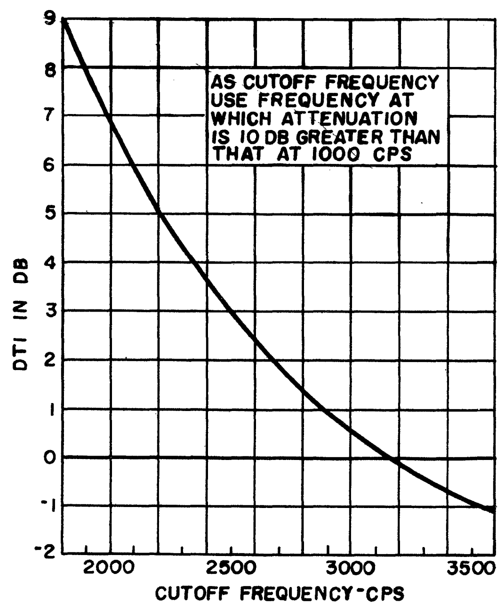
Noise Transmission Impairments (NTI)

Figure 2-3

connected by a variable trunk that is distortionless except for a 3000 cycle low pass filter. At the time it was specified it represented a circuit with characteristics and performance comparable to a good commercial circuit. It was arbitrarily defined as giving 18 db of effective loss in its normal configuration; this value is, in turn, related to evaluations obtained with a still earlier test circuit. To completely specify the circuit, room and line noise as well as circuit loss must be stated.

In order to obtain a measure of the effect of noise on the quality of a circuit, a subjective test was carried out using two working reference circuits. An arbitrary amount of additional noise was introduced into one, and additional loss was then introduced into the other (while still maintaining the 17 dba at the receiver input of the latter, which corresponded to a good local circuit at the time), until the two circuits were judged to be of equal quality. The amount of added loss was the NTI (Noise Transmission Impairment) corresponding to the noisier circuit and measured the amount that the increased noise was judged worse than 17 dba. The results obtained from a series of such tests are shown in Figure 3. Note that the results do not mean that 17 dba of noise into the receiver has no annoying effect. As a matter of fact, previous tests indicated that a circuit with negligible noise would have to be degraded by adding about 3 db of loss to make it equal in quality to the 17 dba normal loss WRS circuit.

Figure 4 gives the effect of limiting bandwidth, again in terms of the added circuit loss (defined as "Distortion Transmission Impairment", or DTI) which would give the same transmission impairment.



Distortion Transmission Impairments (DTI)

Figure 2-4

While the values of NTI and DTI in the above diagrams are useful, caution must be exercised in applying and interpreting them.

1. The addition of NTI, DTI and electrical loss as a means of finding an overall rating of a circuit is seldom valid. The direct addition of subjective effects may provide a rough means for comparing similar circuits under similar conditions where the fact that a distorted scale is being used is not of great importance. (The distorted scale arises from assuming that subjective factors can be added on a db-basis, whereas actually they cannot be so added.
2. The method of obtaining NTI and DTI is to equate these impairments with changes in volume or magnitude of signal. This is valid only over a limited region. Once the volume has reached a reasonable magnitude there is no subjective advantage in increasing it further. For example, high volume coming over a 2 kc channel is not as pleasing to a subscriber as a good normal volume coming over a 4 kc channel.
3. The NTI and DTI were determined on a circuit which is not representative of modern circuits and it does not follow that a given amount of noise will be equally disturbing on the old subset of the WRS and on a modern circuit using type 302 or type 500 subsets.

If the student has now concluded that the Working Reference System is not entirely satisfactory, he is correct. The subject is not merely of historical interest, however. It provides a foundation for understanding the development of a new method of rating telephone circuits.

The WRS was originally established to make possible evaluations of circuits similar to the standard one, with its 144 receiver, 337 transmitter and three mile subscriber loops. When the type 302 subset appeared, an effort was made to evaluate its effective loss, substituting one mile subscriber loops for the three mile loops considered standard when the WRS was first evolved. When we measure the transmitting loop of this newer circuit we find its effective loss to be -11 db. Similarly, the effective loss of the newer receiving loop is measured to be -7 db.*

Using these measurements of effective loss of the prescribed type 302 circuit, it is possible to derive a rule of thumb which will give a good estimate of the total effective loss of any type 302 circuit, provided the electrical loss between loops is known. The total effective loss of a circuit is, of course, defined by an entirely subjective measurement. In the absence of noise and bandwidth impairments it can be thought of as consisting of three factors: effective transmitting loop loss, effective receiving loop loss and the electrical loss of the trunk. In deriving our rule of thumb, the electrical loss is characterized by its value at 1000 cps rather than by its overall subjective effect. This makes it possible to combine an objective measurement of electrical loss with the known subjective effect of the subsets and standard loops. Writing this as an equation:

Effective loss = effective transmitting loop loss
+ effective receiving loop loss
+ electrical loss of trunk at 1 kc

For the values stated above for the 302 subset with standard loops, the resulting rule of thumb is that the effective loss equals the

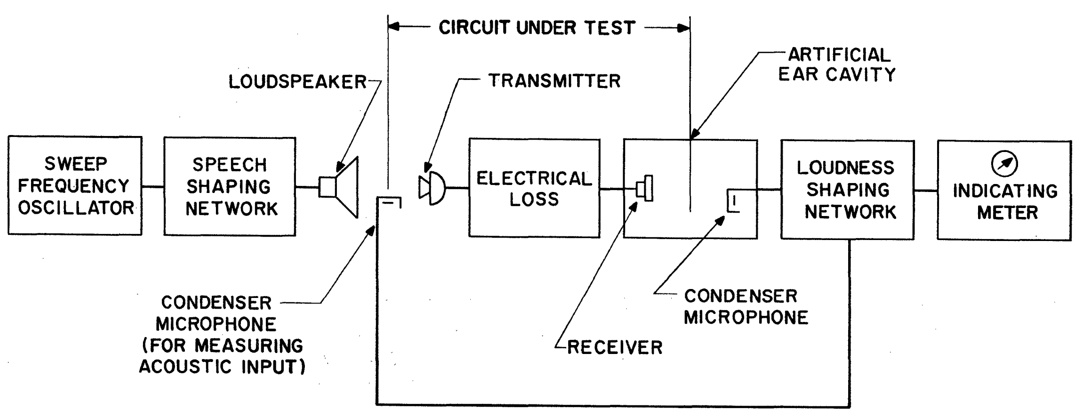
---

*The minus value for effective loss need not be confusing. It arises from the assumption of an arbitrary standard for the reference circuit. In the WRS, the transmitting loop has an effective loss of 7.5 db by definition. The figure of -11 db for the type 302 subset and its loop simply says that it is 18.5 db better than the old circuit. Note that this 18.5 db figure has nothing to do with the 18 db of effective loss arbitrarily assigned to the WRS. The same can be said of the unfortunate coincidence whereby the sum of effective loop losses for the type 302 circuit happens to equal -18 db. These are unhappy tricks of nature, probably deliberately introduced to confuse the student.

1 kc electrical loss minus 18 db. The approximation is sometimes carried one step further and it is stated that the total effective loss of a circuit such as this is equal to the 1 kc electrical loss minus 18 db plus the NTI and DTI. We have already suggested that this sort of addition has limited value. No similar conversion factor has been established for evaluating circuits using the type 500 subset. An entirely new test set for circuit evaluation is now being introduced in the Bell System.

Electro-Acoustic Transmission Measuring System (EATMS)

This new circuit evaluation test set is the EATMS. Our discussion of the Working Reference System indicated the difficulties inherent in the subjective evaluation of circuits as well as the obsolescence of the reference circuit. In recent years noise and distortion impairments in the toll plant have been reduced to very low values. Marked progress has been made in the same direction in the local plant. Our principal interest now is in the impairment caused by transmission inefficiency, which results in low received volumes. This amounts to saying that our attention is currently focused on the one impairment best suited to objective measurement. Why not eliminate the difficulties of subjective measurements, for this purpose at least, and develop a method of testing adapted only to the measurement of efficiency of instruments and of overall connections? At the same time we can define the results in more reasonable units, eliminating measurements of negative "effective loss" for circuits which actually have appreciable real loss.



Simplified Block Diagram of EATMS

Figure 2-5

The EATMS is shown in simplified block diagram in Figure 5. The circuit which it tests consists of a transmitter and its associated subset circuitry, an electrical circuit consisting of loops, trunks or other sources of electrical loss, and a receiving subset circuit and receiver. The test is made by feeding a known acoustic signal to the transmitter under test and measuring the acoustic output from the receiver.

The input to the circuit under test is generated by an oscillator which sweeps through the telephone voice frequency range six times per second. The oscillator output is fed to a loudspeaker, which physically simulates the human mouth, through a weighting network which shapes the acoustic energy spectrum to match that of typical male English speech. The loudspeaker output is measured, in the absence of the circuit under test, with a pressure-sensitive microphone. This permits us to include the acoustic effect of the physical position of the transmitter. The signal passes through the electrical connection under test to the receiver, which is mounted in a cavity simulating the acoustic loading effect of the human ear. Also mounted in the cavity is a microphone which measures the acoustic output.

The signals detected by the two microphones, one measuring acoustic input and the other the acoustic output are fed through a second weighting network which relates the output of the pressure-sensitive microphones to the loudness effect experienced by the human ear. The result is expressed in decibels, derived from taking 20 log of the ratio of the measured output and input pressures.

Obviously, in a telephone circuit combining acoustic and electrical components, we can measure voltages as well as pressures. We can thus define gains and losses between acoustic input and electrical output or between electrical input and acoustic output, at intermediate points in the connection. To express such mixed measurements of gain and loss we must define a convenient unit. The unit chosen is expressed thus:

$$\text{Transducer gain (in db)} = 20 \log \frac{\text{volts}}{\text{millibars}} \; *$$

If we are consistent in using this definition throughout, it is possible to add the transducer gains and losses and the electrical losses in db and come up with the same answer that would obtain from an overall measurement. Again we are violating the purists' insistence upon the db as a unit for expressing power ratios, but 20 log of a ratio is a convenient

---

* One millibar = 10^3 dynes/cm^2

unit to use and the relationship between the numeric and the unit we want to use is identical with what is found in a db table based on current or voltage ratios. For our purpose it suffices to call the units "db".

The purpose of the EATMS, then, is an objective measurement of that characteristic of a telephone connection in which we are most interested -- transmission efficiency. The set can be made portable, does not require subjective judgments and is therefore convenient for work in the field as well as in the laboratory. It falls short of the ideal in two respects, however.

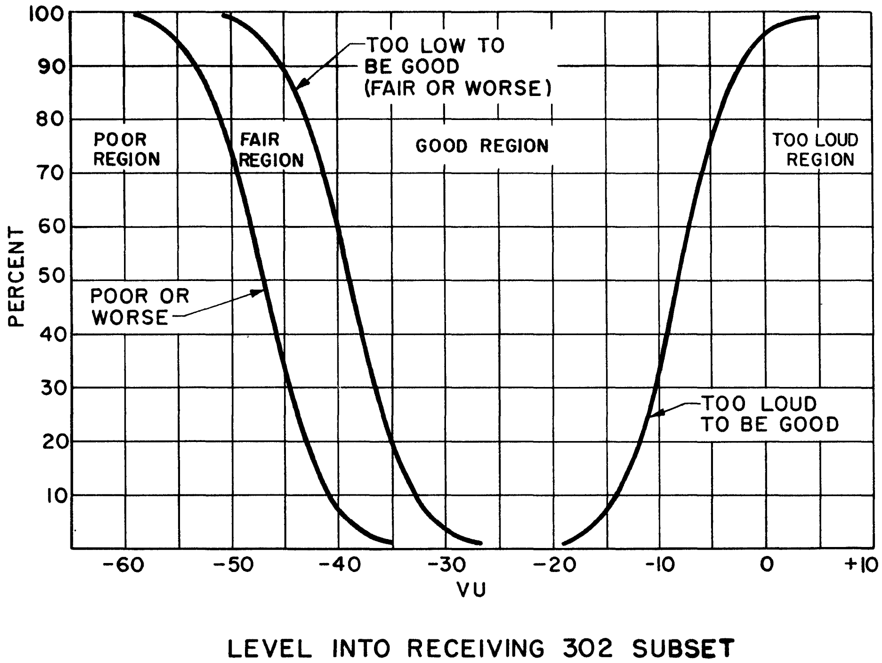
First it is not a complete substitute for the WRS and in those cases where noise and bandwidth impairments cannot be neglected some sort of correction will have to be made. This will not commonly be the case, however. If, when transmission efficiency throughout the System has been improved, subscriber subjective standards rise and once more require an emphasis on rating noise and bandwidth impairments, it is probable that a new approach and a new technique will be required for evaluation. The fact that systems objectives are based on the subjective reactions of customers and that the customers' requirements change as circuit characteristics change is nothing new in the telephone business.

The second shortcoming of the EATMS may prove to be more serious; this is the fact that it does not "judge loudness" in quite the same way as the human ear. The basic problem here is the same one that we noted before in discussing the noise meter. For many circuits, the EATMS error is small, but in some cases may not be negligible. Current development work may result in modifying the detector characteristics to eliminate this difficulty, but at the moment a satisfactory solution has not yet been found.

Customer Reaction to Received Volume

Thus far we have discussed some of the methods for evaluating circuit performance, particularly with respect to loss, noise, and bandwidth. Let us now consider the objectives to which we should engineer.

If, as the previous section suggests, noise and distortion impairment are held to negligible values, then a customer's evaluation of a connection will depend upon the received volume alone. Figure 6 shows the statistical relations between received volume and circuit quality based on the opinions of a large number of observers. To see the kind of information that can be obtained from this figure let us take a couple of examples. Suppose that the received volume into a certain 302 subset is -15 vu. From the curve can be read that 6% of the subscribers would judge



Judgment of Grade of Telephone Conversation vs Received Volume

Figure 2-6

it to be too loud and 94% would judge it good. A received volume of -40 vu would be rated good by 42% (100 minus 58 who think it worse than good) and 6% would rate it poor or worse. The difference between 58% fair or worse and 6% poor or worse is 52%, those who would rate it fair.

From these curves of observers' opinions of specific received volumes, an estimate may be obtained of subscribers' reactions to any distribution of actual received volumes that might be given to them. This is done by properly combining the opinion distributions with the received volume distribution. Such an estimate forms the basis for a business judgment on the grade of service that is to be supplied. A completely rational analysis of the problem would have to include data on the cost of providing the various distributions of received volumes, that is, grades of service, which are contemplated, as well as the dollar value of the customers' satisfaction or dissatisfaction as a result of this grade of service. Even the more tangible of these, the costs, are difficult to obtain, since performance-cost relations are continually changing as new developments come into the plant. As a practical matter the decision rests on a balance of these costs against what is judged to be the value of customers' reactions to the grade of service under consideration.

Estimates of subscribers' reaction to grades of service on toll connections in the present telephone plant, and in the presently projected future plant, will be discussed in a later section of this chapter.

Noise Objectives

Both the local and the toll plant will contribute to the noise heard by the subscriber on many of his calls. Local plant noise is a subject which calls for a more extended discussion than would be appropriate here, depending as it does on the type of office and line involved; it is considered in more detail in the chapter on voice frequency systems which follows. In general, however, it might be said that when possible, local noise is held to rather low values, of the order of 20 dba at the central office. This corresponds to the 12 dba at the subscriber's receiver and is well below the 17 dba value previously found satisfactory and assigned a value of 0 NTI.* These values should be considered as very tentative, however, since no national surveys have been made for many years, nor is there any agreed-on standard to which all operating companies conform. Instead, high noise situations, when observed, are dealt with on an individual basis.

In the toll plant, noise is more subject to specification and control. From the beginning, it has been the practice to design broadband carrier systems so that the transmission impairment due to noise and bandwidth restrictions would be essentially negligible, and the loss as low as possible. There are sound economic reasons for so doing. To meet any particular objective for the effective transmission loss on toll calls, the permissible transmission impairments must be allocated partly to the local and partly to the long-distance plant. Meeting the resulting requirements will call for both initial investment and maintenance expense. It is obviously cheaper to purchase a db of performance, in terms of reduced transmission impairment, in the relatively few broadband systems than in all the millions of subsets and local circuits which they may, from time to time, be called upon to interconnect.

Another reason for making the toll plant good should always be at the back of our minds. As engineers, our essential job is planning for the future. While we cannot know what future requirements will be, a reasonable extrapolation from the past tells us (a) that telephone subscribers will become more, not less, critical, and (b) that new services and new types of signals will impose still more stringent

---

*Assuming no noise is picked up in the loop, noise at receiver will be 8 db less than at the office. This 8 db figure assumes 0.5 db office loss, 4.5 db loop loss, and 3 db loss in the subset circuitry.

requirements. The mere possibility of future needs may not always justify the expenditure of today's dollars, but one of our jobs, nevertheless, is to take every opportunity of building flexibility into the plant.

In discussing toll noise objectives, it is natural to begin with system design objectives. Good maintenance will be required if these values are to be actually realized in the field, but the best maintenance can hardly be expected to get better performance from a system than was designed into it.

The fundamental design objective might be stated thus: the subscriber should, on the average, receive no more than +17 dba of noise at his receiver. It will be recalled that this corresponds to zero db NTI. Working back through the subset and loop to the receiving central office, and ignoring any other sources of noise, we find that 17 dba at the receiver corresponds to about +26 dba at the local central office. (This assumes a loop having about 1 db more than average loss; noise is more important when received volumes are low.) If all this noise comes from the toll circuits through a terminating link (toll connecting trunk) which has about 3 db loss, the noise at the receiving toll office must be 29 dba.

As mentioned earlier, a common transmission level at receiving toll offices many years ago was -9 db. For many years, therefore, the design objective for transmission systems was stated as 29 dba (average noise power) at the -9 level point at the receiving toll office at the end of the longest circuit, which was taken as 4000 miles. If we postulate a fictitious zero level point for convenience of measurement or computation (at the receiving toll office) the corresponding value at such a point would be +38 dba.*

Stated in these terms, this noise objective is mentioned at many points in this text, for the simple reason that it is still being used in

---

*In specifying our objectives, we find it desirable to express them in terms of the measurements which would be made to determine whether or not a circuit was satisfactory. For this reason, noise requirements are given in terms of an absolute noise measurement, not in terms of signal-to-noise ratio. In some special systems like Pulse Code Modulation (PCM), noise performance is specified on a different basis. PCM is subject to a unique type of noise (quantizing noise) which is present only during signal bursts and which is a function of signal magnitude. Under such conditions, a system is engineered so that the noise objective for loud talkers is higher than that for quiet talkers. More will be said of this in the chapters on PCM which follow later in this text.

the design of long-haul transmission systems. It is, however, subject to re-evaluation, and will probably be revised in the near future.

For example, one result of the current drive to reduce loss in loops and in toll connecting trunks which separate the customer from the receiving toll office will be an increase in the noise power actually delivered to the subscriber's ear, assuming toll system noise remains the same at the receiving toll office. The loss from receiving toll office to subscriber's receiver quoted above totals 12 db; if this is, on the average, reduced to about 10 db, then if noise at the receiver is not to exceed +17 dba, we can have not +29 dba but +27 dba at the receiving toll office. If, furthermore, receiving toll offices are not -9 db level points but -4 db level points on the average, then the "zero level point" noise can be only 4 db higher than +27 dba, or +31 dba. This would be 7 db more stringent than the old requirement,* and would probably be too expensive to meet; a compromise solution might be an average of 34 dba at 0 TLP at the output of a 4000 mile toll circuit. This would, under the assumptions made above, be +20 dba at the subscriber's receiver, or about 1 db of NTI. Shorter lengths of long-haul transmission systems would give less noise than the relatively rare 4000 mile case taken as a reference in this discussion.

A revision of design objectives for noise will, of course, have no effect on systems which are already in the field, which were designed to the 29 dba at -9 TLP value (or 38 dba at 0 TLP). In most cases, however, margins were designed into such systems to allow for aging, etc.- which means that performance somewhat better than +38 dba at 0 TLP can be obtained by better maintenance.

The point being made here is that systems objectives are not static, that they are established in a certain transmission environment, and that they must be modified to remain useful in a constantly improving plant serving customers whose requirements are constantly growing more stringent.

In modern carrier plant, with its channel spacing of 4 kc, DTI has been reduced to negligible values. After allowing for adequate filter guardbands, this channel spacing leaves a useful bandwidth extending from 200 cps to 3200 cps. From Figure 4, it can be seen that there is not too much to be gained by extending this bandwidth appreciably.

---

*The increase in the number of 500 sets in use - with their increased sensitivity - will provide pressure to tighten the noise requirement even further.

Statistical Approach to Noise Objectives

So far, in stating the 38 dba noise objective, no mention has been made of the percentage of the time a given channel should meet this objective, or the percentage of the channels that should meet it at a given time. Pretty obviously, all the channels in a large complex telephone plant will not have the same amount of interference all the time. For example, random noise in radio systems will have reasonably low values most of the time, but higher values during deep fades. Carrier systems with less than 100 channels will have high peaks of inter-channel modulation noise a small percentage of the time. These case have generally been handled by permitting the maximum noise from any one source to reach a value not exceeding 49 dba at the zero TL for not more than .01% of the time in any hour.*

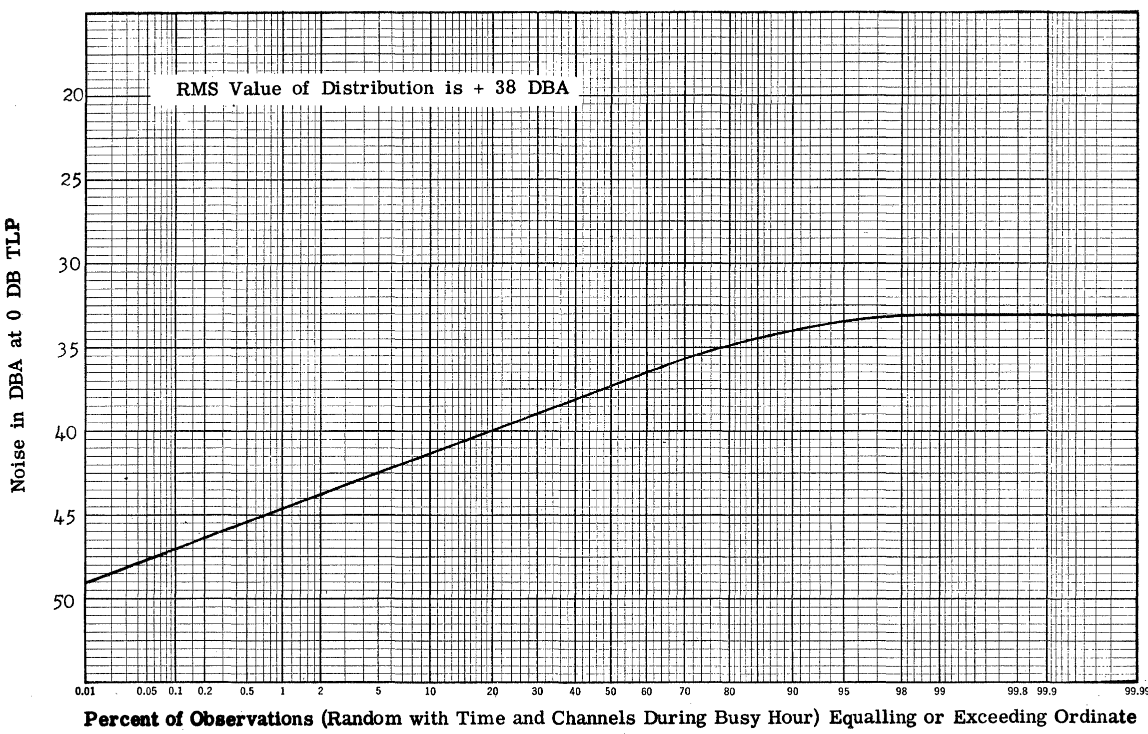
There is, however, no general agreement as to what constitutes an acceptable overall distribution of noise values with time or from circuit to circuit. One method of specifying an acceptable time distribution of noise which has been advanced results in Figure 7.

This curve is based on the assumption that any time distribution of noise is acceptable provided the average zero TLP value (averaged on a power basis) does not exceed +38 dba and the 0.01% value dos not exceed +49 dba. It will be noted that in this plot another condition has been imposed, namely that no 4000 mile circuit should be quieter than 33 dba at 0 TL. This is desirable because noise helps to mask crosstalk and because there is no advantage to reducing the noise below 33 dba at zero transmission level. Such distribution curves should be interpreted as expressing the chance that a subscriber may encounter the specified noise on a random call, and are thus distributions of both circuits and time.

The distribution of noises from circuit to circuit is, however, more complex to consider than time distributions are, and no method has been evolved for stating noise objectives for large multi-channel systems to take advantage of the fact that a few noisy channels might be tolerated. In general, long-haul systems such as coaxial or TD-2 have been designed so that 100% of the channels should be 38 dba or less at 0 TLP after traversing 4000 miles of the system. On the other hand, some advantage of statistics is taken in engineering short systems. Objectives set in recent years for short-haul systems such as the N, O, and TJ have been

- -

*Another special case, mentioned earlier, is that of persistent single-frequency tones. A requirement of not more than 15 dba at the receiving toll office for a 4000 mile system has been used as a design objective for these contributors to the total noise.



An Acceptable Distribution Curve For Busy Hour Noise at Receiving Toll Switchboard (0 DB TLP) At the End of a 4000 Mile System

Figure 2-7

on the basis that a 150-200 mile link would give +32 dba at zero TLP. A circuit which included four such links would give somewhat more than +38 dba at zero TLP.* In effect, this means that some long multi-link switched connections will fail to meet the 38 dba objective. This can be justified by the argument that the subscribers chance of encountering such a circuit is small, so that the net effect on grade of service in terms of transmission impairment of the entire plant will be small.

---

*Off hand, we might expect that if one link gives +32 dba, four would give +38 dba. Actually, somewhat more noise than this will be found at our fictitious zero level point because signal levels would be somewhat lower in each successive link, as indicated by the discussion of Via Net Loss in a later section.

It should be clear that the 38 dba noise objective is not a rigidly established number that must be adhered to without question for all circuits at all times. There are many instances in setting initial objectives on a system or in seeking a suitable compromise on some detailed design problem where some objective should be questioned and subjected to further study. However, any major compromise of this particular objective should be approached with great care and most likely be avoided.

Intelligible Crosstalk Objectives

Crosstalk is a term widely used to describe unwanted coupling from one signal path onto another. Crosstalk may be due to direct inductive or capacitive coupling between conductors, as discussed in the next chapter. It may also be caused by coupling between radio antennas, or by cross-modulation between channels and single-frequency signals (carriers or pilots) in multi-channel carrier systems. Such cross-modulation may occur in any non-linear element, such as the repeater electron tubes or terminal modulators. In many instances, the resulting interference in carrier systems will be unintelligible due to the interfering signal's having been inverted, displaced in frequency or otherwise distorted. In these cases it is generally grouped with other noise-type interferences.

When coupling paths give rise to intelligible (or nearly intelligible) interference, it is necessary to design the cable, open wire line, antenna, repeater or modulator so that the probability that a customer will hear a "foreign" conversation will be less than a prescribed value. In normal practice, a one per cent chance is considered tolerable. This value is not based on customer perference testing, but on an arbitrary judgment.

The probability of a listener's hearing a conversation from a disturbing channel depends on the distributions of a number of parameters. The probability of someone listening on the disturbed channel, the probability of there being signals on the disturbing channel, the acuity of the listener, the magnitude of the signal on the disturbing channel, the magnitude of masking noise, and the loss in the coupling path between disturbed and disturbing channels all enter into the final "probability of intelligible crosstalk".

In designing a transmission system, the problem usually is to adjust the coupling path loss so that the desired probability objective is met. The other factors in the problem depend on the people who use the channel, traffic patterns, and operating practices. In a study made some time ago, listener acuity, speech volumes, channel activities and other parameters were combined and the probability of intelligible crosstalk was derived in terms of masking noise, number of possible sources, and loss in the coupling path. The results of this study are given in the next chapter after further discussion of the cross-talk problem and the methods of measuring coupling paths.

Echoes and Singing

As we have noted, the effective transmission losses caused by noise and bandwidth limitations are currently held to sufficiently low values so that electrical loss has become the most important source of the transmission impairment caused by local and toll trunks. Offhand, it would appear that the electrical loss could easily be made negligible in repeatered trunks, but, in fact, two factors inhibit us. These are the hazard of singing and the penalties associated with echo which are characteristic of the existing telephone plant.*

Although these problems are particularly severe in repeatered two wire voice frequency transmission systems, as discussed in the next chapter, it is simpler to consider the four-wire toll case first. In four-wire toll systems, echoes and near-sing conditions arise because of impedance mismatches, particularly at the junction of the distant four-wire terminating set with the two-wire toll connecting trunk. The return loss** at this point is, on the average, only about 11 db over that portion of the voice frequency spectrum which is most important from the echo standpoint; i.e., 500 to 2500 cycles. (By expenditure of considerable effort, an average return loss of 14 db can be maintained but 11 db is the more realistic value.) This impedance mismatch causes a

---

*Singing may also arise from other causes. We are concerned here only with the difficulties which result from insufficient loss in a circuit which can include excessive noise and crosstalk as well as echo and singing.

**Return loss, in more familiar terms, is defined as,

$$20 \log \frac{1}{\text{Reflection Coefficient}}$$

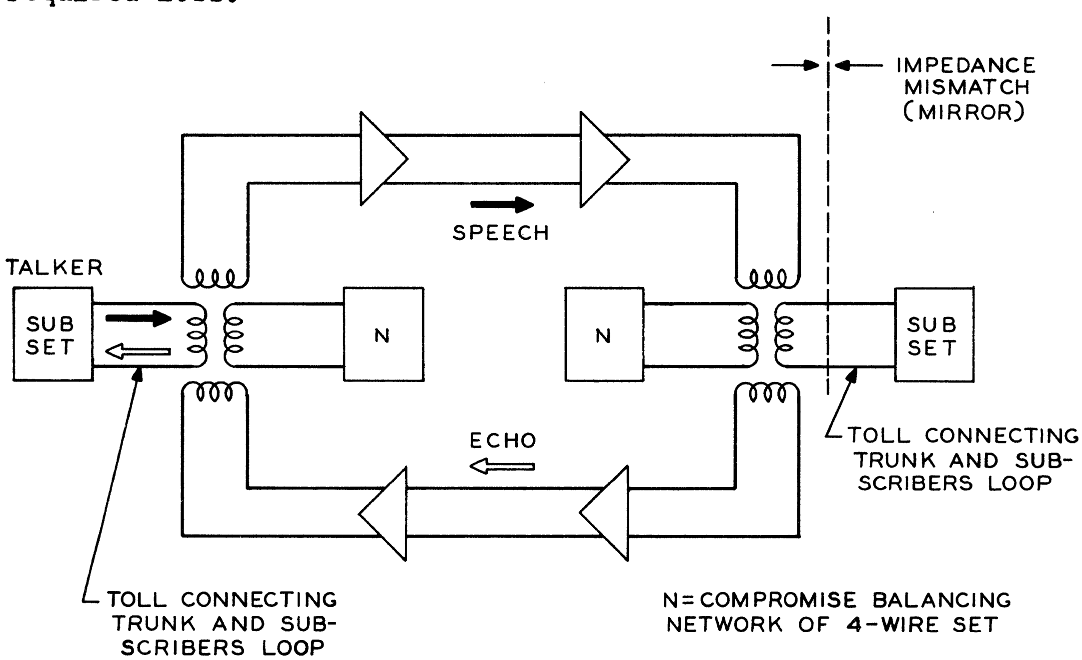
in which

$$\text{Reflection Coefficient} = \frac{\text{Reflected signal}}{\text{Incident signal}}$$

at the point of impedance discontinuity.

reflection, or echo, of a talker's voice to be returned to him on the channel to which he is listening. Figure 8 illustrates such an echo path. If the echo is not delayed (i.e., if the trunk is short enough) it is indistinguishable from sidetone, and is not annoying. If, however, it is sufficiently delayed in making the round trip, it can be quite annoying to a talker and can interfere with his normal process of speech.

At frequencies outside the 500 to 2500 cycle band, the balance of the four-wire set or hybrid may become much poorer than the 14 db figure quoted above, which can lead to a sing or near-sing condition, especially under special termination conditions - for example, when an operator's set is connected to the line. It is not sufficient to prevent singing itself - it is necessary, in a customer connection, to have enough margin to avoid the hollow or "rain-barrel" effect which is characteristic of circuits having the poor transient response associated with a near-sing condition. It is, therefore, necessary to provide some loss (of the order of 4 db, one way loss) even on short toll lines where delay is negligible and echo is unimportant. In general, on longer circuits, echo objectives rather than singing protection determine the required loss.



Echo in a Toll Connection

Figure 2-8

Echo Objectives

The magnitude of the echo that a talker hears will depend on the "echo path loss", which is the sum of the return loss at the distant hybrid and the round-trip loss in the channels.* As mentioned above, however, the subscriber's tolerance of the echo depends not only on the echo magnitude, but also on the round-trip delay between echo and original signal. Immediate echoes sound like sidetone, but long-delayed echoes can be very disturbing. If we cannot reduce the delay, and if we have gone as far as we economically can in improving return losses at the hybrids by impedance balancing, we have only one more way to make the echo tolerable - decrease the echo magnitude by increasing the electrical loss between the talker and the point where the mismatch occurs.

The objective which has been selected is that the probability of a subscriber being dissatisfied with the echo performance of a circuit shall be held to 1% or less. The value of loss which must be inserted to meet this objective for an echo of given delay can be determined by taking into account the following facets of the situation:

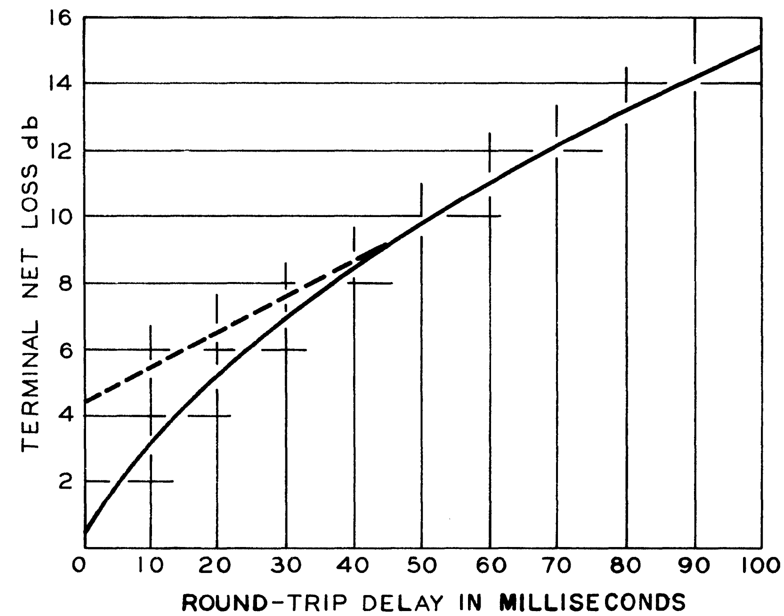
a) The average value and standard deviation of the subscribers' tolerance to echoes as a function of delay;
b) The average return loss at four-wire sets where they terminate in toll connecting trunks, and the standard deviation of this distribution;
c) The variations which toll trunks will exhibit from any nominal value of loss which we assign to them (to control echoes), the variations arising because maintenance is never perfect;

Data on these three variables shows them to be very nearly "normal" or Gaussian distributions; combining them, we can compute the solid curve shown in Figure 9.** This gives the values of toll circuit loss, one-way, not round-trip, which would reduce to 1% the chance of a subscriber being annoyed by the echo performance of a one link circuit. (The toll circuit loss is called "Terminal Net Loss" - abbreviated TNL - in this context. TNL will be defined more carefully later.)

---

* For the present we shall ignore the losses of the loop and toll connecting trunk at the talker's end of the connection.

** The way in which the solid line is derived from data on subscriber tolerance, 4-wire set balances, and variations of trunk loss is a good illustration of the use of probability in setting telephone system requirements, and is therefore covered in some detail in an illustrative example appended to this chapter.



Relation Between Terminal Net Loss and Echo Delay

Figure 2-9

Although the solid line of Figure 9 satisfies the echo requirement for one link, it does not provide enough loss to protect against sings and near-sings, nor does it take into account the impact of multiplicity of links, which has two effects:

d) It forces us to assign more loss to multi-link circuits than we would to a one link circuit of the same delay, in order to meet the 1% objective in the face of the increase in variability from nominal loss (see c above). For example, the TNL for a two link circuit must be 0.4 db greater than for a one link circuit of the same delay; a three link circuit of the same delay would require 2 x 0.4 db more TNL than a one link circuit.

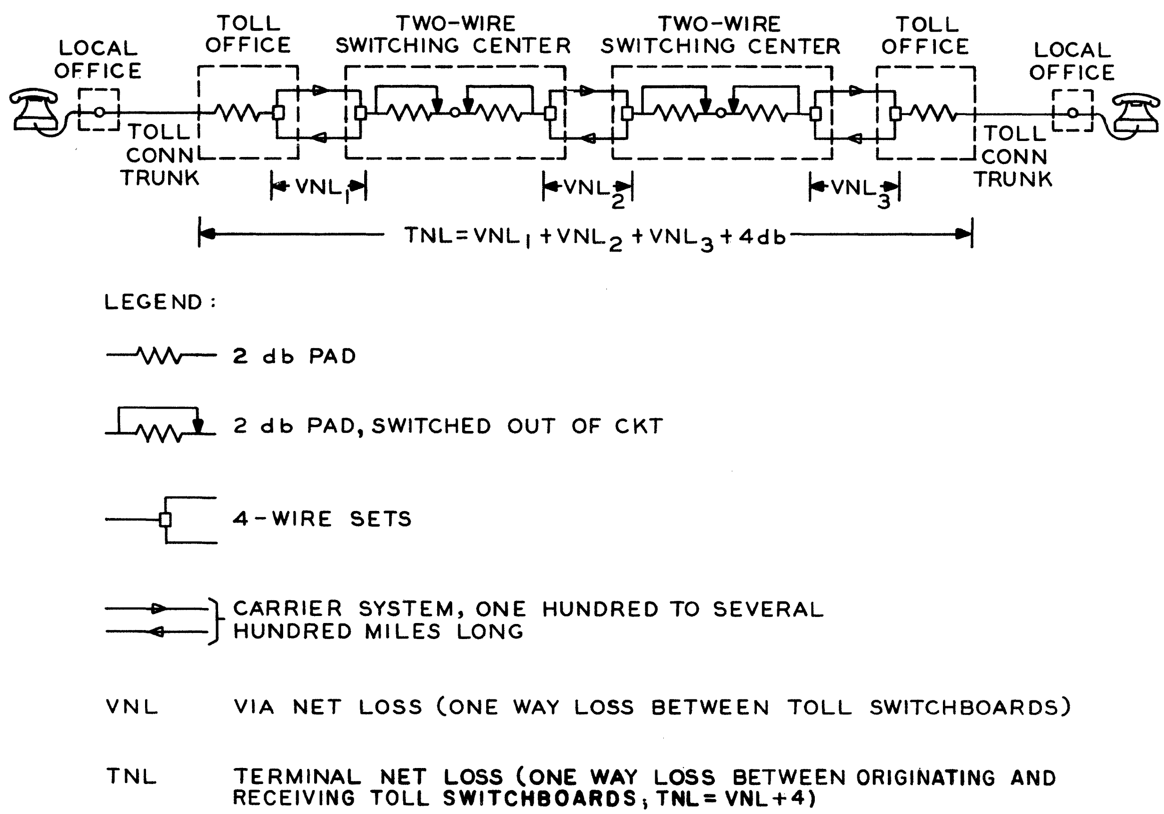
e) It calls for a method of administering the circuits which will permit us to switch in any number of links without building up excessive loss values, while still assuring that single link circuits will have the 4 db minimum loss dictated by singing considerations.

At this point we ask the reader to take on faith the statement that we can meet these criteria by using the dashed line of Figure 9 for one way trunk loss in the single link case, instead of the solid curve.

This line starts at 4.4 db, is linear in db vs delay, and intersects the solid curve at 45 milliseconds delay; above this value of delay we lose interest. The reasons for choosing this dashed line will become clear as we go on with this discussion.

<u>Terminal and Via Net Loss</u>

One way in which circuit components have been arranged in the past to meet the above requirements is shown schematically on Figure 10.



Via and Terminal Net Losses

Figure 2-10

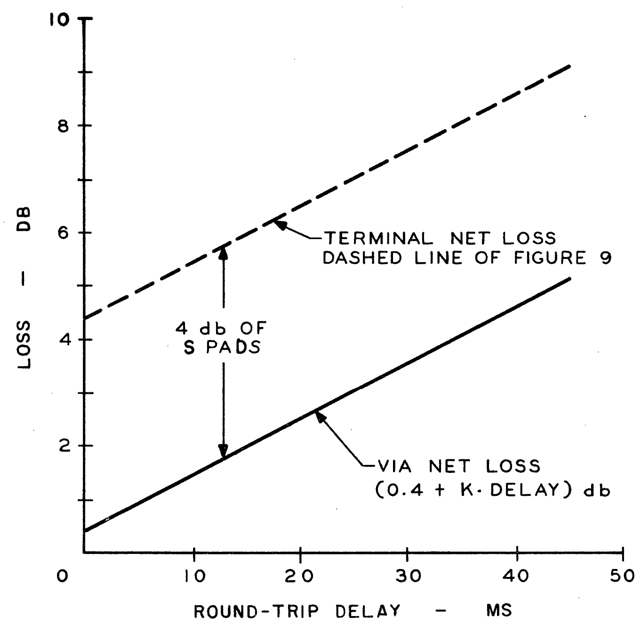
Here a 2 db pad is associated with each four-wire set, to be left in circuit whenever it terminates in a toll connecting trunk, but switched out at intermediate offices. It is assumed, when we do this, that the only echoes of concern are those which occur at the final four-wire terminating set, which faces a two-wire toll connecting trunk in tandem with a subscriber loop. The validity of this assumption depends on the hybrid balance which can be obtained at various points along the circuit. The impedance seen looking into the final four-wire set will depend on the particular trunk, loop, and subset involved. The balancing network of the hybrid must be a compromise choice, and hence the balance obtained on any individual connection is not very good. This produces the dominant echo.

Echoes do arise at other points along the toll circuit where the four-wire toll circuit is brought down to two-wire for switching purposes. The reflection problem at such points, however, is much less severe because the impedances of the following toll link are generally good and the office wiring can be carefully adjusted. The average return loss at two-wire switching points is about 27 db. This has a negligible effect when compared with the return loss average of 11 db at the final four-wire termination. At four-wire switching points, of course, echoes cannot occur, since the circuits provide no return path. At such offices, connection is made to a two-wire facility, such as a toll connecting trunk, by permanently associating the terminating set with the two-wire facility.

Assuming that the echo at the final four-wire set is the only one we have to suppress, then, the circuit of Figure 10 assures us the 4 db minimum loss needed without piling up toll losses at the rate of 4 db per link in multi-link connections. Since at the final four-wire set the reflection really occurs at the connection to the toll connecting trunk, the 2 db pad at that point reduces the echo by 4 db, just as the 2 db pad at the talker's end of the circuit does.

In addition to the 2 db pad at each end of the over-all circuit, each one-way inter-toll trunk is adjusted to a value of Via Net Loss (VNL) which is a function of its delay. The total of the Via Net Losses of the links and the losses of the switching pads is then defined as the Terminal Net Loss (TNL); i.e., TNL = VNL's + 4 db. Like the loss of the pads, each db of VNL is worth 2 db in suppressing echoes - attenuating first the signal on its way to the "mirror", and then attenuating the echo on its return trip.

If we now define the dashed line of Figure 9 as the required Terminal Net Loss for a one-link system, the relationships shown in Figure 11 can be drawn. The values of Via Net Loss that result assure a total circuit loss (VNL + 4 db) which equals or slightly exceeds the one-way loss needed to give a round-trip loss that meets the echo and singing objectives. The VNL curve includes the 0.4 db per link needed to



VNL and TNL for a One-Link System

Figure 2-11

guard against the risk that circuit deviations from nominal may add up in such a direction (too little loss) as to result in too much echo. The fact that the loss-delay relationship is linear (and has a value of only 0.4 db at zero delay) permits us to increase or decrease the number of links in tandem up to 45 ms total delay, without exceeding by very much the TNL needed to make echoes tolerable.

Where delays of more than 45 ms are encountered, echo suppressors are used. An echo suppressor is a voice-operated device which, while one subscriber is talking, inserts as much as 50 db loss in the opposite direction of transmission - the path over which the echo would return. Although they effectively suppress echoes, echo suppressors introduce their own transmission impairments by clipping the beginning and ends of words. Another more serious problem occurs on multi-link connections where two or more circuits equipped with echo suppressors are switched together. It is possible for each subscriber to talk simultaneously and gain "control" over the echo suppressor nearest him. In this case both directions of transmission will have a high loss inserted in them, and a condition of "lock-out" is said to exist, in which neither subscriber can be heard by the other. Because of this possibility, echo suppressors are used sparingly in the plant, and under the present toll switching plan, they are used only between Regional Centers, where the requirement that RC-RC trunks have via net losses of only 0.5 db can only be met by using them.

VNLF - Via Net Loss Factor

On Figure 11, the formula for VNL for one link is given as

$$VNL = K \cdot Delay + 0.4 \text{ db}$$

where K is a constant having a db/ms value. Since the delay of a trunk carried on a given type of facility (such as N carrier) is proportional to length, we can translate from a loss vs delay relationship into a loss vs length relationship for such a trunk if we know the speed of propagation over it. This gives us a db per mile number which is called the Via Net Loss Factor, or VNLF. Thus, we can write

$$VNL = VNLF \cdot Length + 0.4 \text{ db}$$

Various types of systems will have various VNLF's, depending on their speeds of propagation. The total VNL for a particular connection is the sum of the VNL's of the links that compose it. Note that this means an added 0.4 db for each additional link that appears in a connection of a given length.

Thus, if we have a 500 mile link with a VNLF equal to 0.01 db per mile, the via net loss of the system can, in most cases,* be adjusted to

$$VNL = 0.01\ (500) + 0.4$$
$$= 5.4 \text{ db.}$$

Two links extending the same distance would have VNL = 5.8 db.

Via net loss factors for various types of circuits vary from 0.04 db/mile for two-wire voice frequency loaded cable circuits to 0.0014 db/mile or less for broadband carrier circuits. Clearly if we want long low-loss trunks, carrier systems are to be preferred.

Elimination of S-Pads

In conclusion, just to bring the story up to date, it might be mentioned that in a recent effort to minimize high circuit losses** a change of practice in administering the so-called "Switch Pads" (2 db) has been instituted. This is called VNL operation. By associating the loss

---

* Strictly speaking, the "Via Net Loss" is set by echo only, but operation at a higher value, called "Working Via Net Loss" may be necessary because of crosstalk, or to reduce noise at the receiving terminal. In similar fashion, a higher value can be assigned to systems which have impedance discontinuities (hence multiple echoes) along the line as well as at the terminating links.

** As discussed in Chapter 3, the objective is that the loss of any toll connecting trunk should be from two to four db.

of the pads with the toll connecting trunks instead of with the four-wire sets, and using impedance-correcting reactive networks, if necessary, it is possible to omit the pads if the toll connecting trunk is lossy (more than 2 db) since: a) a talker's lossy toll connecting trunk attenuates the echo he hears and b), at the receiving end a lossy trunk tends to give a moderately good termination by masking the possibly poor impedance of the subscriber loop. On a call which involves two lossy toll connecting trunks, omitting the switching pads reduces the total effective loss by 4 db; for example, a 15 db effective loss can be reduced to 11 db. Putting this into effect involves re-wiring offices and also makes it necessary to add impedance correcting reactive networks (particularly in toll connecting trunks whose loss is marginally low) to guard against singing. When a toll connecting trunk has less than 2 db loss, the pads are retained and no impedance correcting network is used.

Overall Grade of Service

The loss associated with intertoll trunks in the plant will vary widely depending on the length and type of facility. In the present plant it is the terminal net loss which characterizes the loss of an inter-toll connection. In estimates for the future plant as based on the toll switching plan, the important loss is the via net loss. This is because VNL operation (described previously) which associates pad loss with terminating links rather than inter-toll trunks, is part of that plan. In the present plant, terminal net loss which may involve up to five links has an average value of 9 db and a standard deviation of 2 db.

One estimate of the distribution of via net losses between toll centers of the future plant has an average value of 4 db and a standard deviation of 1.3 db. This takes into account data, like that given in Table I of Chapter 1, on the relative frequency of occurrence in a connection, of various numbers of links. These averages and standard deviations of inter-toll trunk losses, present and projected, are summarized in Table 1 below, along with similar parameters for talker volume into the inter-toll trunk at the first toll center (the usual point for volume measurements), and losses in the receiving terminating link and subscriber loop.

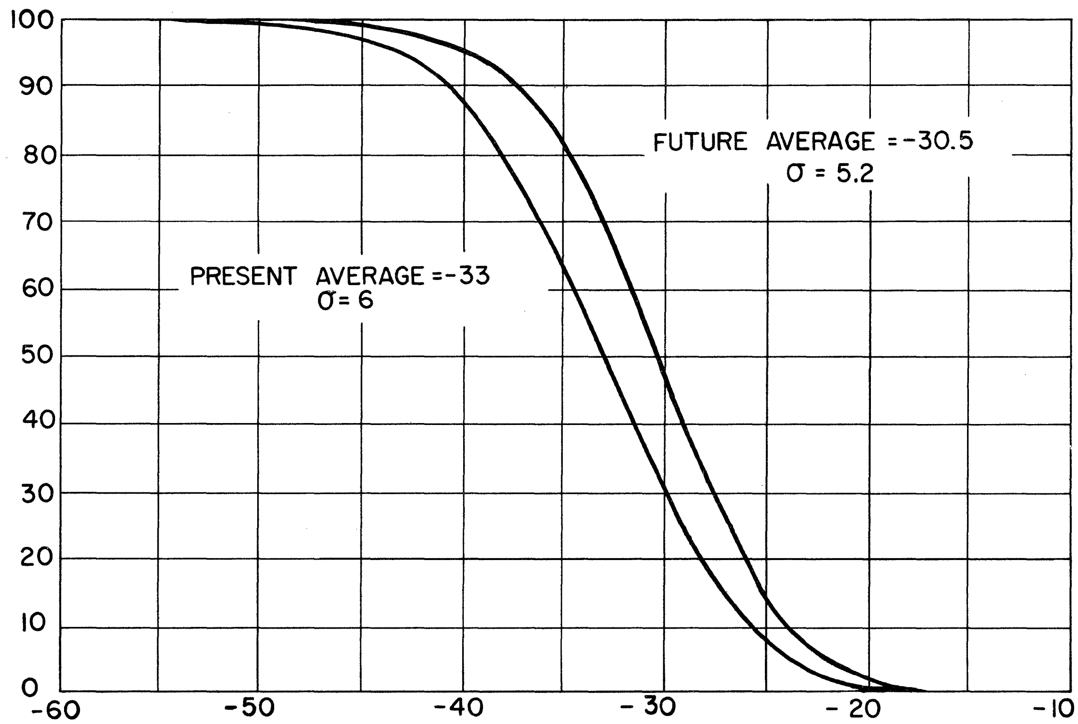
Table 1

	Present		Future*	
	Average	Standard Deviation	Average	Standard Deviation
Talker volume into inter-toll trunk	-15vu	5.3 db	-19 vu	4.8 db
Loss of inter-toll trunks	9 db	2.0 db	4 db	1.3 db
Loss of terminating link	4 db	1.0 db	2.5 db	0.8 db
Loss of subscriber loop (including end office)	5 db	1.5 db	5 db	1.2 db
Received volume	-33 vu	6.0 db	-30.5 vu	5.2 db

The average received volume is obtained by subtracting the average losses from the received volume into the inter-toll trunk, while the standard deviation of the received volume is obtained by adding the squares of component standard deviations and taking the square root of the sum. It should be pointed out that the values given for talker volume into the inter-toll trunk apply to toll connections only. Volumes into trunks on local calls tend to be lower because talkers speak more quietly and, in the case of the present plant, battery supply current is lower than on toll calls. In the future plant the same (lower) battery supply current is expected to be used on both local and toll calls. Because of this lower battery current and because it is expected that subscribers will talk more quietly in the future, talker volumes into the inter-toll trunk

---

*It should not come as a surprise that the set of numbers given here is but one estimate of the performance of future plant and future subscribers. Other estimators arrive at considerably different values. For example, one school of thought believes that talker volumes (at zero level) may well go back up rather than down, as is indicated in the table. Hence, in the design of systems, the load carrying capacity is usually determined on the basis of an average volume at zero level of -12.5. If the lower estimates then prove correct, it will be possible to trade some of this load carrying capacity for improved noise performance which might well be necessary at that time. If we used the lower values of volume the combined effect of having guessed too low on top of tighter noise requirements might well prove disastrous.



Distribution of Received Volumes on Toll Calls

Figure 2-12

are expected to be several db lower than at present. Yet in spite of this, received volumes in the future plant are expected to be higher than at present, because of the lower trunk losses.

The parameters for received volumes in Table 2 describe distributions which may be assumed to approximate the normal law (measured volumes into inter-toll trunks are a very close approximation). These distributions of received volume for the present and future plant are shown on Figure 12.

The observers' opinion data of Figure 6 may now be applied to determine customers' reaction to the grades of service represented by the two distributions on Figure 12. This is a problem in compound probabilities: the probability that a given value of VU will be considered good (or fair, or poor) by observers, and the probability of occurrence of that value of VU in the distribution of received volumes. The compound probabilities must be summed up for all values of VU in the distribution of received volumes. This is a straightforward process for the present plant, but in the case of the future plant an adjustment must be made to allow for the effect of increasing the number of

500 sets in service from the present small number to about 75% of the total. Because of the greater receiving efficiency of 500 sets, a received volume 4 db lower in the future plant will produce the same subjective effect as a given received volume in the present plant. Therefore, for the future plant, the values in the abcissa of Figure 6 should be reduced by 4 db (e.g. -30 to -34).

The results of combining the probability distributions, directly for the present plant and with the 4 db adjustment in the case of the future plant, are given in Table 2. This shows subscribers' reactions to the grade of service now provided on toll connections, as well as estimated reaction to the projected plant.

Table 2

Toll Calls Judged	Present Plant	Future Plant
Too Loud	0%	0.3%
Good	78.6	95.9
Fair	17.9	3.6
Poor or worse	3.5	0.2

To show how the numbers for Table 2 were obtained, we will calculate the percentage of calls judged "fair" in the present plant. Let x = the minimum received volume a particular subscriber will judge "good". Let y be the received volume observed on a particular connection with a given talker. From Figures 6 and 12 respectively we can obtain the fact that for x, the mean is -39 vu, σ is about 4.5 vu and for y, the mean is -33 vu, $\sigma = 6$. Form the new variable $z = y-x$. This will also have a normal distribution with a mean equal to $-33-(-39)= 6$ vu and

$$\sigma_2 = \sqrt{(4.5)^2 + (6)^2} = 7.5 \quad .$$

By definition, a volume will be called too quiet if $y-x$ is negative or, in other words,

$$\text{fraction of calls judged too quiet} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{(z-\bar{z})^2}{2\sigma_2^2}} \, dz \quad .$$

From probability tables this is found to be 21.4%. By following a similar procedure it is possible to calculate the percentage of calls judged lower than fair as 3.5%. Then the number of calls judged fair will be the difference between these two percentages or 21.4-3.5 = 17.9%.

Bibliography

A New Standard Volume Indicator and Reference Level - H. A. Chinn, D. K. Gannett, R. M. Morris - BTL Monograph B-1201.

2B Noise Measuring Set and Noise Measuring Set per D-157641 Bell System Practice, Section E40.459.

Message Circuit Noise - Factors Involved in Determination of Magnitude and Evaluation of Effects. Bell System Practice - AB63.375.

Transmission Rating System - Bell System Practice, Section 980.150.01, (A Discussion and Description of EATMS).

A Forward Look in Transmission - A talk by H. R. Huntley April 30, 1954. A.T. and T. Co.

Appendix on Noise Measurement

I

As this book goes to press, work is being completed on a new noise meter. We will now have come full circle, for this meter reads noise in dbRN. It uses a frequency weighting which is a compromise between F1A and T1U (it is appreciably flatter than F1A) and will read 0 db RN with -90 dbm input at 1000 cps. It is estimated that 0 dbm of white noise will read 88 dbRN. Since this is a new design, the 7 db correction which was necessitated by modification of the 2B meter will be done away with. The new meter will have many other advantages over the old, not the least of which is a great reduction in weight.

II

There are two schools of thought on the validity of adding dba of tone and noise. One school would agree to the procedure we have just gone through. The other, while agreeing that this would be the noise meter reading, would argue thus: "Subjective tests have shown that tones which seem tolerable at first become very annoying if they persist in the background during a conversation of several minutes. Therefore a more stringent requirement should be used; tones which are present all the time should not exceed +24 dba at zero level, if they are to be consistent with a quality of transmission which tolerates thermal type noise of +38 dba at zero level. This being so, the addition of annoyances on a power basis is misleading."

Illustrative Example #1

This problem illustrates how one point on a curve like the solid one of Figure 9 of Chapter 2 might be derived, and should also make clear the reason why 0.4 db is included in the expression for VNL.

1. a) Two cities 1300 miles apart are connected by a 4-link K carrier system. Find the terminal net loss required so that only one percent of the subscribers using these circuits will find the echo unsatisfactory. Do this by combining distribution curves, using the attached data (I, II, III, IV). Do not use Figure 2-9 of the notes, which is based on similar but superseded data. This problem calls for deriving a new Figure 2-9 or at least one point on it. Assume that toll connecting trunk losses are to be neglected - i.e., taken as 0 db.

   b) Same as above, but this time assume a 2-link connection, same total length.

   c) And for one link, still the same 1300 mile length?

Reminders

The standard deviation "σ" of the normal distribution obtained by adding "n" distributions is

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 \ldots \sigma_n^2}$$

where σ_1, σ_2 etc. are the standard deviations of the individual distributions.

For a normal distribution, the 1% point is 2.33σ away from the average or median value.

The methods used in Monograph 2184 are of interest in connection with this problem.

Data

I. The following table gives the loss in the echo path (round trip) required to make echo just tolerable to the average observer. (The data was taken using representative subscriber loops.) The observations at each value of delay showed that subscribers formed a normal distribution with a standard deviation of 5 db.

Table 1

Round Trip Delay, ms	Required Loss, db
0	-2
10	7
20	13.5
25	16.0
30	18.0
40	21.5
50	24.0
100	35.0

II. The return losses at 4-wire sets when terminated in representative toll connecting trunks will form a normal distribution with an average value of 14 db and a standard deviation of 1 db. (This is believed to be an achievable objective even though it may not be true yet.)

III. The maintenance of toll circuits may be expected to result in channel losses which form a normal distribution having correct average values of VNL (as yet unknown, and to be found) with a standard deviation (round trip) of 2 db.

IV. The velocities of propagation, hence, the delay, in various carrier systems (including the effects of filters, etc., as well as the propagation medium) are shown in the following table.

Table 2

Facility	Velocity of Propagation (miles/second)	Delay (milliseconds/mile)
Open wire carrier (J or O)	117,500	8.60×10^{-3}
Cable Carrier (K or N)	105,000	9.55×10^{-3}
Coaxial (L1 or L3)	133,000	7.55×10^{-3}
Radio (TD-2)	142,000	7.10×10^{-3}

Remarks

Let us first recall some of the points discussed in Chapter 2. In the first place, terminal and via net losses are one-way losses between corresponding points (e.g., switch boards, or toll connecting trunks themselves) in the various toll offices involved. Second, since switching pads are designed to have, as their image impedance, that value which would give a good hybrid balance, the use of switching pads moves the point at which echoes arise to the junction of pad and toll connecting trunk. Thus a two db switching pad first attenuates the signal, then the echo, and increases round trip echo path loss by four db. Recall that it is the fact that local plant impedances are not uniform, and cannot be matched by a single choice of hybrid balancing network, which causes reflections. This also means that echoes at intermediate offices, where local plant is not involved, should be negligible.

One assumption made in the following computations calls for comment. It is this: that the round-trip loss of a link (terminal and high frequency line equipment between two toll offices or switching centers) will vary around any assigned nominal value by the same amount (σ=2 db) regardless of link length. This is not quite true; on the other hand, this sort of variation will not be a direct function of link length. It is made up of measuring errors, terminal variations, and high-frequency line variations. Only the last is a function of length, and it might be argued that even this effect is modified by the possibility that maintenance practices may call for more frequent readjustment of the longer links. The assumption made, therefore, seems the most reasonable one to make in the absence of more data.

Solution

The round trip delay is 25 milliseconds, from IV. If the problem of distribution curves were not involved, and we could use merely the average values given in I, II, and III, the answer to a, b and c would be merely:

From I:	Round trip loss required	16.0 db
From II:	Return loss, 4-wire set	14.0 db
Hence:	Loss to be inserted, round trip,	2.0 db
	TNL, one way,	1.0 db

Using this value of TNL would not satisfy our objectives, of course. Even if all four wire sets had exactly 14 db return loss, and trunks never varied from their nominal TNL's, only 50% of the subscribers would be satisfied. To satisfy 99% of them (still assuming 4-wire sets and trunks do not vary) we would have to increase the loss to get to the 1% point on the distribution curve of Figure A, to give a round trip loss of

$$2\ \text{TNL} = 2.0\ \text{db} + 2.33\sigma = 13.6\ \text{db}$$

which would make TNL = 6.8 db.

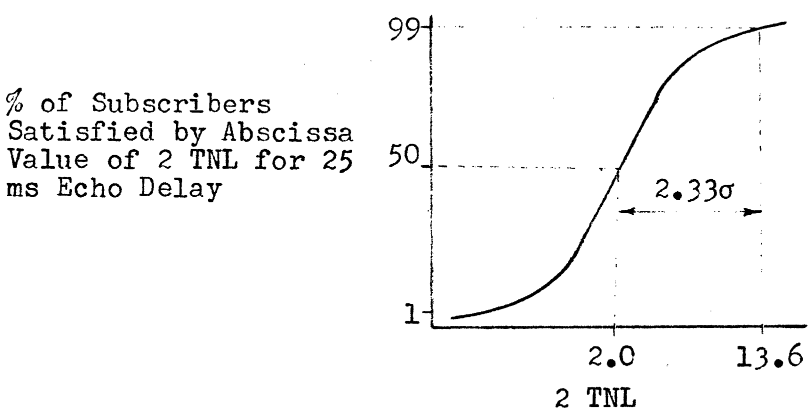


Figure A

We still have not taken into account the fact that trunk losses and 4-wire set return losses will also vary. To some of our readers, the correct method for doing this may be obvious; they will find the following discussion boring, and are to be envied. Most of us have to take such questions in easy steps the first few times we meet them.

Let us consider for a moment how one might obtain the data on which Figure A is based. Figure B shows a possible experimental set up, in which we would vary the attenuator until each observer was satisfied. (The 14 db fixed pad represents the return loss of the 4-wire set.)

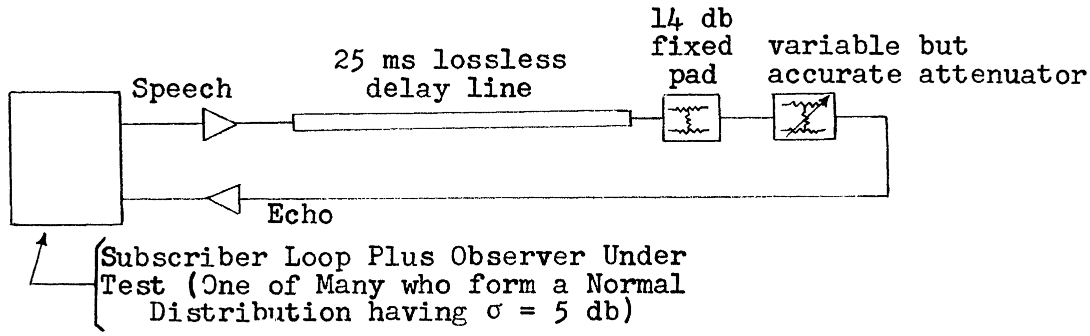


Figure B

Suppose now that instead of an accurate attenuator, we have one with a loose knob or dial, so that when it actually has 4 db of loss, for example, it reads 3.7 or 4.3 or some other number exhibiting a random distribution around the correct value - a normal distribution with a standard deviation of 2 db, to be precise. Even a single consistent observer, repeatedly tested, would now yield a universe of values instead of a single number of db for just satisfactory echo. The curve corresponding to Figure A would spread out. Assuming the attenuator errors to be random, with no systematic error (or bias) the 50% point would still be at 2.0 db. However, the readings would spread out more, and the resultant curve would have a standard deviation corresponding to the addition of the two distributions involved, which is

$$\sigma = \sqrt{5.0^2 + 2.0^2}$$

Similarly if we used a number of different 14 db pads, taken at random from a universe of 14 db average and 1 db standard deviation, our data would spread still more (same average, though) and would have a standard deviation of

$$\sigma = \sqrt{5^2 + 2^2 + 1^2}$$

Finally, if we use four attenuators like the one described* (each attenuator corresponding to a link in our four-link case), the standard deviation of the data would be

$$\sigma = \sqrt{5^2 + 2^2 + 2^2 + 2^2 + 2^2 + 1^2}$$

$$= \sqrt{5^2 + 4\ (2^2) + 1^2} = 6.5 \text{ db}$$

It is on this curve that we must find the 99% satisfied value. Recalling that the average has remained 2.0 db for 2 TNL, we see that

$$2 \text{ TNL} = 2.0 + 2.33\ (6.5) \text{ db} \doteq 17.2 \text{ db}$$

A satisfactory value of TNL, then, is 8.6 db for the 4 link case.

Similarly, for the two and one link cases we obtain

TNL, two link case $\doteq$ 7.8

TNL, one link case $\doteq$ 7.4

This 7.4 db value would be one point on a new solid line (using new data as given above) for Figure 2-9.

We observe that although other factors in the problem remain the same, the required TNL goes up by about 0.4 db when we go from 1 link to 2 links, and by another 0.8 db when we go from 2 to 4 links, because of the changes in σ. Similar results would be obtained if we computed the required TNL for various numbers of links at other values of delay. We conclude, therefore, that the effect of the maintenance variation is to increase the required TNL (or VNL) by about 0.4 db over the value necessary if the circuit loss never varied from its assigned value.

---

*We would be forced into reading negative losses on at least some of the attenuators if we kept adding these variable attenuators to the circuit. If you wish to avoid this, substitute an amplifier plus a larger attenuator for each variable attenuator.

Chapter 3

VOICE FREQUENCY TRANSMISSION

The telephone instrument and its performance characteristics are described. This is followed by a discussion of the objectives for loops and trunks of the exchange area plant, and the interaction of signalling considerations on transmission problems. Voice frequency amplification methods used in both toll systems and the exchange area are reviewed. Sources of impairment in voice frequency circuits are described, with particular attention to noise and crosstalk. Throughout the chapter, consideration is given to the impact of recent developments on the exchange area plant.

Introduction

The generation, transmission, and reception of voice frequency telephone signals are the oldest and most basic problems encountered in the practice of transmission. The evolution over many decades of methods and theory for solving these problems has given us a vast body of knowledge, covering in great detail many of the problems of any consequence. The objective of this chapter is to present a general picture of typical methods used, and of the basic problems encountered in handling voice freouency signals.

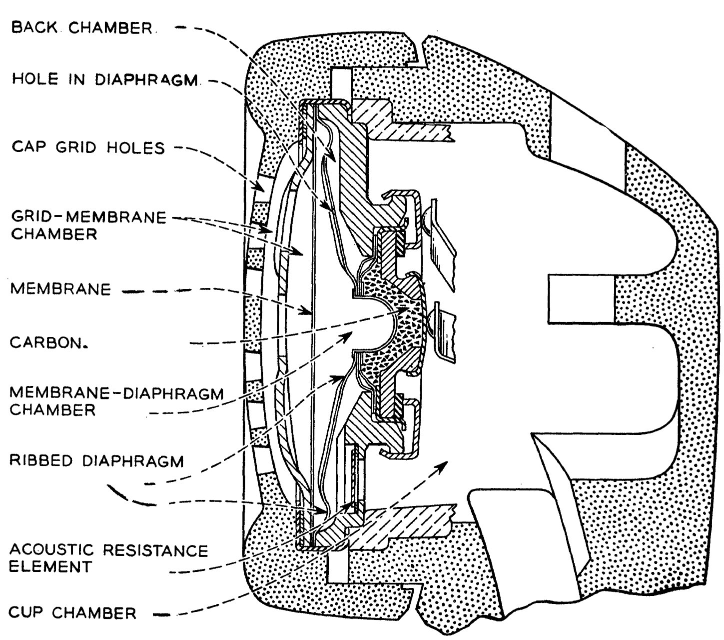
To accomplish this purpose we will first describe the telephone instrument and its associated performance characteristics. Next, we will discuss the losses of loops and trunks, and the bearing they have on the composition of the exchange area plant. This will be followed by a review of voice frequency amplification methods as applied to the exchange area and to voice frequency toll circuits. Following this, the most important types of interferences will be discussed.

The performances and limitations of the elements of the voice frequency telephone plant are gaining a new importance that goes beyond the simple fact that they form a part of every telephone connection. New services such as data-phone, the need for improved transmission performance, the application of multiplex transmission to existing wires, and the impact of electronic switching will require increased understanding of the voice frequency plant.

Telephone Set - Description

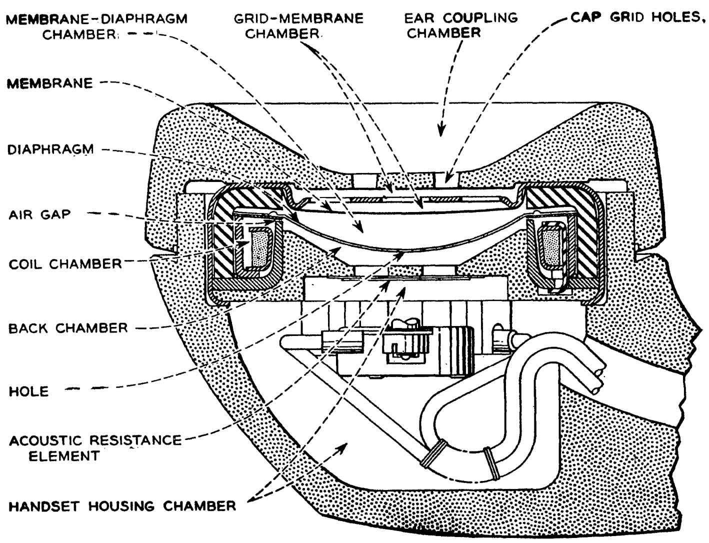
The telephone set is composed of a transmitter, a receiver, and the associated circuitry for suppressing sidetone, connecting power, and signaling. In a modern telephone transmitter, granules of carbon are held between two electrodes - one, a cup holding the granules and the other a diaphram. The contact resistance between the granules is changed by sound pressure on the diaphram. The resulting resistance variation modulates a battery current flowing between the electrodes, thereby translating the acoustic message into an electrical signal. In the receiver, the varying component of this current passes through a winding on a permanent magnet. The alternate strengthening and weakening of the magnetic field causes a diaphram to vibrate, which generates sound waves corresponding to those of the talker. Cross sectional views of the Tl transmitter and Ul receiver used in the 500 type telephone set are shown in Figures 1 and 2.

The transmission circuit of the telephone set must electrically separate the transmitter and receiver to limit the amount of the talker's



Cross Section of Tl Transmitter

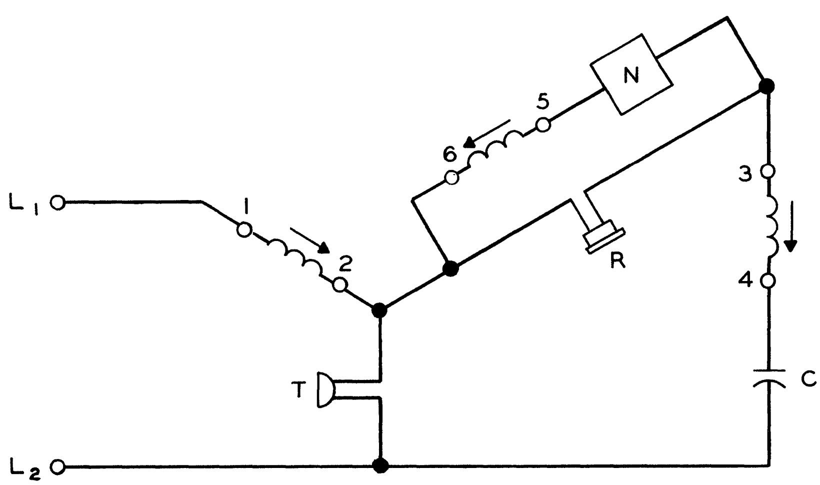
Figure 3-1



Cross Section of Ul Receiver

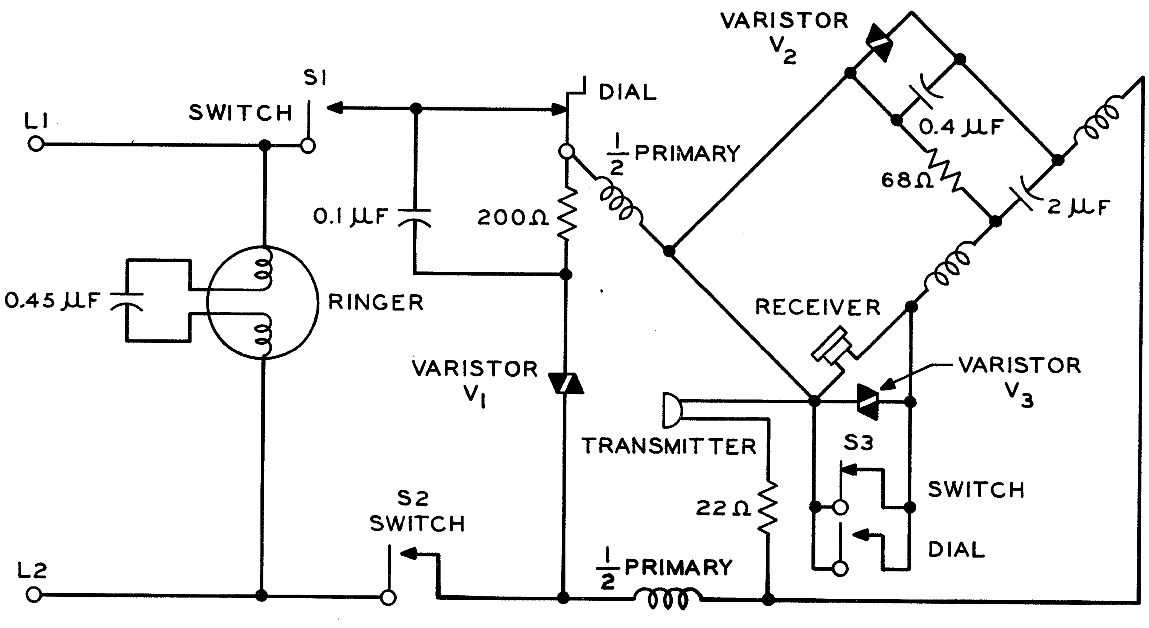
Figure 3-2

signal (sidetone) appearing in his own receiver, and to block the direct current in the transmitter from the receiver. Subjective tests have shown that some coupling must be allowed between the transmitter and receiver to provide a controlled amount of sidetone. However, too much sidetone will cause the talker to lower his voice, thereby reducing the volume which the listner receives. The standard common battery anti-sidetone circuit for accomplishing this purpose is shown in Figure 3. The three-winding transformer and the network N form a hybrid which places the transmitter in conjugate with the receiver. Capacitor C prevents the direct current flowing in the transmitter from appearing in the receiver.



Common-Battery Antisidetone Circuit

Figure 3-3



Schematic Diagram of 500D-Type Telephone Set

Figure 3-4

The schematic circuit of the 500D type telephone set is shown on Figure 4. When the hand set is on its mounting, the "switchhook" contacts S_1 and S_2 are open and S_3 is closed to protect the transmitter and receiver from ringing currents. Removal of the hand set allows direct current from the central office (or a local battery) to pass through the transmitter and removes the short from across the receiver. The dial contacts interrupt the battery current to form the dial pulses required to control the central office equipment. During dialing, contacts across the receiver are closed.

The modern 500 type telephone set incorporates a number of characteristics and features that represent improvements over the earlier 302 type set. The efficiency of the transmitter has been increased about 3.5 db and the receiver efficiency about 5 db. The frequency responses of both the transmitter and receiver have been extended. Components added to the basic common-battery anti-sidetone circuit provide a dial pulse filter to suppress high frequency interference into radio sets, a varistor (V_3) to suppress clicks in the receiver, an improved balancing network (necessitated by the improved transmission characteristics of the instruments), and an "equalizer", employing two additional varistors, to reduce transmitting and receiving sensitivity on short loops.

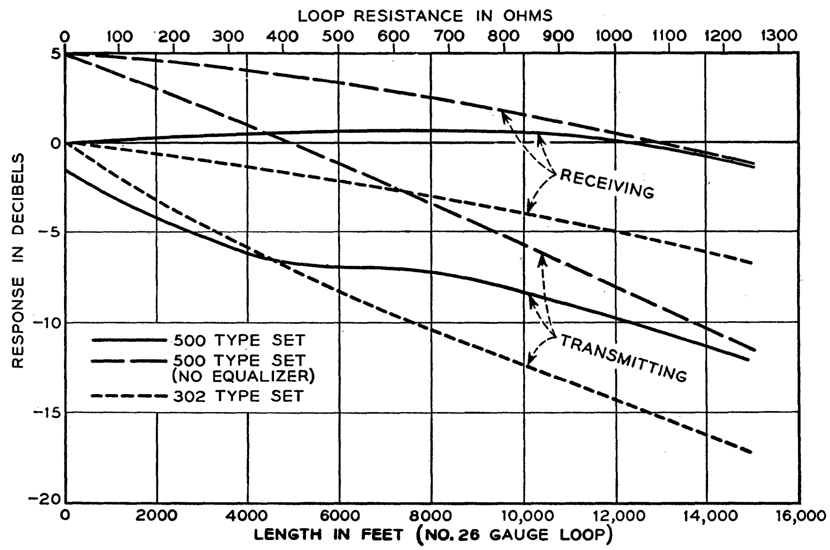
This equalizer helps to solve an important transmission problem in telephone set design, namely the interdependence of the transmitting and receiving efficiencies and the wide range of allowable resistance in local loops. The high efficiency desired for the longest loops would provide excess high frequency transmission on the short loops. The increased sensitivity of the 500 set makes possible increased loop length or the use of wire of reduced size and hence, higher resistance. On low resistance loops, this sensitivity is detrimental and must be reduced. This is a function of varistors V_1 and V_2. On long loops, the direct current from the central office battery is low. The varistor impedances are therefore high, and maximum telephone set efficiency is obtained. On short loops, the high direct current results in low varistor impedances, which shunt the speech currents and reduce the set efficiency. As Figure 5 shows, the combined response of a receiver loop (#26 gauge) and station set with the equalizer is essentially constant for any loop length between 0 and 12,000 feet. For the transmitting case, the response does not vary appreciably for lengths between 4,000 and 10,000 feet with variations outside this range being substantially reduced over those which would occur without the varistors. The overall effect is to produce more nearly uniform volumes at the central office from transmitting loops, and more nearly uniform volumes at subscribers' receivers, or at least to make these volumes less dependent on loop length.

Varistors V_1 and V_2 also serve an additional purpose. By a mechanism similar to the one described for the equalizing function, they compensate for differences in subscriber loop impedances which would otherwise tend to produce unbalance in the sidetone circuit. Prevention of significant unbalance is necessary because the greater efficiencies of the receiver and transmitter makes excessive sidetone more objectionable.

<u>Telephone Connection - Performance</u>

The quality of the transmission over a telephone connection will depend on the received volume, the relative magnitude of the transmission of different frequencies, and the interferences. In a typical local connection the ratio of the acoustic pressure at the transmitter input to the corresponding pressure at the receiver output will depend upon:

1. The translation of acoustic pressure into an open circuit voltage in series with the transmitter impedance.



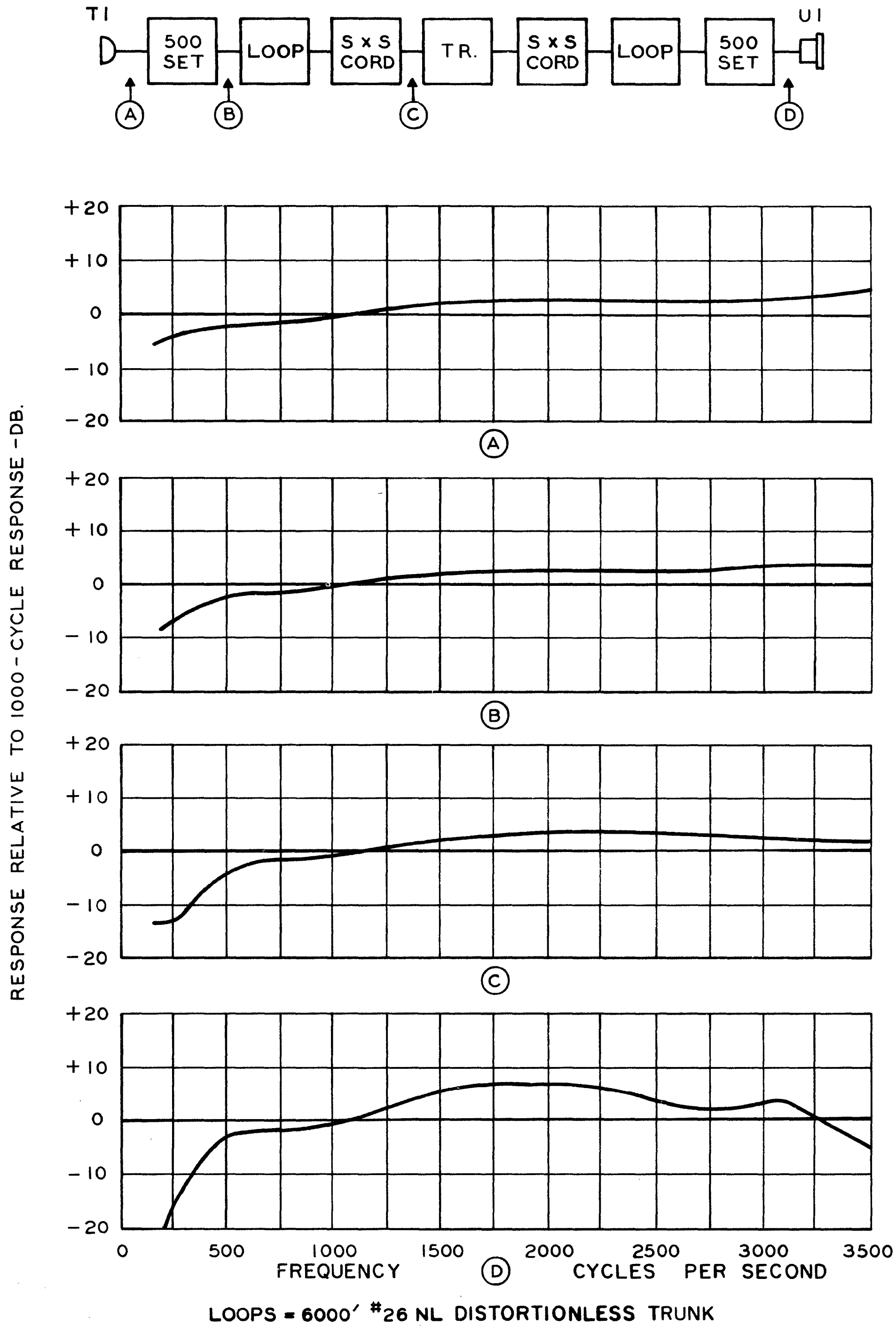
Relative Response of 302 and 500 (A or B) Type Sets (D Type Set has Slightly Higher Gain for Short Loop Lengths)

Figure 3-5

2. The loss between this open circuit transmitter voltage and the voltage across the terminals of the subset, introduced by the circuit of the telephone set.
3. The loss of each subscriber's loop, the central office equipment, and the local trunks.
4. The loss between the terminals of the listener's set and the voltage appearing across the terminals of the receiver.
5. The translation of this voltage, acting across the receiver impedance, to acoustic pressure at the receiver output.

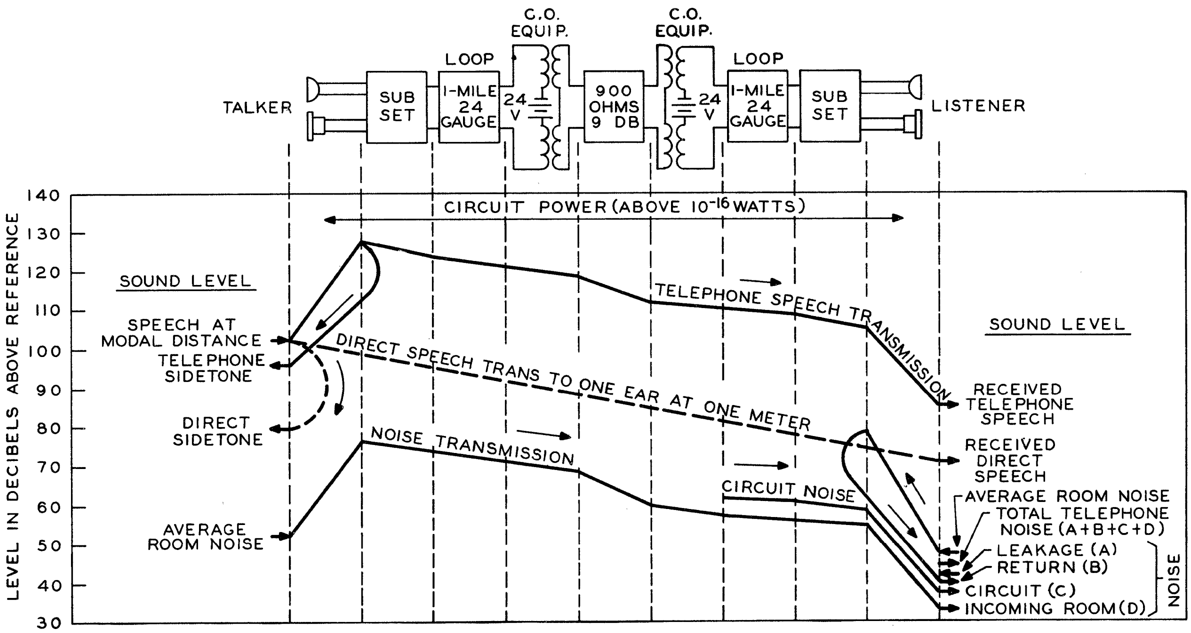
Figure 6 shows the typical transmission vs. frequency characteristics (normalized at 1000 cps) encountered as one progresses through such a connection.

The importance of speech volume and the losses in various parts of a telephone connection can be seen by examining the power level diagram for a typical telephone connection. Figure 7 shows the speech and noise power at various points in a local connection involving a trunk between two central offices. In this particular connection the listener has a signal-to-noise ratio of about 35 db. This is considered good.



Transmission - Frequency Characteristics, Normalized at 1000 cps, of Various Points Along a Typical Telephone Connection

Figure 3-6



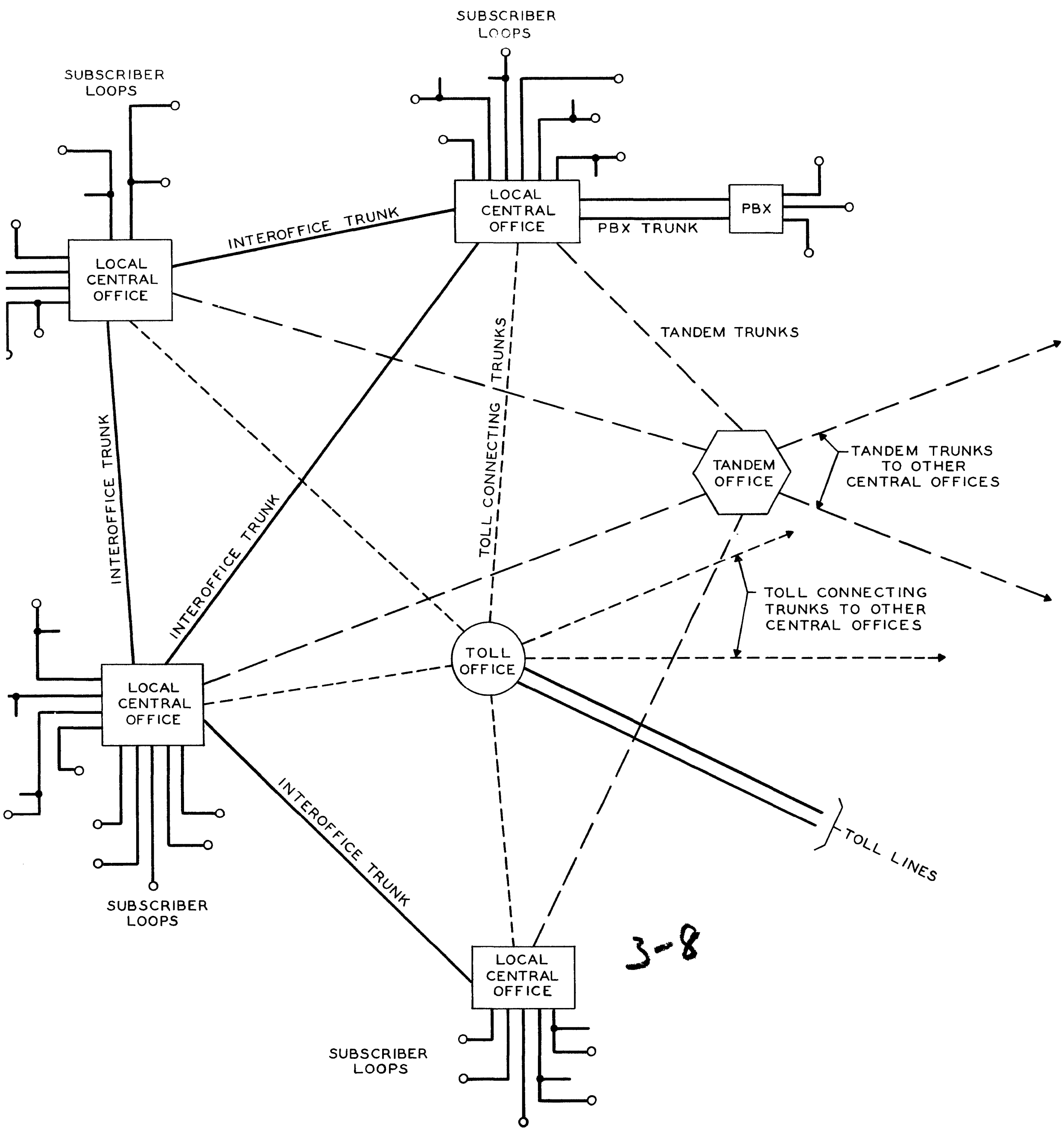
Level Diagram of Typical Telephone Connection

Figure 3-7

With a weak talker this ratio might drop to a low but still usable value of about 20 db. Additional losses in the trunks or loops could further reduce the received volume and, depending on the magnitude of room noise at the receiving end, might lead the listener to ask the talker to speak louder.

Resistance Design of Loops

An "exchange area" is the area served by a single central office. In a more general sense, we may speak of the "exchange area plant" as including the metropolitan area served by numerous central offices or a small city with a single telephone office. In the exchange area, subscriber's loops connect each telephone with its central office and trunks inter-connect the offices. A typical pattern of loops and inter-office trunks is shown on Figure 8. Before the advent of the 500 set, with its improved efficiency and equalization provisions, engineering the local plant required selecting a transmission objective or standard which connections in a given area should meet. The problem then became one of determining the costs of different combinations of wire sizes and office locations required to meet these standards so that the most economic plant layout could be made. Both transmission and signaling had to be considered, and one or the other might govern.



3-8

Possible Layout of Loops and Trunks in Part of an Exchange Area

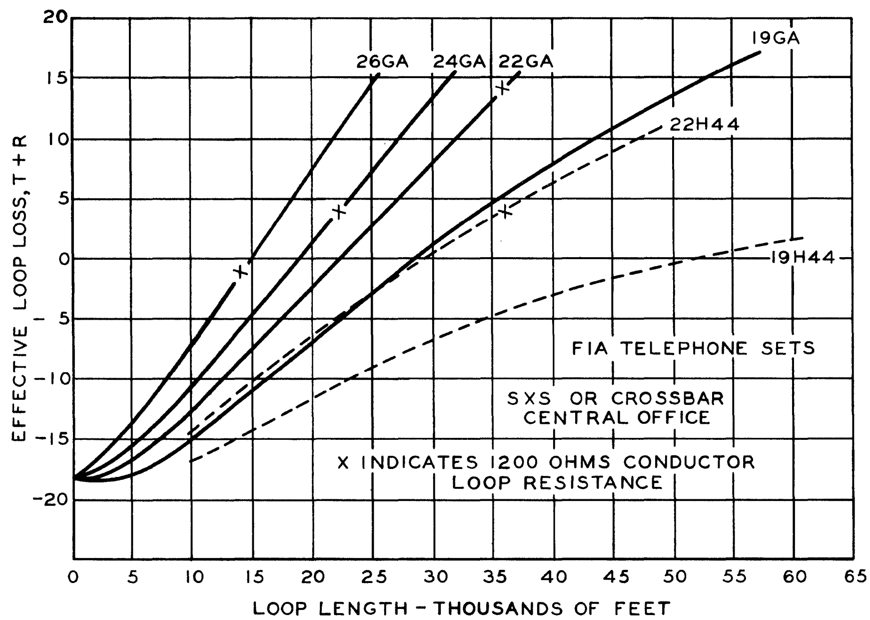
Figure 3-8

Since the introduction of the 500 set, the selection of wire sizes for various loop lengths has been based purely on signaling criteria. Signaling performance is limited by the direct current resistance of the loop. In the older types of manual central offices, signaling restricts the resistance of the subscriber loop and telephone set to a maximum of 635 ohms. In the newer crossbar offices, this total resistance can be about 1500 ohms. Allowing 200 ohms for the set, the total pair resistance can be 1300 ohms. Of this, 1200 ohms is allowed for the pair at normal temperature and without loading, and 100 ohms for temperature effects and loading coil resistance. It is now permissible to engineer local loop lengths and gauges right up to this 1200 ohm resistance limit, providing 500 sets are used on the longer loops (300 sets can still be used on the shorter loops) and provided that loading is used on the loops which exceed 18,000 feet in length to reduce the total attenuation and the attenuation distortion.* The resulting cable economies are considerable; before comparing the results of the two methods of design, however, it is desirable to review some definitions.

In a given telephone connection there will be an effective transmitting loop loss T which is a function of the voice frequency loss of the loop, the direct current flow in the transmitter, and the telephone set characteristics. The actual value of T in db is a rating obtained by comparison of the loop and telephone set under consideration with the reference loop and set of the Working Reference System. Similarly, the effective receiving loop loss R will be a function of the loop loss and the telephone set characteristics. Its value in db is likewise found by comparison testing, using the Working Reference System. In general, the values of T and R are not equal; that is, the effective loss is different for the two directions of transmission. In order that only one effective loss value be assigned to a telephone set and its associated

---

* The loss of a pair of wires below a prescribed frequency can be reduced, and the loss versus frequency characteristic made nearly flat, by inserting inductance periodically. Loading arrangements are specified by a letter designating the distance between loading coils and a number indicating the inductance. Thus, H44 loading means 6000-foot spacing of 44 millihenry coils. H44-25 indicates the loading arrangement for the regular or "side" circuits and a phantom circuit: namely 44 millihenrys on the side circuits and 25 millihenrys on the phantom, both spaced at 6000-foot intervals.



Effective T+R Loop Loss
Versus Loop Length

Figure 3-9

loop, to apply to both directions of transmission, it is common practice in computing the effective loss of a connection to use an average value of instrument and loop loss, namely $\frac{T+R}{2}$.*

Under the old method of loop design, it might be decided that $\frac{T+R}{2}$ in a given area was not to exceed -3 db. Such a decision would be made after weighing the overall grade of service which was to be rendered, the performance of trunks, and the costs involved. At the same time, local signaling considerations might impose a dc resistance limit of, for example, 1200 ohms on the loops. The effect of these two requirements can be seen from Figure 9, which shows the effective loss vs length for various

- -

* The actual value will vary to some extent depending on the battery supply and consequent transmitter current, which is different in local and toll connections. On a local connection, the common battery at the central office is bridged across the line through a high ac impedance circuit to supply dc to the subscriber. On a toll call, the battery power of each subscriber is supplied by inserting the battery at the electrical center of a repeat coil which is used to connect the subscriber loop to the toll trunk. This is done to provide a better impedance match to the toll trunk and to effectively reduce the noise and crosstalk introduced in the toll connection by the central office battery supply. The lower resistance of the toll battery causes more current to flow in the transmitter, thereby increasing the transmitter efficiency on toll calls.

cables. The crosses indicate the loop lengths whose resistances are 1200 ohms. (Although loops are generally made up of a variety of wire gauges, for this illustration it is assumed that each loop is composed of a single gauge.) From Figure 9 and data on wire resistance, the limiting loop lengths and their resistances for effective loop losses of -3 db (i.e., $\frac{T+R}{2}$) can be obtained. These are given in Table I. Obviously the -3 db loss requirement is governing. The corresponding loop lengths which would be obtained using resistance design, 500 sets, and loading above 18,000 feet are also shown.

Table I

	300 Set Transmission Design		500 Set Resistance Design	
Gauge	Maximum Loop Length in Feet	Total Resistance of Loop in Ohms	Maximum Loop Length in Feet	Total Resistance of Loop in Ohms
26	10,500	875	14,400	1,200
24	14,000	725	23,100	1,200
22	17,000	550	37,000	1,200
19	21,000	330	74,500	1,200

It is apparent from this table that for loops longer than about 10,000 feet there will be a considerable saving in cable cost by using the 500 set. Furthermore, when installed on this resistance design basis, the $\frac{T+R}{2}$ effective loss rating of the 500 set and loop will generally fall between -4 and -2 db, regardless of the loop length. Therefore, in engineering the local plant, it is recommended, both from the economic and transmission standpoint, that any 300 set in the office area be kept within about 10,000 feet of the central office. The 500 type sets are then used on the longer loops. This plan leads to a more uniform grade of transmission in the local plant than was possible before the 500 set was available.

Trunks in the Exchange Plant

In contrast to the loop plant, the layout of the trunk plant is governed dominantly by transmission objectives, although supervisory signaling considerations cannot be ignored. Permissible resistance values for supervisory signaling are, for example:

Step-by-step	2000 ohms*
Panel	1300
Crossbar	3000

- -

* New developments may push this to 5000 ohms in the future.

Transmission objectives for trunks between central offices (i.e., inter-office trunks) call for an average loss of about 4 db, with a worst value of 6 db, and a future target, someday, of nearly zero db. Frequency distortion (variation of attenuation with frequency) in the trunks is to be minimized. (In the present plant, many of these trunks still have losses of 8 to 10 db.) Meeting even the present objectives calls for loading all CO-CO trunks which are more than 6000 feet long and the installation of repeaters, when necessary.* To meet the future objective, many of the shorter trunks will also have to be loaded, on a 3000 foot basis. This can be seen from the column in Table II, which shows the effective loss in db per mile of typical cable circuits.

The corresponding objectives for CO to tandem office trunks (i.e., tandem trunks) are 2 to 3 db of loss, at the worst. Trunks between tandem offices have a future objective of nearly zero db, which will call for 4-wire operation, repeatered lines, and careful balancing of 4-wire sets.

For illustrative purposes, Table II tabulates the maximum length in miles for various cable facilities having an effective loss of 6 db, and the corresponding resistance. Clearly either loss or resistance may govern, depending on the type of office involved, but usually, unless repeaters are used, the transmission performance is the dominant consideration

Table II

Gauge and Loading	Effective Loss in db/mi	DC Resistance, Ohms per mile	Maximum Length in Miles for 6 db Effective Loss	Total Resistance of Trunk in Ohms
26 NL	3.4	404	1.7	775
24 NL	2.8	273	2.1	590
24 H 88	1.2	280	5.0	1400
22 H 88	0.8	178	7.5	1340
19 H 88	0.43	92	13.9	1280
19 B 88	0.35	99	17.1	1600
19 B 135	0.27	105	22.2	2320

As discussed in Chapter 2, the present design for toll connecting trunks calls for the elimination of the 2 db switching pads. This is done by associating a minimum loss of 2 db with the trunk, to "pad-out" the generally poor impedances of the local plant; and, when

---

* A repeatered cable may not have to be loaded if the repeater gain-frequency characteristic can be adjusted to compensate for the normal loss-frequency characteristic of the line.

necessary, by adding impedance correcting networks at the toll office to increase the return loss seen looking into the trunk. The maximum loss of the toll connecting trunks, in the present plan, is to be 4 db. These transmission objectives can be achieved by any one of the following means:

1. If the normal loss of the trunk is 0 to 2 db, a 2 db pad is added to the trunk at the toll office. This pad serves the same function as a 2 db switching pad.
2. If the trunk has a normal loss of 2 to 4 db, it will adequately mask the impedance mismatches of the local plant. Its own impedance, however, may not be very good; in such cases it is necessary to add an impedance correcting network at the toll office, to increase the return loss looking into the trunk.
3. If the trunk loss exceeds 4 db* one or more repeaters are added, and the trunk loss is adjusted to equal the quantity (VNL + 2) db. Here, VNL represents the via net loss computed for the terminating link facility as if it were a toll trunk. For typical lengths of toll connecting trunks, the VNL will come out to be 1 db or less, so that the addition of 2 db puts the trunk loss into the required 2 to 4 db range. An impedance correcting network is added to the trunk at the toll office if necessary.

In addition to the above procedures, all toll connecting trunks having lengths greater than 6000 feet are loaded to minimize the attenuation distortion across the voice channel.

Loss in an Exchange Area Telephone Connection

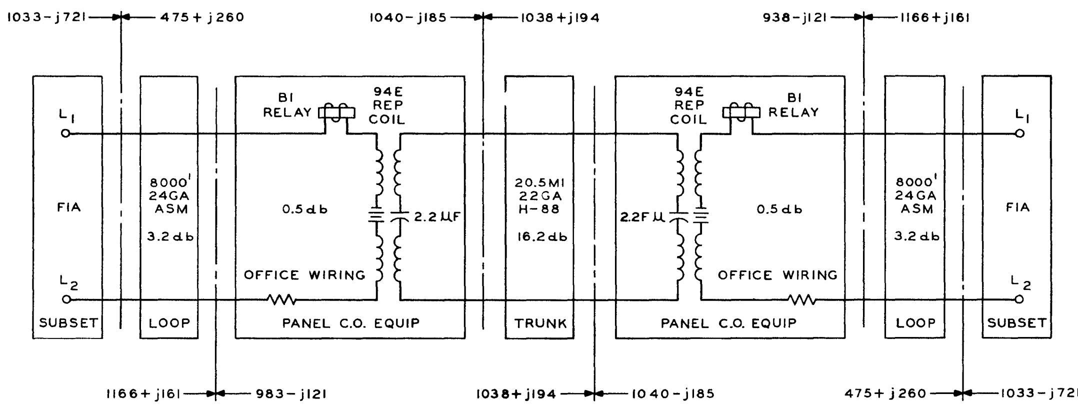
The preceding discussion, which has been used to illustrate the composition of the exchange area plant, is all based on the use of effective transmission loss. As discussed in Chapter 2, this concept includes the effects of frequency distortion and interferences. The true loss of power in a telephone circuit is, of course, a function of frequency and dependent on the actual impedances of the components of the circuit.

Toll trunks are usually electrically long and designed to provide good impedance matches at their terminals and at intermediate points. Furthermore, toll trunks generally introduce no appreciable frequency distortion or noise. (We are ignoring EB banks, of course.) Thus, the attenuation of a toll trunk usually determines its contribution to the effective loss of a circuit.

---

*The toll connecting trunks entering four-wire switching centers will frequently fall in this category, since the four-wire set (approximately 3.5 db loss) is considered part of such a trunk.

In the exchange plant, loops and trunks are normally electrically short and do not have particularly good impedance matches at junctions. Thus, the impedance looking into a particular point will depend on the impedance terminations at remote points. Likewise, the power loss will be a function of the attenuation of the components and the reflection gains and losses at the numerous junctions.



Impedances at Various Points in a Possible Exchange Area Telephone Connection

Figure 3-10

Figure 10 shows an exchange area connection with the 1000 cycle impedance seen at each junction as computed from the equivalent T network for each component. The attenuation of each component is also shown. It is apparent that the impedance requirement for maximum power transfer - namely equal resistances and equal reactances of opposite signs - are approximately met at the junction between the trunk and central office. At the other junctions the situation becomes less ideal as we proceed towards the subscriber. A detailed computation shows the 1000 cps insertion loss between the telephone sets in this connection to be 21.1 db. The sum of the attenuations is 23.6 db. This illustrates the fact that multiple reflections between component parts of an exchange area circuit combine in such a way as to make the summation of individual attenuations an unreliable measure of total loss. For effective loss computations, this has led to the preparation and use of detailed charts which give the junction loss (which may be positive or negative) to be expected when various loop and trunk facilities are interconnected. Using these, the overall effective loss of a connection can be computed as the sum of the $\frac{T+R}{2}$ values of the loops involved, the effective loss of the trunk, and the two junction losses (which may turn out to be gains).

We see, then, that although detailed computations of the 1000 cps insertion loss provide some useful information in the design of exchange area systems, 1000 cps insertion loss alone is a poor criterion to use in judging the performance of a telephone connection. Even the computation of the insertion loss at a number of frequencies in the transmission band will not provide a complete picture of the circuit quality, since no information is obtained about the circuit noise, crosstalk, distortion in the transmitter and receiver - to mention just a few of the factors which determine circuit quality. The engineer, therefore, generally falls back on the effective loss method of evaluating the exchange area plant.

It is apparent from the discussion in the previous section that use of effective transmission loss is an engineering technique designed to fit all possible combinations of exchange area plant components, but with an accuracy which is only approximate. The technique is one which gives good results when applied to the sum total of plant, as in computing the plant performance, for example, to see if transmission standards for the area are being met. Some errors in the rating of individual circuit quality are to be expected, however. Furthermore, all of our "effective loss" values are comparisons to old instruments (circa 1928) which were limited bandwidth, high efficiency, instruments. This can lead to confused thinking. For example, modern instruments might be rated as 18 db better than the old reference instruments for some particular loop length, but careful interpretation of this 18 db advantage is essential. It is largely a consequence of the use of anti-sidetone circuitry and wider bandwidth - but the increased bandwidth is an advantage only if the signal can be heard at satisfactory volume. One cannot argue, for example, that use of modern instruments should permit 18 db higher trunk losses than old instruments did for the same grade of transmission, since under such conditions the received volume would be so low that the bandwidth advantage would be lost. It is evident that there is a great need for a method of actually measuring over-all circuit performance. The EATMS (Electro-Acoustic Transmission Measuring System) is a start towards such a measuring system.*

---

* See Chapter 2 for a discussion of the EATMS and its limitations.

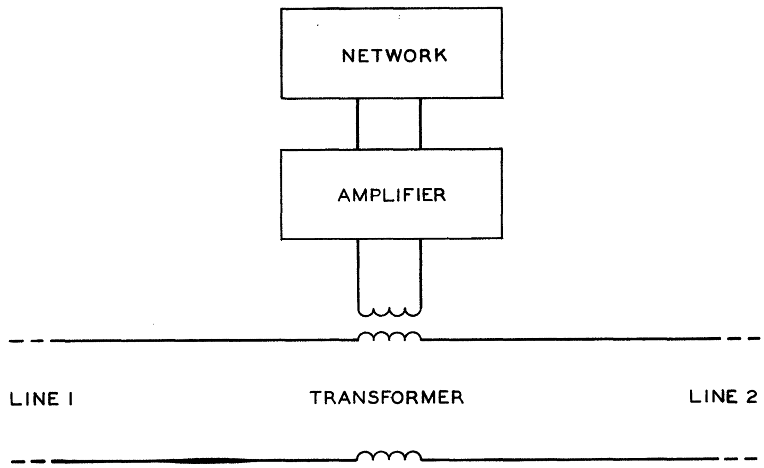
E-Type Repeaters

The preceding discussion has been based on achieving the tandem and toll trunk loss requirement by specifying wire size and loading arrangement. Until the advent of the E-type negative impedance repeater in 1948, voice frequency amplification was available only in the form of relatively expensive hybrid-type repeaters designed for voice frequency toll circuits. Besides being costly, these repeaters were not designed to transmit the dc signaling used in the exchange area. For these reasons, the toll circuit type voice frequency repeaters have been used only in a relatively few special applications in the exchange area plant. The new E-type repeaters provide an economical means of reducing the loss of inter-office trunks and at the same time permit normal signaling to be used. This permits improved plant performance, reduction in wire sizes, and, in general, increases the flexibility in the layout of the exchange area trunks.

A negative impedance repeater operates on the principle of inserting negative resistance (and, if desired, negative inductance or capacitance) in the line, thereby reducing the overall impedance and increasing the current in the line. This results in transmission gain in the same sense as that resulting from a repeater of the conventional type. In addition, repeaters of this type preserve the dc continuity of the circuit and are bilateral; i.e., provide amplification for both directions of transmission. They have the additional advantage that, in the event of a tube failure, transmission along the line is still possible, but at substantially higher loss, however.

Three basic E-type ("E" for "Exchange Area") negative impedance repeaters have been developed; these are designated the E1, E2, and E3. The E1 and E2, which are electrically identical, insert series negative impedance in the line. A block schematic of this type of repeater is shown in Figure 11. The E3 is designed to bridge a shunt negative impedance across the line. The E3 alone has had virtually no application in the plant. It is widely used, however, in conjunction with either the E1 or E2 repeater to form a combined series-shunt negative impedance repeater. Repeaters of this type are designated the E13 or E23 and are electrically identical. Figure 12 shows the E23 type repeater, in block schematic form.

The impedance inserted in the line is approximately proportional to, and the negative of, the impedance strapped in the adjustable network associated with each type of repeater. As in any

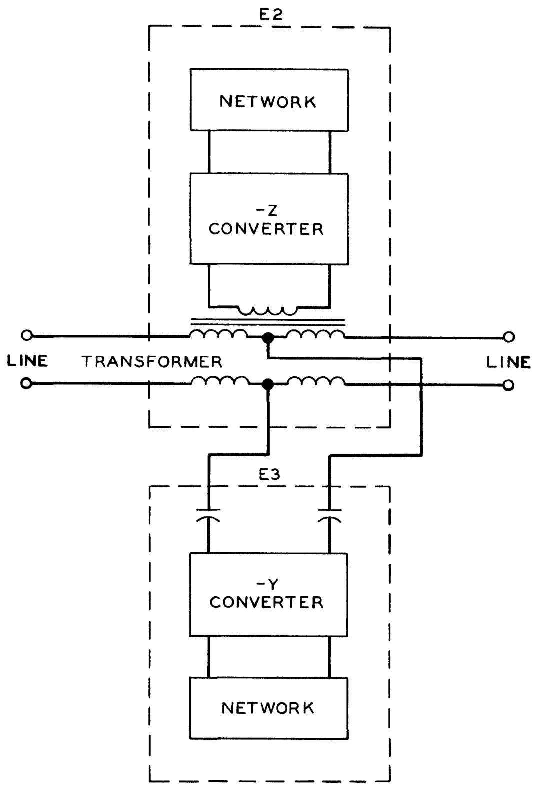


Block Schematic of E1 Repeater

Figure 3-11

repeatered circuit, the amount of negative impedance or "gain" inserted in the line is limited by considerations of such things as circuit singing (oscillation), noise, crosstalk into other systems, overload of the repeater, and echo loss requirements. In most inter-office trunks the necessity of preventing the repeater from singing is the gain controlling factor. The stability of the E-type repeater is dependent on the impedances connected to its terminals.* In a typical trunk installation, the impedances seen by the repeater depend upon the impedances of the circuits connected to each end of the trunk. The impedances across the repeater terminals vary, therefore, as the trunk is switched from circuit to circuit, or is left, for example, with its ends unterminated in the idle circuit condition. The repeater must be stable under all these conditions. Therefore, the worst trunk termination condition, for a particular type of repeater, will generally serve as the basis for setting the repeater gain. The E13 or E23 type repeater, as

* Basically, the series type repeaters are classified as open-circuit stable (short-circuit unstable) devices, while the shunt type are short-circuit stable (open-circuit unstable). The series-shunt type repeaters, therefore, can be either open- or short-circuit stable depending on whether the series or shunt element dominates. It follows, then that the repeater stability is dependent on the impedances connected to its terminals. However, this is just one illustration of the basic problem of repeater stability which exists in any repeatered transmission line, regardless of the type of repeater. A general discussion of this fundamental problem is given in "Some Fundamental Properties of Transmission Systems" by F. B. Llewellyn, which appeared in the March 1952 Proceedings of the I.R.E., (pages 271-83).



Block Schematic of E23 Repeater

Figure 3-12

normally used, is less sensitive to the circuit impedances and, as a result, in a given trunk will usually provide higher gain. The maximum gain which these repeaters can insert in a line is 8 to 10 db. As a rule of thumb, it is generally possible to use E-type repeaters to reduce the loss of a loaded trunk to about one-half its non-repeatered value, or the loss of a non-loaded trunk to about one-third of its non-repeatered value (where the losses are measured in db). Short trunks are usually operated with a single repeater placed either at or near the center of the trunk or at the terminal at either end. E-type repeaters can be worked in tandem on longer trunks. In this case, the loss of the repeatered trunk will approach the optimum given in the rule of thumb previously stated provided the trunk is uniform (i.e., one type of facility) and the repeaters are spaced the equivalent of about 8 db of line loss apart. The strapping network in each type of repeater can be adjusted to provide either shaped or flat gain across the voice frequency band. This permits a repeater to be used to compensate for attenuation distortion in the line, as well as providing net gain.

One disadvantage of the E1 and E2 type repeaters is that the insertion of a series negative impedance in an otherwise uniform line (loaded or non-loaded) introduces an impedance discontinuity which will appear as an echo at the sending end. The magnitude of this echo will be proportional to the magnitude of the negative impedance, or the gain, inserted by the repeater. This echo will not be a serious problem on trunks used for short inter-office connections where the delay is small. On toll connecting trunks used as part of a long toll circuit, however, the delays may be large. This frequently causes the echo requirement, rather than singing considerations, to limit the gain which can be inserted by the series-type repeaters in toll connecting trunks. This impedance discontinuity problem can be markedly reduced by using the E13 or E23 repeater on these trunks. By properly adjusting the E3 portion of the repeater, the impedance looking into the composite repeater can be made to approximate the trunk impedance, and, therefore, the echo which arises at the repeater is minimized.

In the trunk plant 19, 22, and 24 gauge conductors are normally used. For short distances the differential in cost between trunks of different gauges is not large. However, as the trunk length increases, smaller gauge wire with a repeater will be more economical than a larger gauge without a repeater. For example, a 10 mile trunk of 22 gauge wire with a repeater will be more economical than a 19 gauge pair with no repeater. In addition, these repeaters provide an economical means of improving the quality or grade of exchange area service.

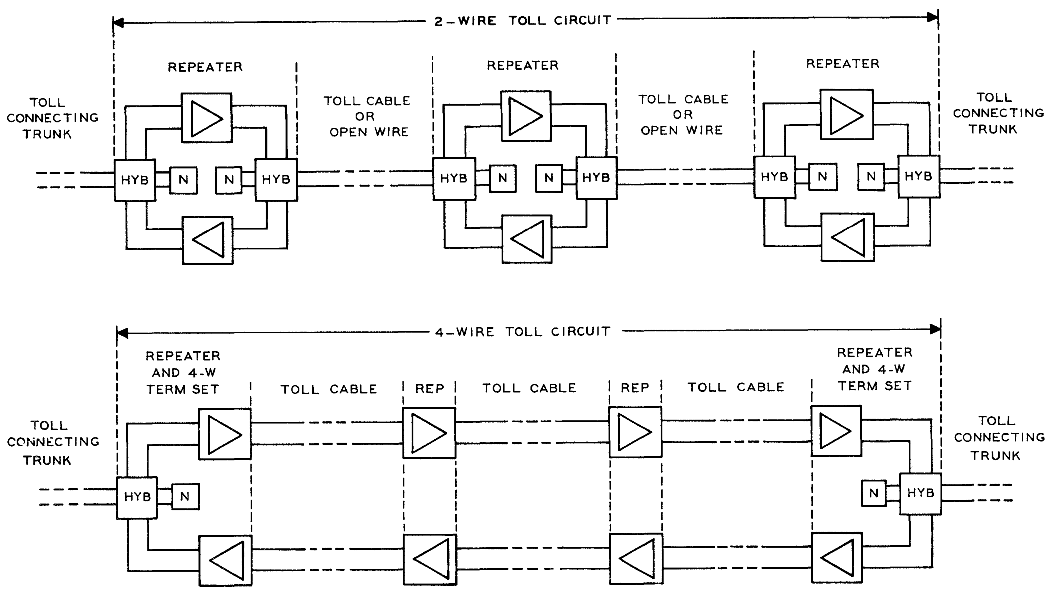
Voice Frequency Toll Systems

Before the development of AM carrier systems, telephone communication over long distances was possible only through voice frequency toll circuits operating on either open wire or cable. Today VF toll circuits are restricted to lightly loaded or secondary type routes having lengths of a few hundred miles or less. Both loading coils and hybrid-type voice frequency repeaters* are used on these circuits to reduce the transmission loss.

---

* We are using the word "repeater" here and throughout the text to denote the aggregate of equipment required to provide gain in a transmission path. Some transmission engineers, and some Bell System Practices, in discussing voice frequency or AM carrier systems, prefer to use terms referring to the component parts of what we are calling a repeater - namely "line amplifier" and "line equipment". In these cases, if the word "repeater" is used at all, it is generally used synonymously with "line amplifier".

Except for negative impedance repeaters such as the E-types, conventional amplifiers are one-way devices. To obtain amplification for both directions of transmission, additional circuitry is required. This may be done on either a two-wire or four-wire basis, as shown in Figure 13; both methods are used in the voice frequency plant. From the diagram it is evident that in each configuration a high order of balance in the hybrids is required to attenuate energy circulating from a repeater output to its input, by way of the hybrids and the opposite direction of transmission path. Failure to provide adequate attenuation will result in "singing" or oscillation of the circuit. The balance of the hybrids is determined, of course, by the degree to which the impedance seen looking into each intermediate section of line and the toll connecting trunks matches the impedance of the balancing network N connected to the opposite arm. Balance must be maintained not only over the voice frequency band but out-of-band as well so as to keep the loss of the circuit greater than the gain at all frequencies. In order to ease the out-of-band balancing problem, a low-pass filter is normally used at the input to each repeater amplifier to provide high loss above the voice frequency transmission band. The net loss around a sing-path is generally referred to as the "singing margin". Voice frequency circuits are usually engineered to provide a minimum singing margin of 10 db under normal conditions (e.g., temperature). Variations in net loss due to temperature changes, battery voltage variations, tube aging or other causes may result in the singing margin under less favorable conditions becoming as low as 4 db, which is about the minimum tolerable value.



2-Wire and 4-Wire Methods of Operation

Figure 3-13

The same paths which can give rise to singing also account for echo returned to the subscriber as he is talking. Echo, as discussed in Chapter 2, represents a distracting annoyance which in extreme cases can impair talking ability. The velocity of propagation at voice frequency, particularly on loaded cable, is relatively slow compared to radio or carrier systems (14,000 miles per second on 19 gauge cable with H88-50 loading, for example). Therefore, the maximum length of a voice frequency toll circuit, and the net loss to which it must be adjusted, are in some cases determined by echo requirements, as we shall see.

Figure 13 shows that in two-wire operation there are a multiplicity of singing (or echo) paths for each repeater. A given repeater is subject to instability due to its own loop transmission, its own plus that of the adjacent repeater, and so on until the end of the system. Thus, the tendency towards instability in two-wire systems increases as the length of the circuit and, therefore, the number of repeaters is increased. As a result, singing and echo performance are the controlling factors in setting the repeater gains in most long two-wire voice frequency toll circuits. In order to achieve reasonably high gains in the repeaters, careful attention must be paid to the installation and maintenance of the circuit, particularly with regard to the balancing networks on each hybrid. This disadvantage in the two-wire method of operation is off-set, in many cases, by the saving in copper afforded by using only one pair of conductors between repeaters.

Two-wire voice frequency toll systems are today operated on both open wire and cable. No loading is used on open wire, and the repeater spacing varies from up to 350 miles with 165-mil conductors to 150 miles for 104-mil conductors. The average length of these open wire systems is around 300 miles, with the maximum length systems extending about 700 miles. For two-wire cable operation, 19 gauge cable with H88-50 loading is standard. The higher attenuation of the cable as compared to open-wire necessitates a repeater spacing of about 50 miles. The difficulty of maintaining stability as the number of repeaters is increased limits the length of two-wire voice frequency cable circuits to 150 miles or less

Four-wire operation is restricted to cable, where a relative abundance of conductors is available. Figure 13 shows that in four-wire operation the major singing path is that which extends around the circuit, from one end to the other. This tends to make four-wire voice frequency systems more stable than two-wire. As a result, repeater gains in four-wire systems are limited more by modulation distortion*, echo, and crosstalk into other systems than by singing considerations. Four-wire circuits are generally operated on 19 gauge cable with H44-25 loading, and may be used over distances of 300 to 500 miles.

It will be recalled from the discussion of Chapter 2 that the performance to which a typical voice frequency toll circuit can be adjusted is neatly summarized by the "Via Net Loss Factor", abbreviated VNLF. The via net loss factor expresses, in db per mile, the minimum loss to which a system should be engineered; and takes into consideration the problems of noise, echo, singing, and crosstalk in a typical installation. The minimum via net loss for a system is found by multiplying the VNLF by the system length and adding a factor of 0.4 db to allow for maintenance variation. Via net loss factors for the various types of voice frequency toll circuits we have been discribing are tabulated in Table III below. Factors for carrier systems are also shown for comparison purposes.

Table III

Facility	VNLF db per mile
2-wire open wire (all wire sizes)	0.01
2-wire 19-H-88-50	0.03
4-wire 19-H-44-25	0.01
Open wire carrier (all types)	0.0017
Type K or N carrier (cable)	0.0019

It may not be possible to meet these minimum loss specifications in some installations, of course, particularly when adverse conditions lead to very poor echo performance. In general, however,

*Modulation distortion refers to the unwanted products formed by passing a signal through an amplifier, for example, which is never ideally linear. This subject is introduced in Chapter 4 and discussed in more detail in Chapters 8 and 12.

voice frequency toll circuits, of the type and average length discussed, will be engineered to have via net losses of 2 to 5 db. Obviously the use of carrier permits lower loss operation of trunks than VF does.

A number of types of voice frequency repeaters have been developed. Their characteristics are tabulated in Table IV below:

Table IV

Repeater Type	Circuit Description	Maximum 1000 cps Gain-db	Application
22	One stage triode amplifier, no feedback	19	2-wire; open wire or cable
44	Two stage triode amplifier, no feedback	43	4-wire; cable only
V1, V3*	One stage pentode amplifier, with feedback	35	2-wire; open wire or cable

The general practice in laying out a voice frequency toll circuit (as in all types of toll circuits) is to locate repeaters at or near the toll terminals of each end of the system. Additional repeaters, as required to meet the overall loss objective for the system, and to control noise and crosstalk, are located at intermediate points along the line, the spacing between repeaters being dependent on the type of facility as indicated in the previous discussion.

The amplifiers used in these repeaters have a gain-frequency characteristic which is essentially flat over the voice frequency band. Line loss, on the other hand, increases with frequency. Therefore, to reduce attenuation distortion, some equalization must be provided at each repeater to make the overall transmission through one section of line and a repeater flat (or as nearly flat as is practical) vs. frequency across the voice channel.** This is accomplished through the use of relatively simple adjustable networks ahead of the amplifier in each repeater.

---

* The V3 is the miniaturized version of the V1.

** Delay distortion (i.e., a phase characteristic which does not increase linearly with frequency) is not a problem in voice circuits.

Equalization compensates for the nominal characteristics of the line and repeater. In addition to equalization, regulation must also be provided to maintain the over-all circuit performance hour after hour as, for example, the line losses vary with temperature or moisture conditions. At voice frequencies, open-wire line loss variation is due principally to moisture changes at the insulators. The total variation, even on long circuits, is generally small, however. As an example, 300 miles of non-loaded 104-mil open wire is expected to show no more than ± 1.2 db variation in loss with typical changes in weather conditions. Therefore, two-wire voice frequency toll circuits operating on open-wire are regulated manually by changing the repeater gains, as dictated by periodic tests (sometimes as infrequently as twice a year) on the 1000 cps loss.

Variation in cable loss is due principally to temperature changes. For example, a change in temperature of ± 55°F is expected to cause ± 25 db variation in the loss of 500 miles of 19 gauge aerial cable with H44 loading. Such large variations in loss require automatic, rather than manual, regulation. Automatic regulation is provided in both two- and four-wire voice frequency cable circuits by pilot-wire regulators.* A pilot wire consists of a pair of conductors in the same cable as the toll circuits which are to be regulated. The dc resistance of this pair will, of course, vary with temperature just as the voice frequency attenuation of the cable does. The regulator is controlled by this dc resistance, and consists of relay circuits which select gain adjusting networks in the repeaters to compensate for cable attenuation changes. In practice, long toll circuits are broken up into regulating sections about 100 or 150 miles long, so that every second or third repeater is regulated. (The intervening repeaters remain at fixed gain.) At any such regulating point, a single pilot wire regulator can be used to control a number of voice frequency systems.

Long four-wire voice frequency circuits in which echo requirements are controlling can sometimes be operated at net losses lower than permitted by the echo limitation by equipping them with echo suppressors. As discussed in Chapter 2, however, echo suppressors introduce difficulties of their own, and their use is usually restricted to RC-RC trunks.

- -

* This is true in general. Some short two-wire cable circuits are manually regulated, however.

Interferences in Exchange Area Plant

The importance of maintaining low losses in the telephone plant has been pointed out. To achieve satisfactory service, it is also important to control and minimize interferences. In speaking of interferences, we include such items as noise, crosstalk, and power line induction. Some interferences, - for example, certain types of noise, - arise within a message channel: others, such as crosstalk and power line induction, are due to external influences. The characteristics, effects, and control of each of these is a study in itself; we shall be able only to touch briefly a few of the high points in the space available to us. We shall first discuss the various types of noise, then the mechanisms by which interferences are induced in telephone circuits, and conclude with a discussion of crosstalk.

Noise

Telephone intelligibility is affected by the total noise which reaches the listener's ears - particularly the ear to which the telephone receiver is applied. This total noise consists of room noise and circuit noise. Room noise on the listening subscriber's premises reaches his ear directly by leakage around the receiver cap and indirectly by way of the sidetone path through the transmitter and receiver of his subset. Room noise from the talking subscriber's premises also reaches the listener over the normal transmission path, and is assuming increasing importance. This is a result of two factors - first, circuit losses are being decreased; second, and more important, room noise has increased and its spectrum has shifted upwards in frequency in recent years, with increased use of such things as air-conditioners. This new spectrum, together with the improved low frequency response of the 302 and 500 type sets, has greatly increased the importance of room noise.

Circuit noise varies greatly. Its sources are many; for example:

Induced Power Hum - The inductive field of power lines may induce sizable signals in message circuits. It is the harmonics of the power frequency, particularly the third harmonic or 180 cps, which are of particular interest since they are frequencies at which the telephone set receiver is fairly sensitive. The use of balanced circuitry for voice frequency telephone connections, as discussed in a following section, is an important first step in reducing all such induced effects.

The metallic sheaths of telephone cables provide some protection against power line induction, so that power line noise may be low in cable circuits. In the case of open wire lines, however, the only step that can be taken (in addition to balancing and transpositions) is to provide physical separation between power and telephone lines by careful coordination of plant construction by the two industries. Where we have joint use of poles, or only roadway separation in rural areas, very high values of power line noise (45 dba) may be observed, however, and 26-28 dba is common.

Battery Noise - As batteries age, they develop resistance; associated with this is a "burning" type noise. However, this noise is normally negligible compared to the noise from the battery charging machinery (either rectifiers or generators). In addition, impulse-type noise from other circuits may be coupled to any particular circuit of interest by the common impedance formed by the battery and its filters. The resulting noise is predominantly low frequency in character. In crossbar and manual offices, large electrolytic capacitors are shunted across the batteries, giving very low noise values - from +2 dba maximum to -10 dba average.

Contact Noise - This type of noise results when currents pass through contacts which have variable resistance. Sporadic in character, it is most pronounced in panel-type exchanges, and can be very annoying. Improved maintenance of contacts is reducing its importance in the system as a whole.

Impulse Noise - Like power line hum, this is a type of noise that is usually induced in a telephone connection. (Impulse noise can be coupled into a circuit through a common impedance, such as the battery filter.) Unlike power hum, the sources of impulse noise are generally within the telephone plant. Impulse noise consists of rather short spikes of energy, to which the ear is fairly tolerant. The spectrum of impulse noise is relatively constant with frequency. However, since the transmission level at repeater input is frequently much lower in a carrier system as compared to a voice frequency system (i.e., carrier system repeater gains are generally much higher than voice frequency repeater gains) impulse noise introduced at a low level point will normally cause more impairment in a carrier system. Impulse noise arises from dial pulsing and other switching circuit transients in dial-switched central offices through which the connection passes. One of the ways of overriding it is to place repeaters somewhat closer to central offices than normal repeater spacing considerations would otherwise dictate, so that the signals will be at a sufficiently high level at the point of exposure to override the impulse noise. The exact characteristics of this type of noise are receiving increased attention lately, since data and PCM systems do not share the ear's tolerance to short bursts of energy.

Static Noise - Open wire lines are subject to static noise whenever a thunderstorm occurs within a distance of many miles; static may be considerable even when no thunderstorm is evident in the region. Extensive studies of storm static and thunderstorm incidence (TSI) as a function of geography have been made, and the engineering of open wire systems in a particular section takes the local TSI and local earth resistivity into account as important factors. A typical objective might be that static noise should not exceed 26 dba at zero level more than 1% of the time, which is 37 hours during the five month static season. Even under these circumstances, the nature of static incidence is such that for hours we might find noise 10 db higher than this value. The spectrum of static field intensity has its maximum value at 10 kc, and falls off, on the average, at 6 db per octave, but it is still important at carrier frequencies, since the pair balance decreases with frequency.

Thermal Noise - This is mentioned here merely for the sake of completeness; it is seldom important in voice frequency systems, though always present. It becomes extremely important in long carrier systems such as coaxial, submarine cable systems, and radio relay.

The diagram in Figure 7 shows the levels of speech and noise for a typical telephone connection. The circuit noise is somewhat less than average but is sufficient to be barely noticeable in the presence of average room noise.

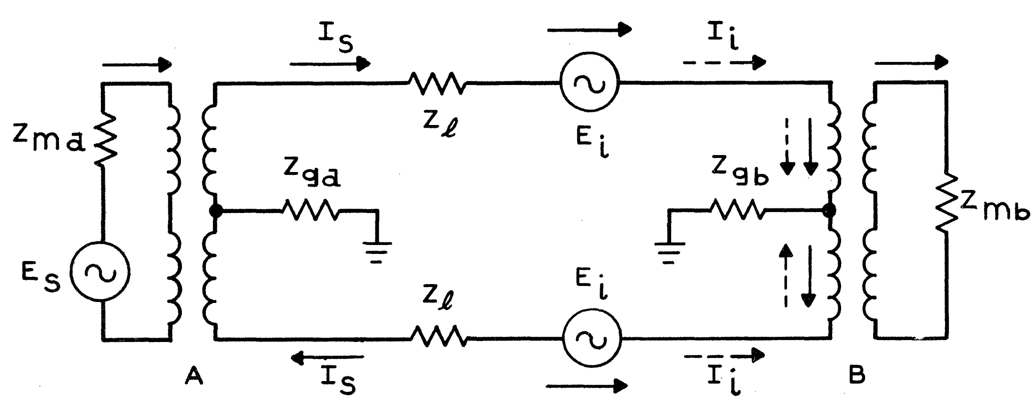
Induced Interferences - Mechanisms

As the preceding discussion has indicated, a transmission path can never be entirely isolated from the external world. It may parallel a power line, it usually lies in close proximity to other similar loops and trunks in cables or on pole lines, and it passes through central offices in which switching apparatus creates sizable transients. While not intended, coupling always exists to these sources of interference. In this section we shall discuss some of the mechanisms involved in this coupling, and the methods which have been used to reduce it.

Although ground-return circuits were used originally in the telephone plant, these were soon given up in favor of the "balanced" pair. The balanced pair has the advantage that interferences induced equally in both wires of the pair are balanced out. This is shown in Figure 14. Repeat coils are used at either end of the circuit. The signal voltage E_s causes signal current I_s, to flow. The current direction in the two wires is opposite, and thus the signal passes through the B repeat coil to Z_{mb}. An interference which is induced equally in each wire is indicated by the E_i generators. In this case the currents flow in the same direction along the pair and cancel in the repeat coil. The currents which flow in opposite directions in the wires of the pair are known as "metallic circuit currents". The currents which flow in the same direction along the pair are called "longitudinal currents".

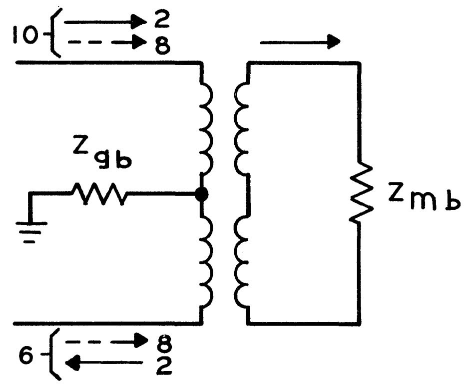
Induced interferences may reach the receiver and disturb transmission in two ways. In one case the coupling between the source and one of the wires of the pair differs from the coupling between the source and the other wire, and in the other there is an unbalance with respect to ground within the pair.

If the voltages induced in the two wires of the pair are not equal, then the interference will appear as both a longitudinal and a metallic circuit current. Consider the case shown in Figure 15 where it is assumed that unequal voltages are induced as a result of differences in exposure to the disturbing source. These voltages may be analyzed into a pair of voltages which would cause balanced longitudinal currents, plus a residual which would cause a metallic current. The circuit presents quite different impedances to these two types of



Signals and Interferences in Balanced Pair

Figure 3-14



Metallic and Longitudinal Currents

Figure 3-15

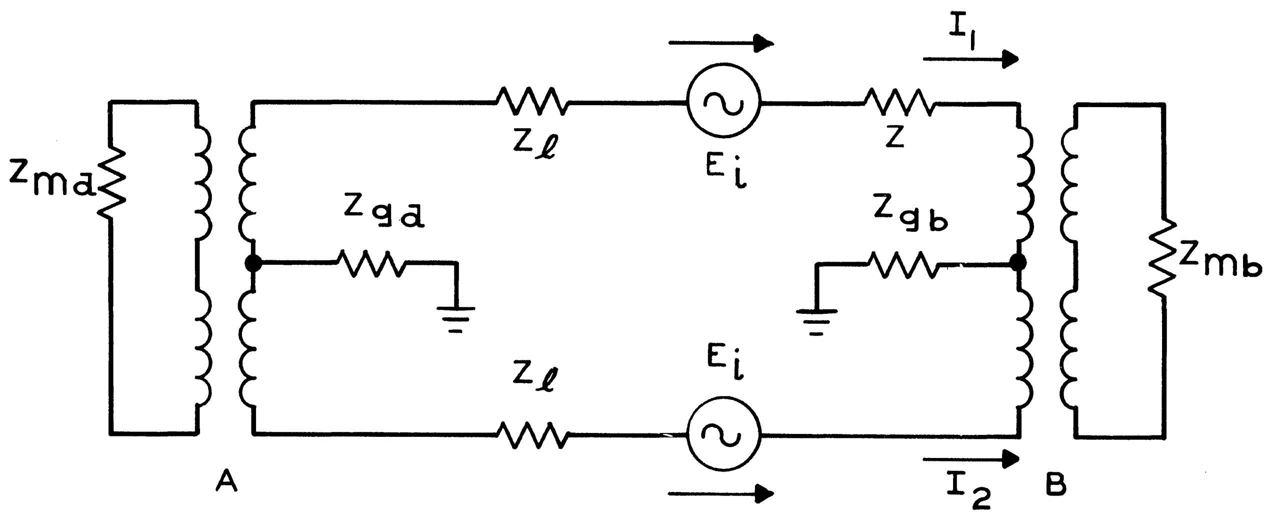
generators. In Figure 15, the total resulting current in the upper wire is shown as 10 units and that in the lower wire as 6 units. As shown by the solid and dashed arrows, these induced currents can be divided into metallic circuit currents of 2 units and longitudinal currents of 8 units. The 2 units of metallic circuit current will, of course, appear as an interference in Z_{mb}.*

In the second case, the voltages induced in the two wires are identical, but there is an unbalance in the pair. This unbalance may result from differences in the resistance of the two windings of the repeat coil or from a poor splice in one of the wires. The effect of this series unbalance can be seen by considering Figure 16. The impedance, Z, represents an added resistance in the upper wire. Because of this additional resistance the currents I_1 and I_2 will not be equal. The unequal currents may be analyzed as in Figure 15 to show that metallic circuit currents are produced.

Consider now that the circuit of Figure 15 or 16 is one of two pairs being used to provide a phantom circuit. The message currents of the phantom circuit appear in either pair as longitudinal currents. It follows that the induced longitudinal currents in the two pairs must be equal in amplitude and phase if the phantom is to be free of induced interference. Phantom circuits are, therefore, particularly susceptible to this sort of interference, since getting balance between two pairs is more difficult than for the two wires of a single pair.

---

*A 4 to 1 ratio of longitudinal to metallic current represents a totally unsatisfactory circuit. Ratios of 500 to 1 or 1000 to 1 would be more representative of plant performance.



Effect of Unbalance Within a Pair

Figure 3-16

The coupling between a source of interference and the wires of a pair may be capacitive, inductive, conductive, or various combinations of these. Coupling to power lines is usually predominantly inductive, and would be of trivial importance if it were not for the great energy of the disturbing source. The coupling in cables and open wire lines which produces crosstalk at voice frequencies is predominantly capacitive, although inductive coupling becomes important as we approach carrier frequencies.*

It might be pointed out that raising the impedance level of a circuit (the impedance from each wire of a pair to ground, say) will cause it to absorb more power from a source to which it is coupled by a very high impedance path (e.g., capacitive coupling). The source and coupling impedance act like a constant current generator. Loading causes such an increase in circuit impedance, so that the wider use of loading in the exchange area plant has the consequence of greater vulnerability to induced interferences - including crosstalk.

Methods of Reducing Induced Interference

The above brief discussion of mechanisms is sufficient to suggest how induced interferences may be minimized. It is clearly advantageous to maintain balance within the pair itself - good splices will help to avoid the situation shown in Figure 16. Similarly, the apparatus to which the pair connects, such as repeat coils, should be well balanced; a common criterion is that the balance of such apparatus should be 10 db better than the best pair with which it will be used. Another obvious method, to be used when practicable, is to reduce the coupling between line and source by adequate spacing. This is done in

---

* Another type of coupling, not germane to this discussion, will be considered when we come to carrier systems, - namely, intermodulation, or unwanted modulation products.

the case of power line induction, as already mentioned; it is one of the reasons for keeping a reasonable spacing between pairs of open wire lines, thus reducing the crosstalk between them.

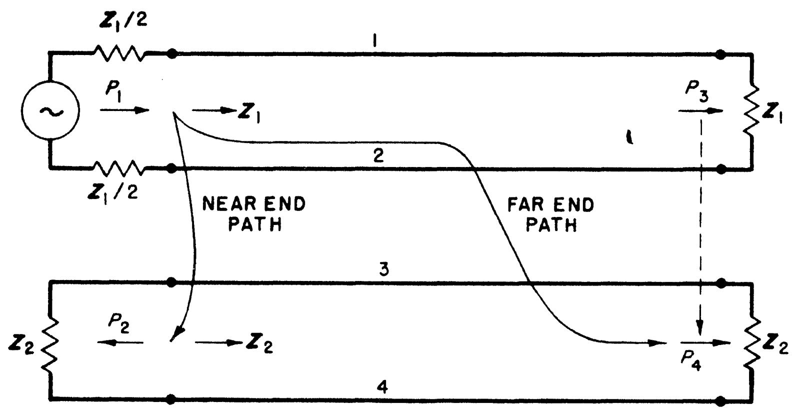
In addition to maintaining balance within the pair itself, it is important, as the preceding discussion has shown, to expose each wire of the pair equally to the interfering source. In open wire lines this is done by complicated transpositions systems. Two ideas are involved here. In the first place, all the lines must be transposed relative to adjacent power lines; since long wave lengths are involved here, this can be accomplished by relatively infrequent transpositions. Secondly, however, the various telephone lines must be transposed relative to each other; in this case, if carrier frequencies are involved, the wave lengths are shorter, and transpositions must be more closely spaced. (Transposing the lines relative to each other may, in some cases, degrade the balance relative to the power line, unfortunately.) By careful design it has been possible to arrive at transposition patterns which permit the use of open wire up to frequencies of the order of 140 kc with transpositions as far as two to four poles apart. In cables, the wires of a pair are twisted around each other, and the various pairs are further twisted relative to each other at various rates (i.e., various twist lengths), again in order to achieve as much balance and randomization of coupling paths as possible.

In repeatered systems, a satisfactory over-all signal-to-noise ratio can be obtained by judiciously spacing repeaters so that signal levels never become so low as to be comparable to the induced noise. (The same approach is used in systems where thermal, rather than induced, noise is controlling, as discussed in detail in subsequent chapters.) On the other hand, the signal levels must not be set so high as to make the repeatered circuit a source of excessive crosstalk.

For the sake of completeness, although they are not economically applicable to voice frequency telephone circuits, two other methods might be mentioned. One is shielding, as exemplified by shielded video pairs and by coaxial.* The other is "crosstalk balancing" - the deliberate introduction of coupling paths between pairs of a cable, the magnitude and phase of the coupling being chosen to balance out the already existant unwanted coupling. The use of this method is pretty much confined to K carrier, which will be discussed in Chapter 5;

---

* The use of two cables to carry the opposite directions of transmission in a section of 4-wire carrier, as in "K", might be considered as an example of shielding, or of space diversity.



Near-End and Far-End Crosstalk

Figure 3-17

coupled inductances, and small capacities (40-200 mmf) are used to obtain this adjustable coupling.

Crosstalk

Crosstalk occurs when signals on one telephone circuit appear on another circuit as an interference. The circuit which is the source of the signals is known as the disturbing circuit, and that on which the signals are heard as the disturbed circuit. On voice frequency message circuits the crosstalk is inherently intelligible. This is contrasted to a carrier system, for example, in which the crosstalk frequently consists of speech which has been inverted or otherwise shifted in frequency so as to be unintelligible. Intelligible crosstalk is particularly objectionable because of the real or fancied loss of secrecy which it implies.

While all crosstalk to be considered here is caused by the sort of unwanted coupling that we have just been considering, the subject is of such importance that we go on to further distinguish between various types of crosstalk paths. One such classification is illustrated in Figure 17, which is mainly concerned with distinguishing between where the crosstalk is measured. It might be mentioned, however, that unless all the factors in the situation are known, including the geographical distribution of the coupling, the value of far-end crosstalk cannot be predicted from a knowledge of the near-end crosstalk even in a non-repeatered circuit.

Crosstalk may also be classified according to whether or not a third circuit is involved in the coupling, and how, leading to the following four definitions:

<u>Direct Crosstalk</u> is the summation of the crosstalk induced by the disturbing circuit directly in the disturbed circuit without involving any tertiary (i.e., third or intermediate) circuit.

<u>Indirect Crosstalk</u> is the sum total of all the crosstalk which reaches the disturbed circuit by two paths in tandem, one from the disturbing circuit to tertiary circuits and the second from the tertiary circuits to the disturbed circuit.

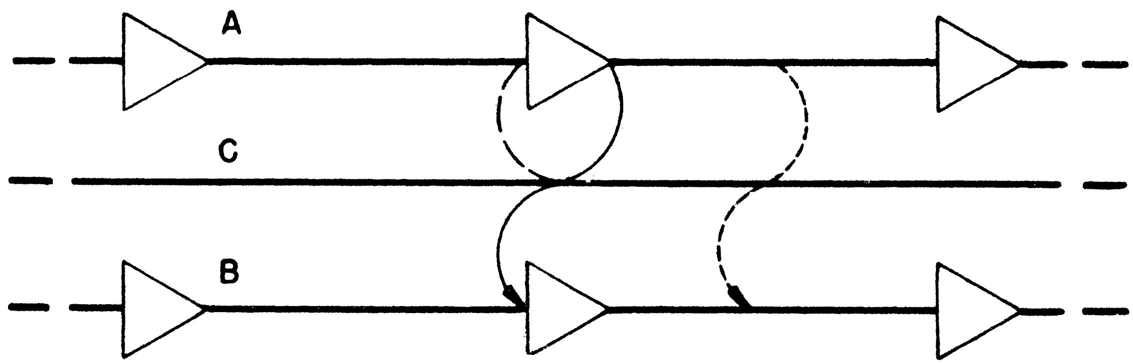
<u>Transverse Crosstalk</u> is the sum total of the direct and indirect crosstalk. It does not include that portion of the crosstalk which involves transmission in tertiary circuits.

<u>Interaction Crosstalk</u> refers to all of the indirect crosstalk which involves tertiary transmission between transverse sections. It is thus the result of direct crosstalk from disturbing circuit to tertiary, which produces currents and voltages in the tertiary circuits; these, in turn, produce direct crosstalk into the disturbed circuit.

Interaction crosstalk may be serious in repeatered circuits, as indicated in Figure 18. Circuits A and B represent two one-way circuits transmitting in the same direction; C is a circuit which is not repeatered at the points of interest to us here. The solid line connecting A and B indicates an interaction crosstalk path between the output of a repeater in A and the input of a repeater in B. This path involves the sum of (1) near-end crosstalk loss between A and C in the second repeater section and (2) the near-end crosstalk loss between C and B in the first repeater section. The dotted line connecting A and B indicates a similar path at the center of a repeater section; here the repeater gain would not be involved, so the dotted would be less serious than the solid line. The dashed line represents a path around a repeater itself by way of the tertiary circuit; if repeater gains are high enough (as in carrier systems) such a path may cause severe distortion of the repeater insertion gain characteristic.

<u>Effects of Transmission Levels</u>

The crosstalk loss which would be measured between two circuits at their terminals due to a known crosstalk coupling in a particular



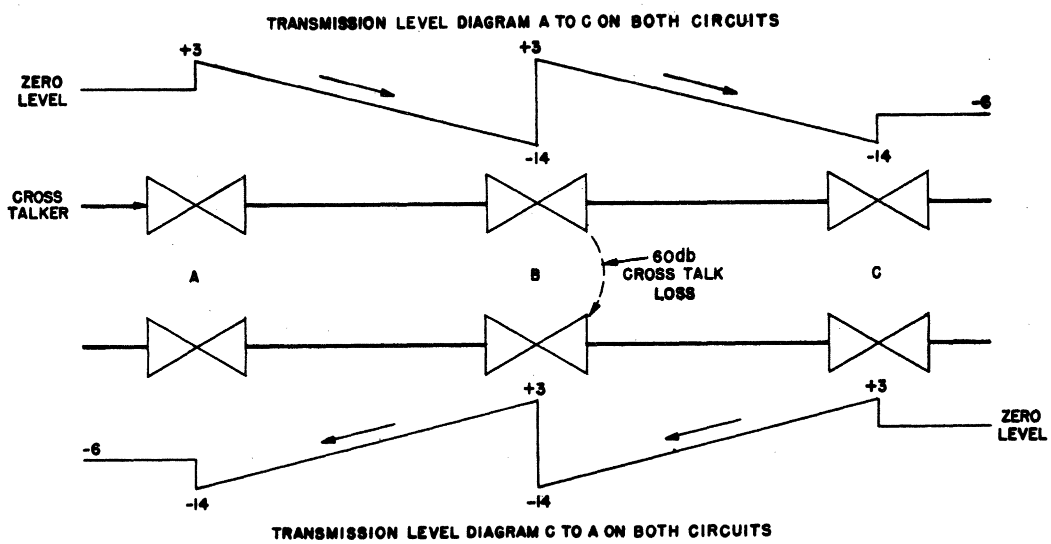
Amplification of Interaction Crosstalk

Figure 3-18

section of line or piece of apparatus, is, in general, a function of the relative transmission levels of the circuits at the point where the coupling exists and at the terminals where the measurements are made. When the circuits contain gains (amplifiers or repeaters), as do many voice-frequency, carrier, and video circuits in exchange area and toll plant, the level differences between circuits, and hence the effects of level differences, may be greatly increased.

By way of example, consider the near-end and far-end crosstalk between the two circuits shown in Figure 19, which are typical of short 2-wire voice frequency toll cable circuits. The symbols at A, B, and C are conventional representations of 2-way telephone repeaters which have gain for both directions of transmission. The transmission level diagrams show, for the two directions of transmission, the levels of the speech at each point in decibels above the level at the sending end. Since the circuits are assumed to be alike, the same level diagrams apply to both circuits. The gains or losses between any two points are easily obtained from the level diagrams by subtracting the levels at the two points.

Now assume that we are interested in the near-end crosstalk at the A terminals of the upper and lower circuits, due to the crosstalk path shown at B having a loss of 60 decibels. This crosstalk path might be a lumped coupling at the point indicated, or it might represent the near-end crosstalk measured at B on the section of line between B and C due to crosstalk coupling distributed along that section.



Crosstalk on Repeatered Circuits

Figure 3-19

The crosstalk loss between the two circuits at A is evidently equal to the sum of the loss in the upper circuit from its terminal at A to the point of coupling, plus the 60 decibel loss between the circuits, plus the loss of the lower circuit from the coupling to its terminal at A. From the level diagrams there is evidently a gain of 3 decibels (or a loss of -3 decibels) from A to the coupling, and a gain of 14-6 or 8 decibels (or a loss of -8 decibels) on the lower circuit from the coupling to its A terminal. Therefore, the near-end crosstalk loss between the circuits at A is -3 + 60 -14 + 6 or 49 decibels. Evidently, a crosstalk loss of 60 decibels at B appears as a crosstalk loss of 49 decibels when measured from A. The apparent gain of 11 decibels is called the crosstalk amplification. If there is a known crosstalk loss between two circuits in a particular section or piece of equipment, its importance cannot be judged without knowing the cross-talk amplification. Different crosstalk couplings along a circuit must be reduced to a common base by correcting for the crosstalk amplification before they can be compared or combined.

If we were to increase the gain of the terminal repeater of the disturbed circuit by six db, the output of this circuit would be a zero level point, and the near-end crosstalk loss would be 49 - 6 or 43 db. This would be the equal-level crosstalk loss; i.e., the coupling that would be measured between equal level points on the disturbing and disturbed circuits.

Coupling at various points in the circuits may be reduced to a common basis by computing the equal level crosstalk loss for each. For an exact computation of the total coupling between the two circuits, it would be necessary to have information on the phase relationships involved, but fortunately it turns out that this is not generally necessary in the practical case. Usually random addition can be assumed in finding the total rms value of coupling, and it can be taken as a good rule of thumb that, for toll cables at carrier frequencies, the 1% value of maximum coupling will be about 7 or 8 db above the rms value. For voice frequency exchange area cable the 1% value of maximum coupling is about 12 db above the rms value.

Measurements and Units

The magnitude of the crosstalk coupling between two circuits is fundamentally a matter of how well speech energy is transferred from one to the other; a single frequency measurement does not tell the story unless the coupling is flat vs frequency. This condition is often satisfied over the bandwidth of any particular channel in carrier circuits, but at voice frequencies, where capacitive coupling is usually the dominant mechanism, the coupling loss has an average slope of 6 db per octave. Furthermore, the discontinuities caused by splices and different gauges of wire cause the coupling loss to vary sharply around this average slope, the deviations being as great as 6-8 db. For the laboratory measurements made during investigations into basic requirements, actual speech and listeners are used; the criterion of intelligibility is taken as the ability to understand four words during a seven second time interval. For field measurements of voice frequency circuits, thermal noise, shaped to have the same power spectrum as speech, is often used as the input to the disturbing circuit. The output of the disturbed circuit can then be measured using a 2B noise meter with F1A weighting. (A zero vu speech volume applied to a noise meter thus weighted gives a reading of about 82 dba.) For smooth 6 db per octave capacitive coupling, this method gives a reading which differs by 2.8 db from the 1000 cycle coupling, the 1000 cps coupling loss being greater than the noise coupling loss. If the coupling loss had a smooth 6 db per octave slope, 1000 cps measurements could be made and corrected. The noise measurement protects us against errors which would be involved in this slope assumption, and gives a single integrated value for the jagged curve of coupling loss vs frequency.

A unit often used in crosstalk computations is the dbx - which was invented merely to permit coupling paths to be expressed in positive numbers. It is equal to the difference between 90 db loss and the

SECTION AB61.010

RELATION BETWEEN COUPLING, NO. OF DISTURBERS
AND CROSSTALK INDEX

1. This table is based on low values of noise, that is, below 19 dba of induced noise or below 11 dba of tube or thermal noise. For the usual conditions, where the noise is greater, the couplings must be corrected in accordance with the Noise Correction table below.

2. For unintelligible crosstalk, subtract, in addition, "Staggering Advantage" before entering chart.

3. Where rms coupling only is known for less than 10 disturbers, read part 3-D before using chart.

NOISE CORRECTION

Induced Noise - dba	Tube or Thermal Noise - dba	Correction - db (Reduce Coupling - dbx by this Amount)
19 - 20	11 - 12	1
21 - 22	13 - 14	2
23 - 24	15 - 16	3
25 - 26	17 - 18	4
27 - 28	19 - 20	5
29 - 30	21 - 22	6
31 - 32	23 - 24	7
33 - 34	25 - 26	8

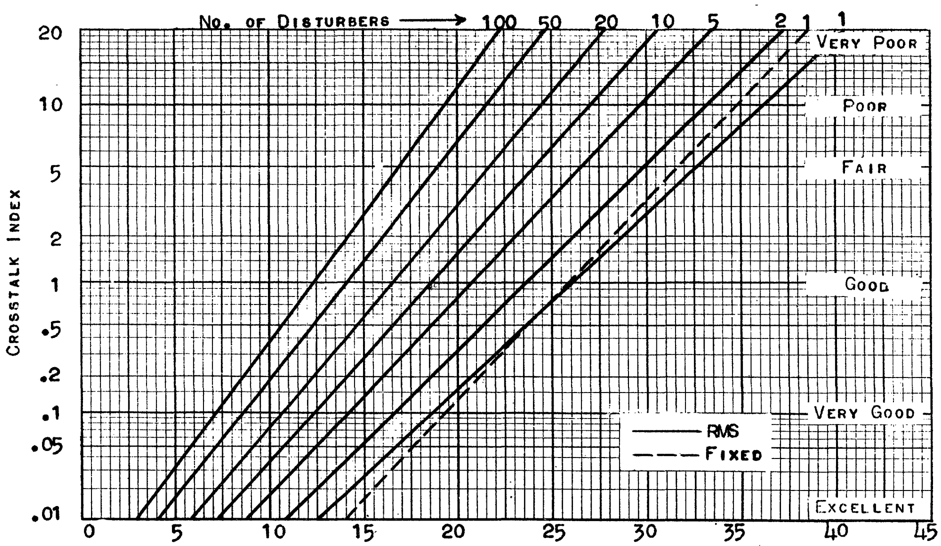


Fig. 3

Chart for Rating Trunk Crosstalk

Figure 3-20

transmission of the coupling path: thus if the coupling between two circuits is -60 db, we express this fact by saying that the coupling is 30 dbx. The smaller the number of dbx coupling between the circuits, the better.

Performance vs Requirements

When the coupling between circuits has been measured, or the expected coupling computed (sometimes from measurements on short sections of cable or short systems), we have the problem of deciding whether or not it represents acceptable performance. This is an extremely difficult question to answer. The economic penalties of too stringent a requirement are great; the service consequences of too lenient a requirement could be painful in terms of customer dissatisfaction. The discussion of objectives in Chapter 2 lists the complex factors involved in arriving at a relationship between the loss of the coupling path and the chance of intelligible crosstalk.

For trunks, at present, we use the curves of Figure 3 of BSP 61.010, given here as Figure 20.* Obviously there is considerable room for judgement even if we take this chart as gospel - for example, shall we insist on "good" performance, or be content with "fair", in a given instance? It has been stated that a 1% chance of intelligible crosstalk (1% index, on this chart) is a system objective, but this is not an ironclad rule.

If we look back of the chart, and ask how it was obtained, we find some experts who are critical of the mathematics used in combining all the probability functions involved, and others who question more fundamental assumptions. For example, should we predicate our requirements on observers who are striving to hear the crosstalk, and have their receivers tightly sealed to their ears? They hardly represent normal subscribers - but it is such observers whose judgements form the basis of the values given on the chart.

For the trunk plant, these matters are being re-examined. When we consider the loop plant, we find the situation even less clear. One point should be stated categorically: the criteria for loops should be different than for trunks, and the chart of BSP 61.010 should not be

---

*Before entering this chart, dbx should be corrected by -5 db to reflect changes in talker volume and a re-estimation of the masking effect of noise, according to a BSP addendum of October 1953. Other corrections apply when either the disturbing or disturbed circuits are compandored; the effects of compandors are discussed in a subsequent chapter of this text.

used for the loop plant. A loop is permanently assigned to one subscriber, who may be a very loud talker; the trunk plant is used by many subscribers, and probability enters to protect us against bull talkers. The activity factors for the two types of plant are very different - 2% for loops as compared to 17% for four-wire toll trunks and as much as 70% for two-wire voice frequency inter-office trunks. Where 30 dbx may be fairly satisfactory for a trunk, 15 dbx may be quite poor for loop. At the moment, there are, unfortunately, no good guides to the engineering of loops with respect to intelligible crosstalk.

Effects of Plant Developments

Changes in the plant affect the crosstalk problem as they do all our problems. Mention has been made, for example, of the fact that the increased use of subscriber loop loading (increasing line impedances) has made the crosstalk problem more severe. The increased efficiency of the 500 set has had the same effect, since the listener has a more efficient receiver with which to hear the crosstalk, the disturbing talker a more efficient transmitter with which to put a high signal on the line. A similar increase in the chance of intelligible crosstalk results from lower net loss operation of trunks. The reduction of noise - as a result of improved battery filtration and contact maintenance, to mention only two ways in which noise has been reduced - is unfavorable, since it decreases the masking. It seems that every improvement we make in the plant makes the crosstalk problem tougher.

There are, however, some factors on the favorable side. The improvements in transmission result in some compensating lowering of volumes - people don't have to shout quite so often, and they slowly find that out. The use of negative impedance repeaters lowers the line impedance, which is favorable. The use of 500 sets with their equalizers reduces the spread in volumes, which is favorable because it is the tails of the various probability distributions which are most important in the crosstalk problem. A similar effect in decreasing the standard deviation may be expected from the use of new low current subsets which are being developed for line concentrators and ESS. Line concentrators will ease the engineering of the loop plant, since they essentially turn loops into trunks, giving us some probability protection.

Conclusion

Two basic points should be apparent from this survey of typical methods and basic problems encountered in handling voice frequency signals.

1. The wire plant used for exchange area and short-haul toll transmission is composed of a vast variety of wire sizes, cable and open wire facilities, and repeater and loading arrangements. These facilities have grown up through the years based on engineering studies made to find the most economical solution to specific situations.

2. The losses, impedance discontinuities, and interferences have been engineered and adjusted to be more or less adequate for voice frequency transmission.

During the coming years there will be increasing efforts made to adapt this extremely valuable plant to provide improved transmission and new types of service. The performance, variability and limitations of this existing plant will strongly affect the design and application of new systems such as PCM carrier, data transmission, and electronic switching.

Chapter 4

AMPLITUDE MODULATION

An amplitude modulated signal can be obtained by passing both a carrier and the modulating signal through a non-linear device. A number of forms of AM signals - double sideband, single sideband, twin sideband, and vestigial sideband - have found application in various carrier systems.

Non-linearity, which is capitalized on to form and detect the AM signal, also gives rise to unwanted products at the output of system repeaters, terminal amplifiers and other devices which are never ideally linear. These products constitute noise and crosstalk which add to the other noise sources in the system and degrade the signal-to-noise ratio.

Introduction

Modulation is defined as the process or the result of the process whereby some characteristic of one wave is varied in accordance with another wave.* As an illustration, consider a wave represented analytically by the expression

$$e = A \cos \theta\,(t) \tag{4-1}$$

This expression will represent a modulated wave if, for example, either A or $\theta(t)$ is made to vary as a function of the modulating wave. Thus, in amplitude modulation (AM), which is the oldest and perhaps most familiar form of modulation, the modulating wave is used to vary the amplitude coefficient A. The properties of the AM signal and its applications are discussed in some detail in this chapter. Another form of modulation, known as angle modulation, occurs when the modulating wave is used to vary $\theta(t)$. Two important types of angle modulation are phase modulation (PM) and frequency modulation (FM), and are discussed in Chapter 19. Both amplitude and angle modulation are widely used throughout the telephone plant in the various carrier and radio systems. It is expected that in the future a form of phase modulation known as pulse code modulation (PCM) will find increasing areas of application

---

* "IRE Standards on Modulation Terms", Proc. IRE, May 1953, pages 612-615.

in the plant.* This modulation scheme is described in Chapter 25.

In general, the modulating wave may be either the electrical signal representing a message, such as a subscriber's conversation or a television picture, or it may be an information bearing signal obtained from a previous step of modulation. The wave which this signal modulates is often referred to as the carrier wave, or simply as the carrier.

The inverse of modulation, known as demodulation, is defined as the process of recovering the modulating wave from a modulated carrier. It should be pointed out that in the broadest sense the term modulation can encompass the process of demodulation, and some modern texts often use the word modulation when referring to either operation.** However, for clarity it is convenient to retain the word demodulation.

Modulation is useful when we want to translate an information bearing wave into a signal which is suitable for transmission over a particular medium. It will be demonstrated in this chapter that modulation occurs whenever a signal is sent through a circuit containing a non-linear element, such as an electron tube, varistor, or even an iron-cored inductance coil. In many cases, therefore, the modulation is unwanted, and considerable engineering effort is spent in suppressing it. In modulators and demodulators, however, the non-linearity is capitalized upon, and the modulation products are utilized.

The Nature of the AM Signal

As stated in the previous section, amplitude modulation is that process by which the amplitude of a carrier is made to vary as a function of a modulating wave. The discussion here will be restricted to the case of a sinusoidal-type carrier, which is in accord with normal carrier systems practice. Let this sinusoidal carrier be defined as

$$C(t) = A_c \cos \omega_c t \qquad (4\text{-}2)$$

in which A_c = amplitude of unmodulated carrier

$f_c = \omega_c / 2\pi$ = carrier frequency

---

* In PCM, however, Equation 1 does not apply, and a more generalized expression is necessary.

**Similarly, the word encoding, introduced in pulse systems, is basically synonymous with modulation and has been used by some writers to include the process of decoding, which is synonymous with demodulation.

For simplicity of analysis we will first examine the case when the modulating wave is a single sinusoid, defined by

$$m(t) = A_m \cos \omega_m t \qquad (4\text{-}3)$$

in which A_m = modulating signal amplitude

$f_m = \omega_m/2\pi$ = modulating signal frequency

It will now be demonstrated that if m(t) and C(t) are simultaneously applied to a non-linear device, amplitude modulation will occur. Let the characteristic of the device be described by the power series

$$E_{out} = a_o + a_1 E_{in} + a_2 E_{in}^2 + a_3 E_{in}^3 + \dots + a_n E_{in}^n \qquad (4\text{-}4)$$

where, for any particular operating point on the characteristic, the coefficients a_i are independent of the amplitude and frequency of E_{in}.

The power series characteristic is chosen because many non-linear devices closely follow such a law. In general, only the first few terms will be important, since the higher order terms will either be negligible or the resulting outputs can be eliminated. For purposes of this analysis it is assumed that terms higher than the third order are negligible.

Let the input to the non-linear device be given by

$$E_{in} = C(t) + m(t) \qquad (4\text{-}5)$$

Substituting Equations 2 and 3 for C(t) and m(t) respectively gives

$$E_{in} = A_c \cos \omega_c t + A_m \cos \omega_m t \qquad (4\text{-}6)$$

Equation 4 then shows the output voltage to be

$$\begin{aligned} E_{out} = a_o &+ a_1 (A_c \cos \omega_c t + A_m \cos \omega_m t) \\ &+ a_2 (A_c^2 \cos^2 \omega_c t + A_m^2 \cos^2 \omega_m t \\ &\quad + 2A_m A_c \cos \omega_c t \cos \omega_m t) \\ &+ a_3 (A_c^3 \cos^3 \omega_c t + A_m^3 \cos^3 \omega_m t \\ &\quad + 3A_c^2 A_m \cos^2 \omega_c t \cos \omega_m t \\ &\quad + 3A_c A_m^2 \cos \omega_c t \cos^2 \omega_m t) \end{aligned} \qquad (4\text{-}7)$$

Carrier Input = $C(t) = A_c \cos \omega_c t$ Modulating Signal = $m(t) = A_m \cos \omega_m t$

Combined Input = $E_{in} = A_c \cos \omega_c t + A_m \cos \omega_m t$

$$\text{Combined Output} = E_{out} = a_o + a_1 E_{in} + a_2 E_{in}^2 + a_3 E_{in}^3$$

$$= a_o + a_1 (A_c \cos \omega_c t + A_m \cos \omega_m t)$$

$$+ a_2 (A_c^2 \cos^2 \omega_c t + A_m^2 \cos^2 \omega_m t + 2A_m A_c \cos \omega_c t \cos \omega_m t)$$

$$+ a_3 (A_c^3 \cos^3 \omega_c t + A_m^3 \cos^3 \omega_m t + 3 A_c^2 A_m \cos^2 \omega_c t \cos \omega_m t + 3 A_c A_m^2 \cos \omega_c t \cos^2 \omega_m t)$$

The table shows the various frequency components and their relative amplitudes.

Term	Zero Order (dc)	First Order	Second Order	Third Order
0	a_o			
1		$a_1 [A_c \cos \omega_c t + A_m \cos \omega_m t]$		
2	$1/2\, a_2[A_c^2 + A_m^2]$		$1/2\, a_2[A_c^2 \cos 2\omega_c t + A_m^2 \cos 2\omega_m t]$ $+ a_2 A_m A_c[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$	
3		$3/4\, a_3 A_c[A_c^2 + 2A_m^2] \cos \omega_c t$ $+ 3/4\, a_3 A_m[2A_c^2 + A_m^2] \cos \omega_m t$		$1/4\, a_3\left\{A_c^3 \cos 3\omega_c t + A_m^3 \cos 3\omega_m t\right\}$ $+ 3/4\, a_3\left\{A_c^2 A_m[\cos(2\omega_c + \omega_m)t + \cos(2\omega_c - \omega_m)t] + A_c A_m^2[\cos(\omega_c + 2\omega_m)t + \cos(\omega_c - 2\omega_m)]\right\}$

Trigonometric identities used in the expansion of E_o:

$\cos^2 x = 1/2\, (1 + \cos 2x)$

$\cos x \cos y = 1/2\, [\cos (x+y) + \cos (x-y)]$

$\cos^3 x = 1/4\, (\cos 3x + 3 \cos x)$

$\cos x \cos^2 y = 1/2 \cos x\, (1 + \cos 2y) = 1/2 \cos x + 1/4\, [\cos (x + 2y) + \cos (x - 2y)]$

Output of Power-Series Law Modulator

Figure 4-1

If use is now made of some trigonometric identities (which are listed in Figure 1), Equation 7 can be expanded into a number of terms which can be collected and tabulated as shown in Figure 1.

The tabulation in Figure 1 shows that the output of the non-linear device includes dc and linear terms as well as terms involving the squares, cubes, and cross-products of the original signals. Out of this assortment, the terms which are of interest are those in which a), the coefficients are linearly related to the amplitude of the modulating function, and b), the frequency of the term is shifted with respect to the original frequency of the modulating signal. Two terms satisfy both of these requirements, one arising from the square factor in the power series and the other from the cube factor. These terms are

$$f_2(t) = a_2A_mA_c\ [\cos\ (\omega_c+\omega_m)\ t + \cos\ (\omega_c-\omega_m)\ t]$$

$$= a_2A_mA_c\ \cos\ (\omega_c+\omega_m)\ t$$

$$+\ a_2A_mA_c\ \cos\ (\omega_c-\omega_m)t \qquad (4\text{-}8)$$

and $$f_3(t) = 3/4\ a_3A_c^2\ A_m[\cos\ (2\omega_c+\omega_m)t + \cos\ (2\omega_c-\omega_m)t]$$

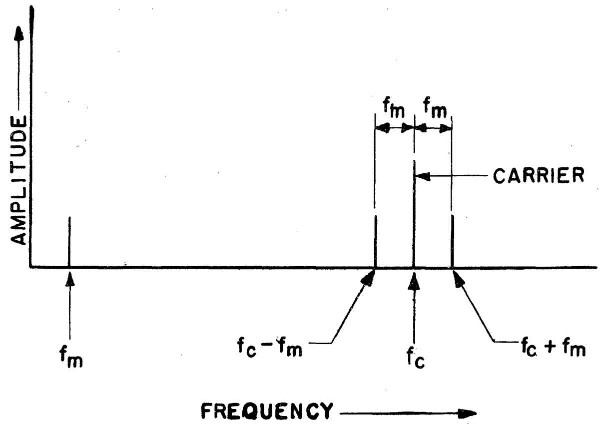
$$= 3/4\ a_3A_c^2\ A_m\ \cos\ (2\omega_c+\omega_m)t$$

$$+\ 3/4\ a_3\ A_c^2\ A_m\ \cos\ (2\omega_c-\omega_m)t \qquad (4\text{-}9)$$

In most modulator circuits a_3 is generally very much smaller than a_2. In practice, then, the third order term is usually too small to be useful and is eliminated, along with the other unwanted frequency components (terms involving $2\omega_c$, $3\omega_c$, etc.), by filters at the output of the device. The term of primary interest is, therefore, $f_2(t)$. Equation 8 shows that $f_2(t)$ consists of two frequency components, one above the carrier frequency and the other below, as plotted in Figure 2. Note that the carrier frequency component, $a_1A_c\ \cos\ \omega_ct$, is also shown.* The two components of $f_2(t)$ are commonly referred to as the

---

* In this and the succeeding analysis the other term falling at ω_c, namely $3/4\ a_3A_c[A_c^2 + 2A_m^2]\ \cos\ \omega_ct$, is, for convenience, being neglected. This is reasonable so long as a_3 is small, which is generally the case.



Spectrum of A Carrier Modulated
By A Single Sinusoid

Figure 4-2

sidebands of the carrier, and the three components together represent a double sideband amplitude modulated signal. Analytically, this signal can be written as

$$f(t) = a_1 A_c \cos \omega_c t + a_2 A_m A_c \cos (\omega_c+\omega_m)t$$

$$+ a_2 A_m A_c \cos (\omega_c-\omega_m)t \qquad (4\text{-}10)$$

which, after some trigonometric manipulation, becomes

$$f(t) = [1 + \frac{2a_2}{a_1} A_m \cos \omega_m t] a_1 A_c \cos \omega_c t$$

$$= [1 + m_a \cos \omega_m t] a_1 A_c \cos \omega_c t \qquad (4\text{-}11)$$

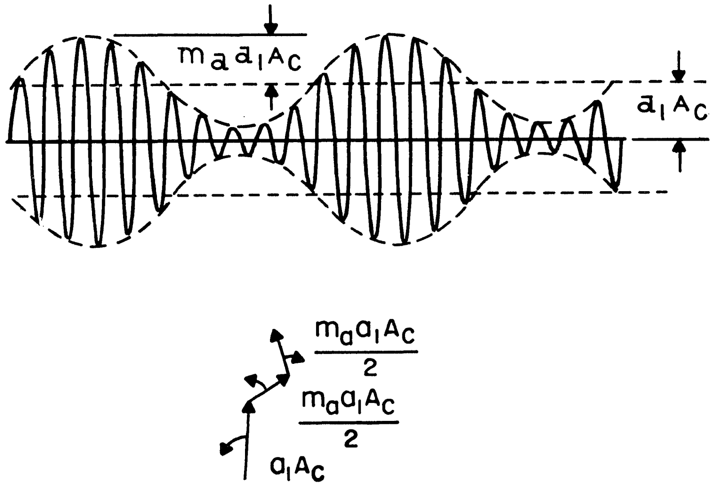
where

$$m_a = \frac{2a_2}{a_1} A_m \qquad (4\text{-}12)$$

Written in this form, it is perhaps more obvious that the carrier has, in fact, been amplitude modulated by the modulating signal as a result of passing both signals through the non-linear device. The factor m_a is known as the "modulation index", "modulation factor", or "degree of modulation". The percentage modulation is given by

$$\text{Percentage modulation} = 100\, m_a \qquad (4\text{-}13)$$

A plot of Equation 11 is shown in Figure 3 along with a vector representation showing how the sidebands combine with the carrier to give this result. This vector representation is further discussed on Page 19-18.



Waveform and Vector Representation of Double Sideband AM Signal

Figure 4-3

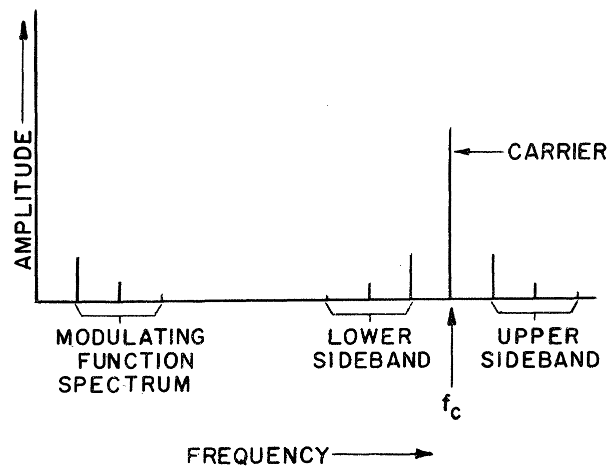
It might be pointed out that these results could not have been obtained if all coefficients of Equation 4 except a_o and a_1 had been zero. Such a device is described as linear, and is not useful as a modulator.

It was stated previously that filters are used to eliminate the other unwanted terms which arise as a result of the modulation process. If the carrier and modulating frequencies are properly chosen, terms of the type $2\omega_c t$, $3\omega_c t$, $2\omega_m t$ and $(2\omega_c \pm \omega_m)t$ are easily suppressed. However, terms of the form $(\omega_c \pm 2\omega_m)t$ can be bothersome since some of them may fall within the sideband frequency range of interest.* In this case the modulator must be designed to minimize the magnitude of these terms.

So far the analysis has dealt only with a single sinusoidal modulating function. What can be said about more complex functions? It is well known that any periodic function can be broken down into a series of sinusoidal components, using the Fourier Series analysis. By extending the method discussed above, it is not too difficult to see

---

*Suppose, for example, that ω_{min} and ω_{max} represent the lowest and highest frequencies, respectively, that the modulating signal can assume. The band of interest for the double sideband AM signal will, therefore, lie between $\omega_c \pm \omega_{max}$. Any modulating signal from ω_{min} to $\frac{1}{2}\omega_{max}$ will then give rise to an $\omega_c \pm 2\omega_m$ component which falls within the band of interest. After demodulation, this unwanted term will appear as a second harmonic of the wanted component.



Spectrum of A Carrier Modulated
By A Periodic Function

Figure 4-4

that the resulting pattern would be as shown in Figure 4.* On the left the original periodic function is shown in terms of a number of its Fourier sinusoidal components. Each of these components is subjected to the same modulating operation that occurred for the single sinusoid case, and each is therefore shifted upward in frequency in a like manner. Thus, each sideband now consists of a number of sinusoids instead of just one. Note that the upper sideband, which consists of terms of the type $(\omega_c+\omega_m)$ is no more than the original modulating function shifted upward in frequency. The lower sideband, consisting of terms of the type $(\omega_c-\omega_m)$ is, on the other hand, not only translated upward in frequency but also inverted. In other words, the lowest modulating frequencies always appear closest to the carrier frequency while the highest modulating frequencies are furthest from the carrier.

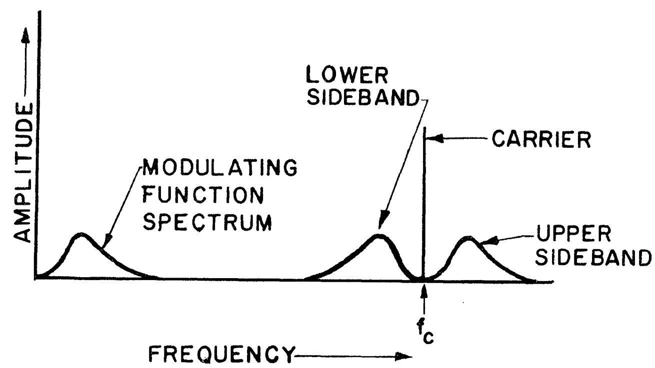
Speech, however, is not a periodic function, so it cannot be broken down into a number of discrete sinusoids following the pattern of the Fourier Series. To deal with non-periodic time functions such as speech, the Fourier Integral and Fourier Transform must be introduced. A discussion of these functions can be found in a later section of the text.** Suffice it to say that it can be shown that a non-periodic time

---

* Only the modulation products similar to $(\omega_c \pm \omega_m)$ are shown; as before, we assume other products are eliminated by filters or by designing to make a_3 negligibly small.

** Chapter 21.

function, such as a speech signal, can be equated to a frequency spectrum. What the Fourier Series does for a periodic function, the Fourier Transform does for a non-periodic function. The difference is that, whereas a periodic function can be portrayed as a discrete series of spikes in the frequency plane, the spectrum of a non-periodic function will be continuous; that is, its frequency components are separated by infinitely small increments of frequency. Using these more advanced methods, it can be shown that for the complex speech signal the result is as shown in Figure 5. Conventionally, a continuous spectrum is portrayed by drawing the envelope of its frequency components.** The original signal is so shown in the figure, at the left in its original voice frequency range and then at the right in the two sidebands. As before, the lower sideband is inverted.



Spectrum of A Carrier Modulated
By A Speech Signal

Figure 4-5

Forms of AM Signals

Up to this point only the double sideband AM signal has been considered. It is evident that the bandwidth required to transmit such a signal is equal to twice the highest frequency component of the modulating signal. Since it is the variation in amplitude of the modulating function which contains the information, and since these variations ap-

- -

* It should always be remembered that a spectrum applies to a signal analyzed over some finite period of time and not at some instant. See the last section of Chapter 21 for further discussion of this point.

pear equally in each sideband, it is reasonable to suppose that the transmission of both sidebands is unnecessary and that the elimination of one would halve the required bandwidth. Furthermore, since there is no message information in the carrier, it too appears unnecessary in so far as the transmission of the original message information is concerned.* The primary virtue of double sideband AM is that it requires the least complicated terminal equipment. This is particularly true at the receiving terminal. However, the Bell System makes use of other forms of AM signals which are derived on the basis that the message information resides in the sidebands. The important forms are single sideband with carrier, single sideband with suppressed carrier, vestigial sideband, and twin sideband.

In single sideband with carrier, a filter is used to suppress one sideband, so that the transmitted signal consists of the carrier and remaining sideband. This method permits increased utilization of the bandwidth of the transmission medium, and the presence of the carrier makes it possible to use relatively simple demodulators at the receiving terminal.**

There is an additional advantage to be gained if the carrier is also suppressed, thereby forming a single sideband suppressed carrier signal.*** The carrier contains no message information, and its elimination means that none of the power handling capacity of the transmission system need be used to transmit it. Putting it another way: if the carrier is suppressed, the useful signal power can be increased, thus giving a better signal-to-noise ratio.**** The suppression of the carrier makes it necessary, however, to reinsert carrier at the receiving terminal for demodulation of the signal. The reinserted carrier frequency

- -

* A carrier is necessary for demodulation and recovery of the message, however, as discussed in the next section.

** Provided one is willing in these single sideband systems to tolerate what is known as quadrature distortion, as discussed in the next section.

*** An easy way to amplitude modulate a signal is by means of the ring modulator discussed on Page 4-17. For more details see Monograph B-1140.

**** Here we assume that our problem is overload. Crosstalk into other systems might forbid our raising the power of the useful signal to override noise.

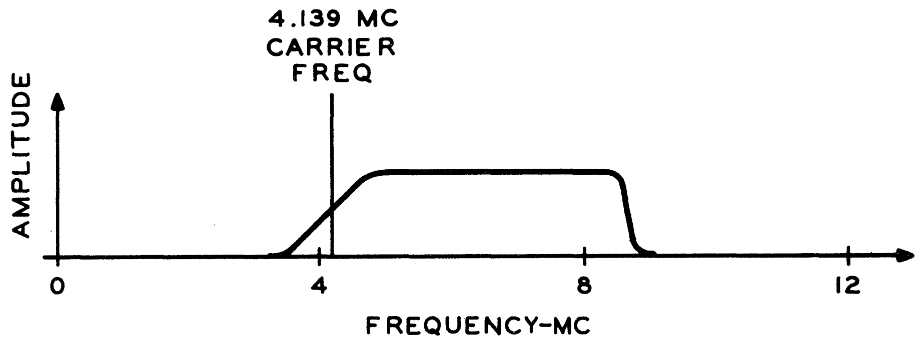
must match the original carrier frequency very closely in order to avoid a frequency shift in the demodulated signal. This is especially critical in program and data transmission.

The restrictions of practical filter design for carrier and sideband suppression produce problems when certain types of modulating signals are used. When the modulating function contains components at very low frequencies, it is not always possible to properly separate the desired sideband from the carrier and the other sideband. This is particularly true when the message is sensitive to changes in the phase relationships of the components of m(t) - that is, when it is sensitive to delay distortion. In telephony this is not a problem because voice frequencies below 200 cycles per second are not transmitted and, besides, delay distortion does not noticeably impair the quality of speech transmission. Filters can be designed, therefore, which offer high attenuation to both the carrier and unwanted sideband without impairing the desired sideband. A TV video signal, however, contains important components down to dc and is very sensitive to delay distortion as is also true for data transmission. No practical filter could include the low frequency components of an upper sideband of the modulated signal without also introducing intolerable delay distortion, unless it also includes some of the unwanted sideband and carrier. Since transmission by double sideband AM would be extremely wasteful of bandwidth, a method known as vestigial sideband transmission is used to transmit such a signal.* The form of this signal, as used in the L-3 coaxial system, is shown in Figure 6. A filter is used to suppress the lower sideband completely except for those frequencies that are within 500 kc of the carrier. From 500 kc below the carrier to 500 kc above, the transmission characteristic of the filter is shaped so as to achieve a response which is symmetrical about the carrier within this band. Frequencies more than 500 kc above the carrier are transmitted unattenuated. The required terminal equipment for this method of transmitting television is complex and expensive.

One of the latest schemes introduced into the Bell System is twin sideband modulation. Simply stated, it involves associating two

---

* The discussion of TV in this chapter is confined to coaxial cable practice and TV broadcasting. Amplitude modulation is not used for TV transmission in such systems as A2A video or radio relay; the problems encountered in such systems are not considered in this chapter.



Vestigial Sideband TV Signal Used in L3

Figure 4-6

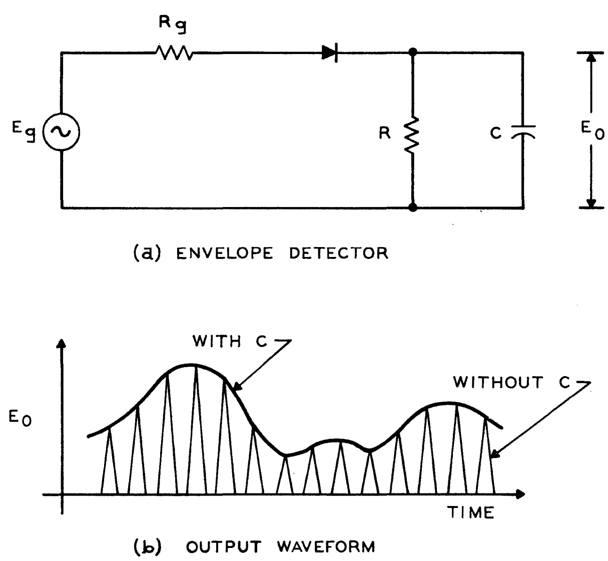
different sidebands with the same carrier. This means taking two different double sideband signals having the same carrier frequency, obtained from a common source, selecting the lower sideband of one and the upper sideband of the other and then transmitting them together with their common carrier. The result is twin, rather than double, sideband modulation and is applicable to telephone transmission. At the receiving end the two sidebands are separated so as to form two signals, each consisting of a carrier and a single sideband, which can then be demodulated.

Demodulation

The previous discussion has shown how the combination of a non-linear device and filter can be used to produce various forms of amplitude modulated signals. At the receiving terminal,means must be provided for recovering the original modulating function from these signals. This is the operation performed by the demodulator. An examination of Figure 1 should make it evident that a non-linear device, in this case characterized by a power series input-output relation, will also act as a demodulator. For this application, the output filter on the device is designed to pass those components representing the original modulating function and reject the others. It is because a non-linear element serves both as a modulator and demodulator that the term modulation has sometimes been broadly used to encompass the process of demodulation. In general, the operation of most demodulators is so highly non-linear that the power series model, so convenient to use in the study of modulation, is not satisfactory for most demodulator analysis.

Therefore, as will be seen in the following discussion, other ways of looking at the operation of a demodulator are generally used.

One of the simplest demodulators is the envelope detector, commonly used to demodulate double sideband AM signals. In its basic form the envelope detector is a rectifier which consists of a diode in series with a parallel combination of load resistance R and capacitance C, as illustrated in Figure 7a. In the absence of the capacitor the diode acts as a simple rectifier, so that with a carrier input, for example, the diode will conduct during one-half the carrier cycle and become nonconducting during the other half. If the carrier is amplitude modulated, the voltage appearing across R then becomes a series of half-sinusoids (ideally), the amplitude of which vary slowly in accordance with the amplitude of the modulating signal.* When the capacitor is placed across R, it will charge up to approximately the peak voltage during the conduction cycle and leak off into the load resistor during the nonconduction period. By properly selecting the RC time constant the voltage across the capacitor can be made to follow, to a good approximation, the envelope of the modulated wave. This is illustrated in Figure 7b. The capacitance serves the additional function of by-passing the carrier frequency and its harmonics, thereby eliminating their presence in the output.



Basic Circuit and Output Waveform of Envelope Detector

Figure 4-7

* This assumes the carrier frequency is very much higher than the highest component of the modulating signal, which is a requirement for envelope detection.

One form of distortion in the envelope detector occurs when the RC time constant is too large, so that the capacitance does not discharge sufficiently during the troughs or minimum values of the modulated wave. High frequency detail in the modulating signal is then lost. Other distortion terms arise because of non-linearity in the diode conduction characteristic. A number of means have been devised for eliminating these distortions. It can be easily demonstrated, for example, that the distortion increases as the modulation index increases, so that in combating distortion either: a), the index of modulation is kept low or b), additional carrier power is supplied at the receiving terminal to effectively produce a low index modulation signal at the detector input. This latter technique is often referred to as "enhanced carrier" operation, and the detector is described simply as a diode detector or, sometimes, linear detector.

The envelope detector can also be used to demodulate single sideband with carrier and vestigial sideband AM signals. The absence of one sideband, however, leads to a form of distortion known as quadrature distortion. Quadrature distortion is discussed in some detail in Chapter 16 so that only a brief example of it will be given here. If, for example, the lower sideband is completely eliminated, Equation 10 becomes

$$f(t) = a_1 A_c \cos \omega_c t + a_2 A_m A_c \cos (\omega_c + \omega_m) t \qquad (4\text{-}14)$$

which, after some trigonometric manipulation, can be written as

$$\begin{aligned} f(t) &= [1 + \frac{a_2 A_m}{a_1} \cos \omega_m t]\, a_1 A_c \cos \omega_c t \\ &\quad - [\frac{a_2 A_m}{a_1} \sin \omega_m t]\, a_1 A_c \sin \omega_c t \\ &= [1 + \frac{m_a}{2} \cos \omega_m t]\, a_1 A_c \cos \omega_c t \\ &\quad - [\frac{m_a}{2} \sin \omega_m t]\, a_1 A_c \sin \omega_c t \end{aligned} \qquad (4\text{-}15)$$

where m_a is the modulation index, defined in Equation 12. The first term of Equation 15 has the same form as Equation 11 and represents a double sideband AM component of the signal. The second term is the quadrature component, which derives its name from the fact that it is 90° out-of-phase with the first component. It constitutes a distortion term which will appear in the output of the envelope detector. The magnitude of the quadrature distortion component at the output can be reduced, however, by using a low index of modulation. In order to see this, let $V_i(t)$ represent the envelope of Equation 15, given by the square root of the sum of the squares of the amplitudes of the $\cos \omega_c t$ and $\sin \omega_c t$ terms in 15. Thus,

$$V_i(t) = \left\{ [1 + \frac{m_a}{2} \cos \omega_m t]^2 (a_1 A_c)^2 + [\frac{m_a}{2} \sin \omega_m t]^2 (a_1 A_c)^2 \right\}^{\frac{1}{2}} \tag{4-16}$$

The desired envelope is, of course, $[1 + \frac{m_a}{2} \cos \omega_m t] a_1 A_c$. Factoring this term out of Equation 16 gives

$$V_i(t) = [1 + \frac{m_a}{2} \cos \omega_m t] a_1 A_c \left[1 + \left(\frac{\frac{m_a}{2} \sin \omega_m t}{1 + \frac{m_a}{2} \cos \omega_m t} \right)^2 \right]^{\frac{1}{2}} \tag{4-17}$$

The term

$$\left(\frac{\frac{m_a}{2} \sin \omega_m t}{1 + \frac{m_a}{2} \cos \omega_m t} \right)^2$$

is the distortion component. This term can be reduced by making m_a small, either by using a low index of modulation in the system or by effectively reducing the modulation index by supplying additional in-phase carrier* at the input to the envelope detector.

Since the ear is not sensitive to quadrature distortion in message circuits, an envelope detector can be used in a single sideband with carrier telephone system. Television signals, on the other hand, are very sensitive to quadrature distortion.** Where economy is of prime importance, envelope detection is nevertheless used, as in home

---

* Any phase difference between the original and re-inserted carrier will produce additional quadrature distortion components at the output.

** We have here another example of the fundamental thesis that a knowledge of the "structure and sensitivity of the message" is essential to transmission engineering. Without experimental evidence, one could hardly assume that telephone is tolerant to quadrature distortion and television is not.

reception of broadcast TV. The amount of quadrature distortion in such TV transmission systems can be reduced by extending the vestigial band farther beyond the carrier, still making the sloped region symmetrical about the carrier frequency. Where the requirements on quadrature distortion are tight, however, as in long coaxial systems with many terminals in tandem, a product-type detector can be used. This form of detector is also used in demodulating single sideband suppressed carrier signals.

The product demodulator is illustrated in Figure 8, where $f_i(t)$ represents an input such as a single sideband suppressed carrier signal and $C(t)$ is a carrier signal. Analytically, these signals can be expressed as

$$f_i(t) = A \cos (\omega_c+\omega_m)t \qquad (4\text{-}18)$$

and

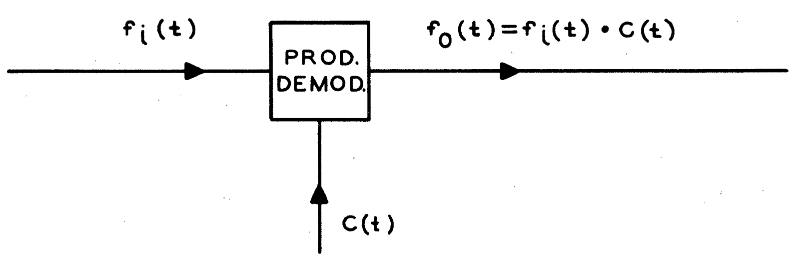
$$C(t) = A_c \cos \omega_c t \qquad (4\text{-}19)$$

The output of the product demodulator is obtained by multiplying Equation 18 by 19; it is, therefore,

$$\begin{aligned} f_o(t) &= f_i(t)\, C(t) \\ &= A \cos (\omega_c+\omega_m)t[A_c \cos \omega_c t] \qquad (4\text{-}20) \end{aligned}$$

which can be written as

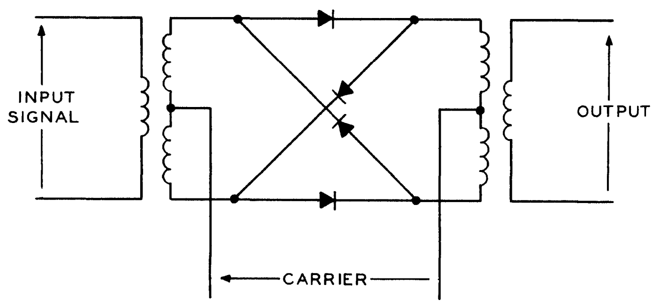
$$f_o(t) = \frac{AA_c}{2} \cos \omega_m t + \frac{AA_c}{2} \cos (2\omega_c+\omega_m)t \qquad (4\text{-}21)$$



Functional Diagram of Product Demodulator

Figure 4-8

By using a filter on the output of the detector, the second term in Equation 21 can be eliminated and the first term, representing the original modulating function, is recovered.*



Ring-Type Modulator Circuit

Figure 4-9

Probably the most common form of product detector is the bridge modulator, one form of which (often called the ring modulator) is shown in Figure 9. This circuit can be used both in the formation of a single sideband suppressed carrier signal and its demodulation. A relatively large carrier voltage is used to "switch" open and closed the diodes in the path of the input signal. When the circuit is used as a modulator, an analysis** of the output signal shows the useful output term to consist of the upper and lower sidebands on either side of the carrier frequency. The single sideband signal is then formed by using a filter to suppress the unwanted sideband. For demodulation of a single sideband input signal, the output filter selects the term formed

---

* The above analysis has assumed, of course, that the supplied carrier has exactly the same frequency and phase as the original carrier. This is difficult to achieve, and the result of a phase difference between the original and supplied carriers is to introduce quadrature distortion components. In telephony, where quadrature distortion is not a problem, satisfactory reproduction of the speech signal requires that the supplied carrier for a single link system (i.e., one pair of terminals) be only within about 20 cps of the original carrier frequency. For the more general case, however, where the system is expected to operate in tandem with a number of other systems, the supplied carrier must be held within about 2 cps of the original. TV requirements are, of course, considerably tighter. These requirements and the methods used for meeting them are discussed in Chapter 16.

** An analysis of this circuit can be found in Reference 4 listed at the end of this chapter.

by the difference between the carrier and sideband frequencies - in other words, the original modulating function.

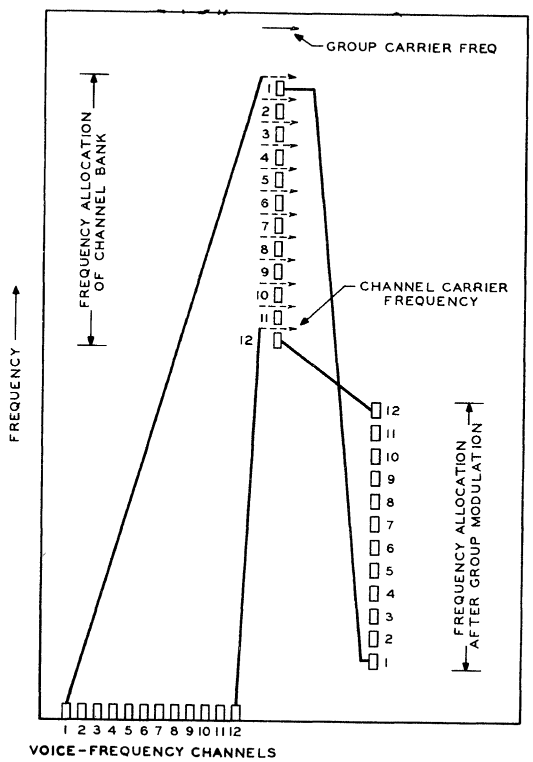
Purpose of Amplitude Modulation in Telephony

The preceding discussion has shown how an information bearing signal can be translated, by the process of amplitude modulation, to a new position in the frequency spectrum. This is the primary application of amplitude modulation in the telephone system - to shift the frequency range occupied by a signal to make it more suitable for transmission over a particular medium. Television transmission in the L3 system serves as a good illustration of the application of this principle. The L3 system is not designed to transmit below 300 kc. Since a television signal contains frequency components down to dc, some translation of the signal must be made. In the system, therefore, the television signal is modulated with 4.139 mc carrier and formed into a vestigial sideband signal which is then transmitted in the band extending from 500 kc below the carrier up to 8.5 mc.*

Amplitude modulation also makes it possible to translate a number of different signals, each originally occupying the same frequency range (e.g., voice channels), to new positions in the frequency spectrum so that all of them can be transmitted over the same medium without interfering with each other. This is the principle of frequency division multiplex, widely used in telephony. One example of a frequency division multiplex terminal is illustrated in Figure 10. Here each of twelve voice frequency channels is modulated by a carrier supply so as to form a "channel group" consisting of twelve single sideband suppressed carrier signals. The equipment which performs this frequency multiplexing operation on the voice channels is generally called the "channel bank" equipment.

It would seem logical, of course, to translate the voice-frequency signal directly to its assigned channel on the line by means of a single step of modulation. However, two factors generally make this unfeasible. The first of these is the need to economize on the number of different channel oscillators and filters required, particularly in systems handling a very large number of channels.

* This particular choice of frequency allocation is influenced by many factors which are beyond the scope of this chapter.

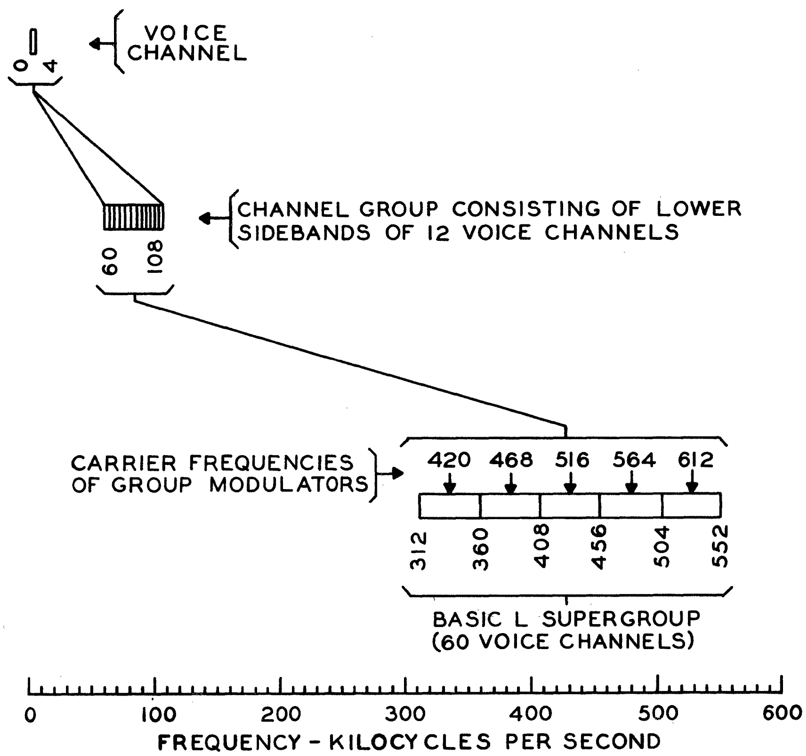


Frequency Multiplexing in A
Typical 12-Channel Terminal

Figure 4-10

Equally important in any size system is the problem of designing the channel filters used after the channel modulators. In many cases the frequency range for optimum design of these filters is not coincident with the frequency range which is most suited for transmission of the channels on the line. Therefore, at least two or more steps of modulation are generally required to translate the voice frequency signals to the proper transmission frequency. The usual practice is to make this translation by using "group" modulators and carrier supplies to amplitude modulate the channel group to the line frequency. For the case illustrated in Figure 10, a single step of group modulation is used.

The type A channel banks are the basic building blocks in most carrier systems. In these units each of twelve voice frequency circuits is allocated to a 4 kc channel in the group, and the entire group occupies a 48 kc bandwidth extending from 60 to 108 kc. This 12 channel group is basic in the Bell System and provides a common base for the ready interconnection of unlike systems. In the J and K systems a single step

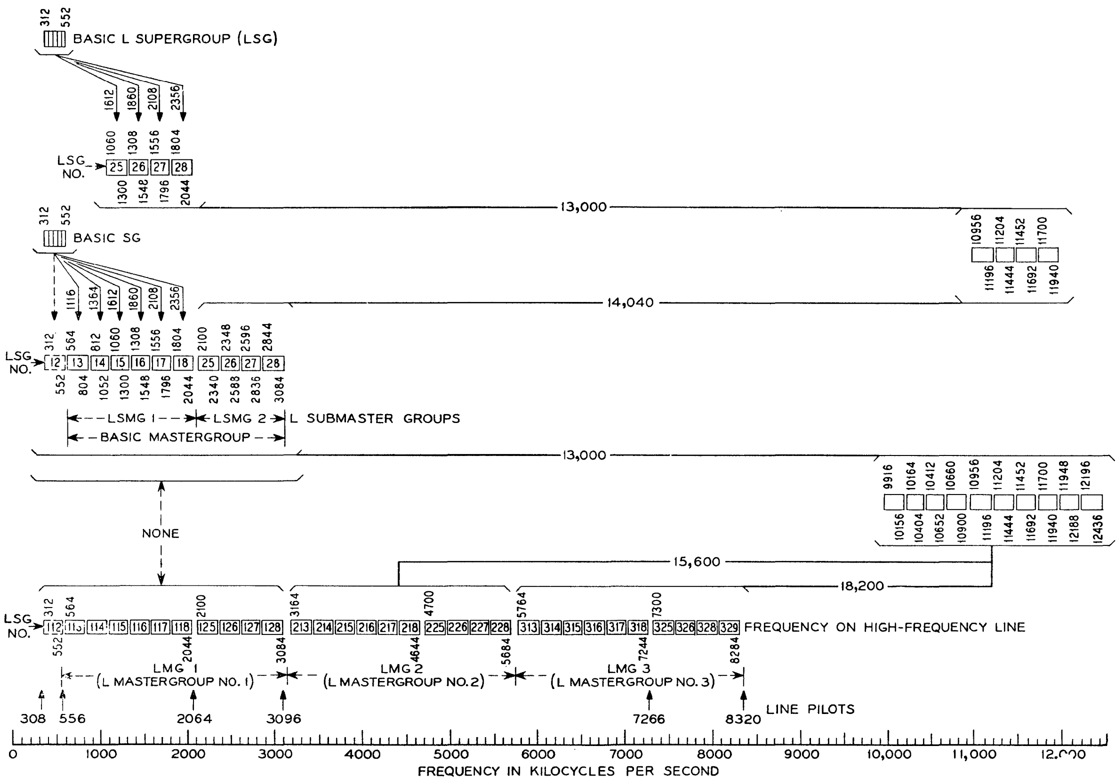


Basic L System Supergroup

Figure 4-11

of group modulation then translates this group of channels to the line transmission frequency. Systems of larger capacity, such as the L, are built up by applying the principle of group modulation a number of times to form large arrays of channels. First, the outputs of a number of channel banks are stacked up to form an "intermediate" size group; then these groups are stacked up to form still larger groups, and so on. Certain of these larger groups have been given special names. For example, an array of 60 channels (5 channel groups) is called a supergroup. Ten supergroups (600 channels or 50 channel groups) is a mastergroup. Figure 11 illustrates the formation of the basic L system supergroup. and Figure 12 shows how these supergroups are used in the L3 terminals to form an array of 1860 telephone channels.

Single sideband suppressed carrier transmission is by far the best method of transmission in long-haul multi-channel carrier systems, both from the economic and performance standpoints. There are a number of reasons for this. A single sideband occupies the same



L3 Frequency Allocation

Figure 4-12

bandwidth as the original voice-frequency signal. With modern filters it is possible to stack the single sideband signals on a spacing as low as 4 kc and still provide a useful bandwidth which is somewhat wider than 3 kc. Therefore, single sideband transmission achieves maximum utilization of the available bandwidth.

Another factor favoring this form of transmission is that it, of all forms of AM transmission, requires the least power handling capacity of the repeaters, since the carriers are not transmitted. In a large system carrying a thousand or more channels this may lower the power handling requirement by 40 db or more.

Another advantage of suppressing the carrier is that the unwanted intermodulation products* produced in the system tend to be noise-like in character. If carriers are transmitted, many of the

---

* The problem of modulation distortion is introduced in the next section of this chapter and is discussed in detail in Chapter 8.

intermodulation products will be either tones or intelligible crosstalk, and the resulting interference is more annoying and must therefore be more rigidly controlled.

For these reasons then, single sideband suppressed carrier transmission finds application in such long-haul systems as the J, K, and L, as well as submarine cable systems.

Single sideband transmission does, of course, require relatively sharp cut-off filters for the formation and extraction of the channels. The filter complexity increases as the channel spacing is reduced. These filters represent an appreciable cost in the terminals. In addition, carriers of quite precise frequency must be reinserted at the receiving terminal for demodulation. This also adds to the terminal cost. Therefore, when terminal costs (rather than line costs) must be kept low, as in short-haul carrier systems, single sideband transmission generally looks unfavorable.

The simplest and often cheapest signal form to use when terminal costs dominate is conventional double sideband with carrier. Channels can be spaced as close as 8 kc without unduly increasing the channel filter costs. Of course, if the filters are cheapened by permitting more gradual cut-offs, crosstalk within and between systems may be aggravated if the channels are too closely spaced. This form of signal has found application in the short-haul N system and the rural carrier P system.

The O system illustrates a case of a short-haul system where double sideband AM did not prove to be the best choice. In cable systems, where there are many cable pairs available, the double sideband signal may prove attractive. However, in open wire, the cost per pair is considerably higher and there are usually few pairs available on a particular route. Therefore, over certain length routes it may be cheaper to use single sideband or something approaching it in order to economize on bandwidth and allow more channels to be stacked up instead of installing additional plant. Such is the case with the O system, which uses twin sideband to stack up four channels using two carriers 8 kc apart. Some savings are made in the number of filter designs required by frequency multiplexing only four voice-frequency channels to form a group, and then using group modulators to translate the banks of four channels to the proper frequency allocation. This also permits the system to be installed in terms of basic four-channel units when it is not desired to put a maximum capacity (16 channels) system into service.

From this discussion, it should be clear that there is no one best choice as to the type of AM signal to be used. In approaching a new system design, the engineer must weigh the relative importance of terminal costs to total system cost, the system requirements, and the nature of the message, before selecting the type of signal to be used in the system.

Modulation Distortion

Up to this point modulation has been viewed as a desirable process, but it often occurs when it is not wanted. Unwanted modulation will arise in amplifiers, for example, because no matter how well designed, an amplifier is never ideally linear. In fact, the relation between the input and output of a typical amplifier can be quite well described by the power series of Equation 4. In a well designed amplifier a_1 will be large compared to the higher order coefficients, but these other factors will be present, nevertheless. Their effect will be felt more and more as the input drive on the amplifier is increased since the amplitude of the unwanted terms varies as the product of the amplitude of the fundamentals. For example, the third harmonic of the fundamental varies as the cube of the fundamental amplitude. Generally, it is possible to disregard terms above the cubic in the series, as was done in the modulator analysis. This is because a_4 and higher coefficients are usually small, so that fourth order (and higher) terms become important only at overload. Thus, if the input signal consists of two sinusoids (a carrier and one component of a sideband, for example), the output of the amplifier will consist of the sort of products shown in Figure 1.* Only the first order or linear term which is a function of a_1 is wanted. All the other terms - second and third harmonics as well as cross products of the original signals - represent distortion products. The harmonics are usually referred to as harmonic distortion and the cross products as intermodulation products. Thus, the whole problem of this form of distortion is generally given the name "modulation distortion", "intermodulation", or simply "modulation".

- -

* At this point the reader may want to refer to Table 1 of Chapter 8 which shows the power series expansion when three sinusoids are present. As long as terms above the cubic in the power series are not considered, the effect of now adding a fourth (or any additional sinusoids) to the input signal is to simply increase the number (i.e., quantity) of products shown in Table 1 without introducing any new forms or types of terms. New forms of cross products and harmonic combinations of the input signals result only by considering higher order terms in the power series.

The important point to be gained from this discussion is that these distortion terms can either a), impair or interfere with the desired signal itself, or b), cause interference and impairment to signals which pass through the same device at other carrier frequencies. Stating b) in another way: Modulation distortion can give rise to a form of crosstalk between channels. In the next chapter the importance of modulation distortion in AM carrier systems is discussed in more detail, while Chapter 8 gives a method of analyzing the distortion terms in a carrier system design.

It might be pointed out that the dc term which, in Figure 1, is a function of a_2 is commonly referred to as the "dc shift" in power amplifiers. Another term which bears a special name is the first order term which is a function of a_3. Because a_3 is usually negative, this term will subtract from the desired output. As a result the power output from an amplifier will compress or fall off as the input drive is increased and this unwanted term increases in magnitude. Hence the term is known as the "compression term".

Summary

An amplitude modulated signal can be obtained by passing both a carrier and the modulating signal through a non-linear device. A number of forms of AM signals - double sideband, single sideband, twin sideband, and vestigial sideband - have found application in various carrier systems. Single sideband is most favorable in long-haul multichannel carrier systems, where costly terminals can be justified. In shorter systems, or systems of lower capacity, double sideband AM is generally the most economical choice, although some situations warrant the use of twin sideband. Vestigial sideband transmission is used for AM transmission of TV to minimize the required bandwidth without introducing excessive quadrature distortion. Other variants of AM transmission have found application outside the Bell System; these are considered to be outside the scope of this chapter

Non-linearity, which is capitalized on to form and detect the AM signal, also gives rise to unwanted products at the output of system repeaters, terminal amplifiers and other devices which are never ideally linear. These products constitute noise and crosstalk which add to the other noise sources in the system and degrade the signal-to-noise ratio. In the following chapters we will see how modulation distortion is taken into consideration in the design and operation of an AM carrier system.

Bibliography

1. H. S. Black, "Modulation Theory", D. Van Nostrand Co., Inc., New York, 1953.

2. S. Seely, "Electron-Tube Circuits", McGraw-Hill Book Co., Inc., 1950.

3. S. Goldman, "Frequency Analysis, Modulation and Noise", McGraw-Hill Book Co., Inc., 1948.

4. D. G. Tucker, "Modulators and Frequency Changers for Amplitude Modulated Line and Radio Systems" MacDonald, London, 1953.

5. J. F. Honey, D. K. Weaver, "An Introduction to Single Sideband Communications", Proceedings of the IRE, December 1956, pages 1667 - 1675.

Chapter 5

INTRODUCTION TO AM CARRIER SYSTEMS

The problems encountered in the design and operation of AM carrier systems are surveyed. Four-wire and "equivalent four-wire" methods of operation are compared. Crosstalk problems encountered in cable and open-wire systems using either method of operation, and various solutions to these problems, are described. The functions of terminals and repeaters, and the characteristics of transmission lines, are reviewed. Problems which influence repeater spacing - noise, crosstalk, overload, modulation distortion - are introduced. Misalignment, equalization and regulation are also described. The chapter is intended to serve as an introduction to the more detailed material on AM systems in Chapters 6 to 15.

Introduction

In the early days of telephony, speech signals were transmitted over the wire lines only at their natural voice frequencies. It was soon realized, however, that this was a very inefficient use of the costly wire plant since a non-loaded line was capable of transmitting a much wider frequency range than the 0 to 3 kc or so required for a voice frequency signal. There were other disadvantages to voice frequency operation. For example, the relatively low velocity of propagation on loaded cable (about 13,000 miles per second on 19H-88, for example) made it necessary to either operate at high via net loss or use echo suppressors to maintain satisfactory echo performance on long toll circuits. Either of these solutions to the echo problem simply exchanges one form of transmission impairment for another.* There was, therefore, a strong incentive to develop a transmission system which would utilize some of the wasted frequency band above 3 kc. This would provide additional telephone channels on the existing wire plant. In addition, these channels would be transmitted in a frequency range where the propagation velocity, for all types of non-loaded facilities, was higher than at voice frequency.** On

---

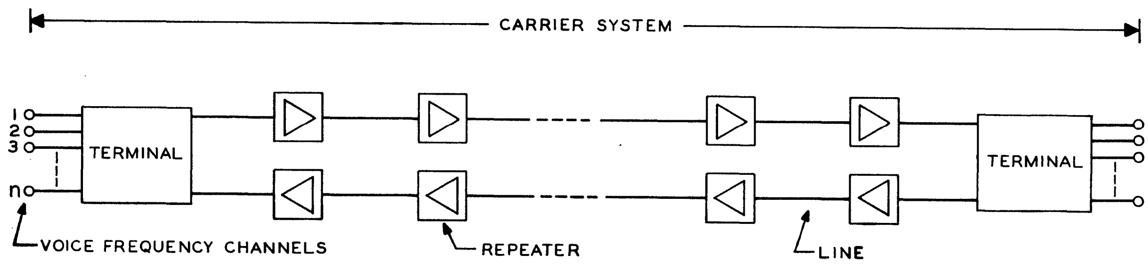
* The disadvantages of echo suppressors are discussed in Chapter 2.

**This is particularly true for cable. The propagation velocity for 19 gauge cable, for example, is 76,000 miles per second at 3 kc, 126,000 miles per second at 100 kc, and 140,000 miles per second at 1 mc.

the basis of echo requirements, then, these systems could be operated at lower via net loss than an equal length voice frequency system using the same facility. This made it possible to improve the plant performance by permitting a reduction in the loss of a typical toll connection. The advent of the electron tube, improved filter design techniques and components, and the principles of amplitude modulation of a carrier provided the means for such a development. The first AM carrier system, designated the type A, made its appearance during World War I. This system, now obsolete, has been followed by a succession of carrier systems designated by letters of the alphabet, the most recent being the P carrier system designed for rural service application. Table I attached at the end of this chapter tabulates the important characteristics of the carrier systems in use today.

It should be pointed out that the term "carrier system" is generally used in a restricted sense in the telephone industry. Broadly speaking, a carrier system is any form of communication system which makes use of a wave which can be amplitude, frequency, or phase modulated so as to "carry" an information-bearing signal. As such, the term carrier system applies not only to those systems which transmit the modulated wave over open-wire, cable pairs, or coaxial cable, but to radio systems as well. However, the term is usually restricted to refer only to those systems which transmit the modulated wave over a metallic facility. It is this meaning of the term which will be used in this and the following chapters on AM transmission.

Except for the type L systems which require coaxial cable, all of the present carrier systems are designed to be applicable to one or more of the existing standard types of line facilities. The application of a carrier system to an existing line requires the installation of the carrier terminals and repeaters, as illustrated in Figure 1. In most cases, special line treatment is also necessary, such as the use of suitable transposition arrangements on open-wire or the balancing of cable pairs. The cost of both the equipment and the line treatment represents, therefore, the cost of the telephone channels furnished by the carrier system. In addition, many of the carrier systems are not designed to permit operation of the voice-frequency circuit on the same facility. This is an important consideration in some cases since the net increase in channels added to the line is then one less than the channel capacity of the carrier system, and the cost per channel is correspondingly increased.



Basic Components of an AM Carrier System

Figure 5-1

If a carrier system is to prove in economically, either when it is added to an existing facility or installed as part of a new plant expansion, the cost of obtaining the additional channels by means of the carrier system must be less than the cost of obtaining the same number of channels by other means, such as a radio system, or simply additional voice frequency circuits.* The terminals, for example, represent an important part of the cost of a carrier system. For a particular system, this is a fixed cost, regardless of the system length. When expressed on a cost per channel-mile basis, however, the terminal cost looms as a much larger part of the total cost on a short as compared to a long system. It follows that for each type of carrier system there is some minimum length of system below which the carrier costs per channel-mile are so great that the system does not prove in, and it is more economical to obtain the needed telephone channels by other means.

The prove-in range of carrier systems leads to the natural classification of "long-haul" and "short-haul" systems. The long-haul systems, such as the J, K, and L, are designed to meet all the transmission requirements of a toll link for the longest system which would be

---

* To be completely fair in this cost comparison, one would have to somehow take into account the transmission advantages - e.g., lower VNL - of carrier channels as against voice frequency channels.

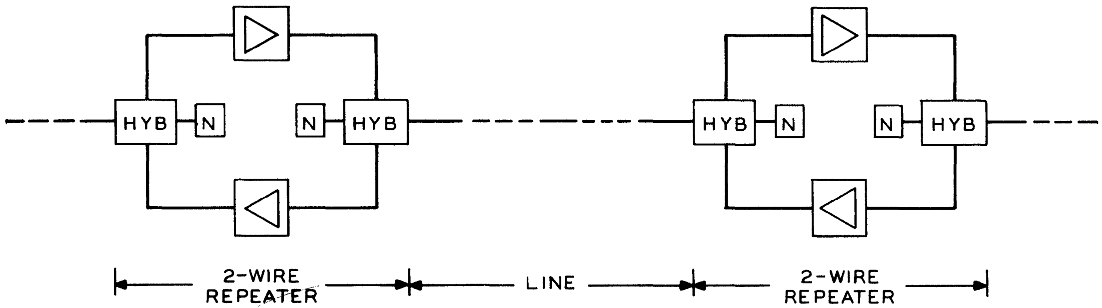
encountered in the United States (4000 miles). Because of the resulting complexity of the terminal and repeater equipment, the minimum prove-in length for these long-haul systems is of the order of 75 to 100 miles. In recent years the need for more circuits on shorter toll trunks and in the exchange area plant has led to the development of short-haul or end-link type systems such as the N, O, and P. Since the transmission requirements of these systems can be eased in view of the shorter end-link type distances to be spanned, cheaper terminal and line equipment can be built which bring prove-in distances down to 10 or 15 miles. Conversely, these short-haul systems must not be extended beyond a maximum distance, of the order of 200 miles, if the transmission impairment introduced by each such system in a multi-link toll connection is not to exceed the transmission requirements of the nationwide toll-dialing plan.

The purpose of this chapter is to acquaint the reader with some of the important problems encountered in the design and engineering of an AM carrier system. Most of the discussion is of a general, qualitative nature. It would be highly desirable from the reader's viewpoint if the features of a single existing carrier system were woven into the discussion to illustrate specific solutions to the problems under consideration. However, no one system seems to adequately illustrate all of the factors that are discussed, so that, instead, frequent reference is made to characteristics of a number of systems. A more complete description of these and other AM carrier systems is beyond the scope of this text, and the reader interested in more detail should refer to the references at the end of this chapter.

Method of Operation

A carrier system can be operated on a wire transmission facility in a number of different ways. As will be seen in this discussion, the problem of crosstalk between systems (except in those systems which use coaxial cable), combined with economic considerations, determines to a large extent the choice of the method of operation for a particular system.

The simplest and oldest of these methods is known as two-wire operation and is illustrated in Figure 2. Here both directions of transmission make use of the same carrier-frequency band and the same pair of conductors between repeaters. At each repeater a pair of hybrids is used to split the two oppositely directed signals into separate paths



2-Wire Operation

Figure 5-2

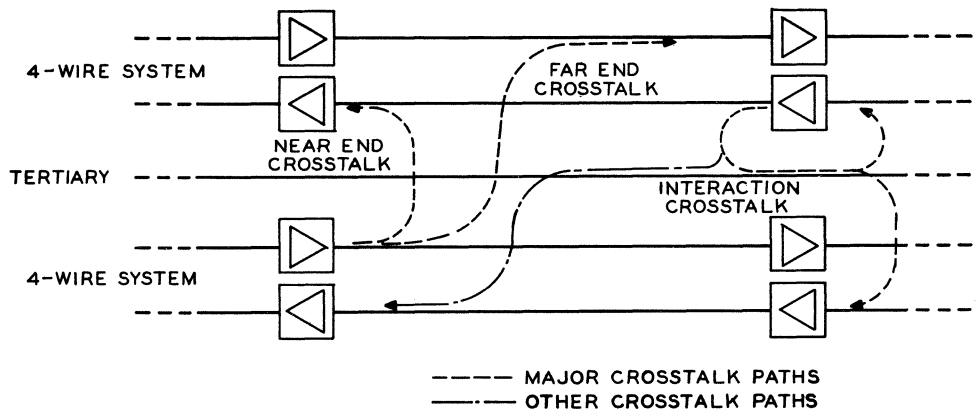
for amplification. At first sight, two-wire operation seems to permit maximum utilization of the wire plant. As in the case of two-wire voice frequency operation,* however, the major problems with this method are those of obtaining suitable balance in the hybrids (i.e., high loss between opposite arms) when the system is first installed and then maintaining adequate balance as the line impedance varies with time (temperature, moisture conditions, etc.). Because of these difficulties it is necessary to keep the repeater gains low to insure circuit stability. This, in turn, leads to high via net loss for these circuits. These problems become more severe as we consider higher and higher carrier frequencies. In consequence, we cannot multiplex many channels on a pair, so the cable or wire cost of two-wire systems turns out to be high on a per channel basis. As a result, all of the carrier systems used on wire lines or coaxial cable, with but two exceptions,** have been designed to operate on a four-wire instead of a two-wire basis.

There are two forms of four-wire operation: "physical" four-wire (which will be referred to simply as four-wire) and equivalent four-wire. In both methods of operation each voice frequency telephone channel

---

*See Chapter 3 for a discussion of two- and four-wire operation on voice frequency toll systems.

**The A system (4 channels, single sideband suppressed carrier, open-wire), now obsolete, and the short-haul G system (1 channel, double sideband with carrier, open-wire), which has found very limited use, operate on a two-wire basis.



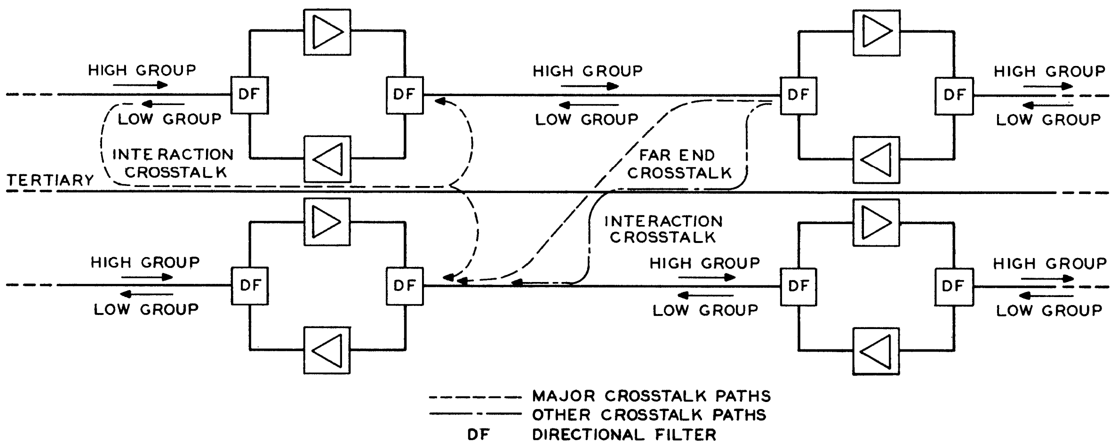
4-Wire Operation and Some of the Important Crosstalk Paths

Figure 5-3

handled by the system is divided by means of hybrid circuits (4-wire sets) at the system terminals into two oppositely directed one-way channels which, by either of two means, are kept separate and distinct from one end of the system to the other.

In four-wire operation, a separate pair of conductors is used for each direction of transmission, as shown in Figure 3. This is, of course, exactly the same principle studied earlier in the text under four-wire voice frequency toll systems. In "equivalent four-wire" systems, different frequency bands are used to form a "high group" and "low group" for the two directions of transmission, thereby permitting operation over a single pair of conductors between repeaters, as shown in Figure 4. The two directions of transmission are thus separated in frequency rather than by physical location as in the case of four-wire operation. At each repeater, high-pass low-pass filters, usually referred to as directional filters, are used to separate the two directions of transmission for amplification.

Each method of transmission has its advantages and disadvantages. Equivalent four-wire, because of the frequency difference between the two directions of transmission, eliminates the near-end crosstalk problem.



Equivalent 4-Wire Operation and Some of the Important Crosstalk Paths

Figure 5-4

It provides a system (of a limited number of channels) requiring only one pair of conductors. When these are important considerations, equivalent four-wire operation may be the optimum choice; most open-wire systems, for example, are of this type.

On the other hand, equivalent four-wire has serious drawbacks. Guard bands must be left between the pass-bands of the directional filters, so that some of the spectrum which would otherwise be available cannot be utilized for transmission. The filter impedances are not good terminations for the hybrids, especially outside the filter pass bands, leading to reflections and transmission irregularities. If the crosstalk problem can be ignored (e.g., in a well-shielded coaxial system) and if in the long run the maximum number of channels are needed, physical rather than equivalent four-wire will usually be found the best choice.

Growth rate may dictate the final decision in many cases. Consider, for example, the problem of providing telephone service by submarine cable. Assume that we have, with reasonable repeater spacing, a 1000 kc bandwidth available. With physical four-wire (two cables) we can transmit 250 two-way channels (4 kc spacing). Using equivalent four-wire, we might lose about 160 kc in guard bands, giving us 840 kc of useful band, or 105 two-way channels per cable. Eventually, then, two such equivalent four-wire systems would provide only 210 channels as against 250 for physical four-wire. But if we do not expect traffic to require more than 100 channels for many years, equivalent four-wire (assuming it to be technically feasible) might be the better economic choice.

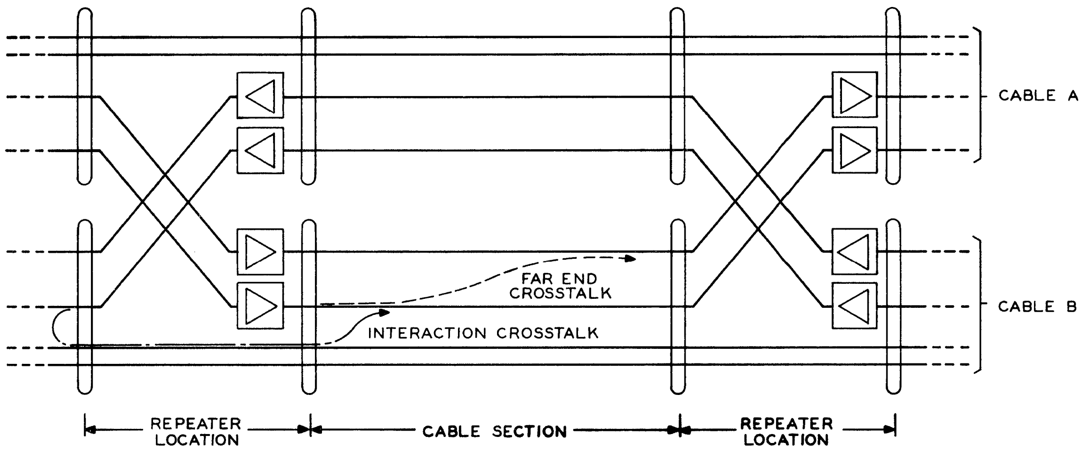
Reliability is another consideration. In our submarine cable example, suppose that traffic studies indicate that we need 150 channels as soon as the system is put into operation. This will require two cables regardless of which method of operation is used. If we use equivalent four-wire, we still have 105 channels left if, for example, a trawler cuts one cable; if we use physical four-wire, all 250 available channels go when either cable goes.*

We have used the submarine cable case for this discussion because it is in many ways a simple one compared to land systems. When we introduce additional constraints - crosstalk, for example, or the necessity of locating repeaters at particular points where power is available, or getting from open-wire into cable in order to get through a city - the decision between various methods of transmission may become even more difficult. On the other hand, we may find one factor making the decision for us, as in the case of crosstalk on open-wire lines.

As the reader knows, crosstalk is a serious problem in four-wire cable systems unless special measures are taken to control it. Three major crosstalk paths between systems are shown in Figure 3. These are 1), the near-end path between the opposite directions of transmission; 2), the interaction crosstalk paths from the output of one repeater into a paralleling cable pair (a voice circuit perhaps) and then into the input of the same or another repeater; and 3), the far-end path from the output of one repeater to the input of another. A method which is used in the K system to eliminate the first two of these paths is shown in Figure 5. Two cables are used alternately to provide the pairs for each direction of transmission. By using physical isolation, the near-end crosstalk paths between the opposite directions of transmission are automatically eliminated. The interaction crosstalk path is effectively broken up by alternating the two directions of transmission between the two cables in successive repeater sections, as shown in Figure 5. In this way, the

---

* All this is predicated on the feasibility of designing satisfactory broadband equivalent four-wire repeaters for submarine cable systems; such feasibility has not been established yet.



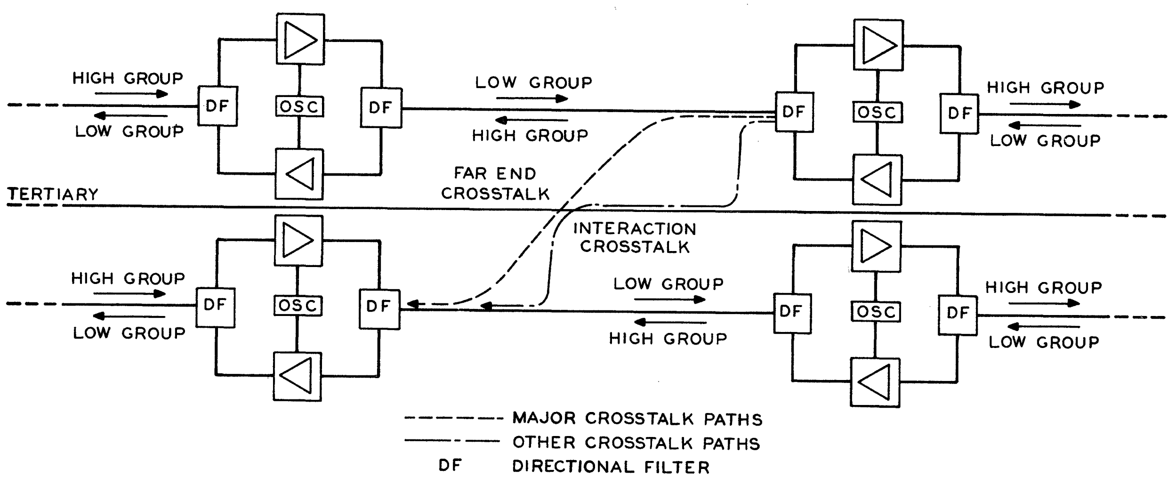
K-System Method of Near-End and Interaction Crosstalk Control, and Some of the Crosstalk Paths that Remain

Figure 5-5

interaction path is made to terminate at the high level point at a repeater output and is, therefore, less serious by the gain of a repeater. These measures do not, of course, affect the far-end crosstalk between carrier systems in the same cables. This last path can be reduced only by special carrier-frequency balancing of the cable pairs.

In open-wire systems the conductors are completely unshielded from electric and magnetic fields. It is not possible, therefore, to electrically isolate groups of pairs for crosstalk control as one can by using two cables on a cable route. As a result the equivalent four-wire method of operation is generally used on open-wire. As can be seen from Figure 4, equivalent four-wire eliminates the near-end crosstalk between the two directions of transmission because of the frequency separation. However, there still remain the problems of interaction and far-end crosstalk within and between systems using the same facility.

One method which can be used to reduce both of these forms of crosstalk between equivalent four-wire systems is to stagger the frequency allocations of the systems using the same facility. This can be done, for example, by using slightly different (1 kc or so) carrier frequencies for each system, or, in a single sideband system, by using the lower sidebands



Frequency-Frogging in Equivalent 4-Wire for Interaction Crosstalk Control, and Some of the Crosstalk Paths that Remain

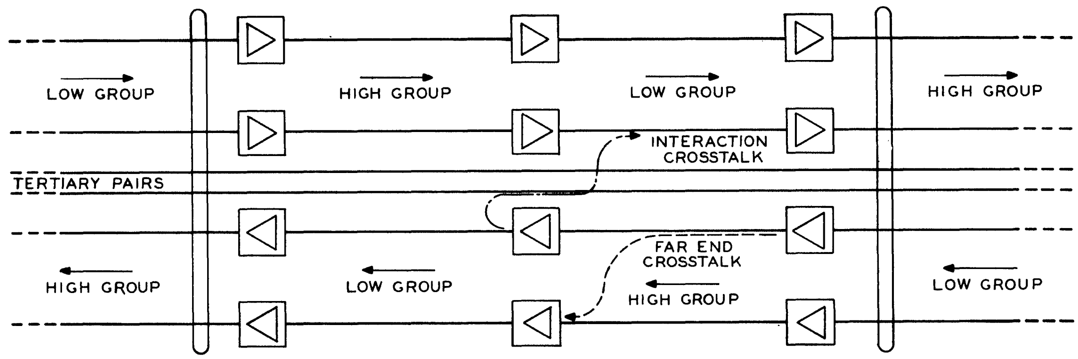
Figure 5-6

of one set of carriers in one system and the upper sidebands of the same carrier frequencies in the adjacent system. These methods have the effect of reducing the coupling between systems by either shifting or inverting the spectrum of the disturbing crosstalk with respect to the signal in the disturbed channel, thereby reducing or destroying the intelligibility of the crosstalk. A combination of these methods, together with a suitable transposition scheme, is used for crosstalk control on the long-haul open-wire J system. On a light-route system using P carrier, and where two systems are needed but the cost of line treatment is to be avoided, it is planned to control inter-system crosstalk by the use of a staggered frequency plan such that the channels of the "staggered group" system are inserted in the wide (6 kc) guard-band spaces between the channels of the regular system.

Another method which, in an equivalent four-wire system, controls interaction crosstalk from a repeater output to its own or another repeater input is that of frequency frogging, illustrated in Figure 6.*

---

*A further advantage is attained by frogging by using the lower sideband each time a group is modulated from high to low and vice versa. Since the lower sideband has the frequencies inverted with respect to the modulating signal, this cancels the slope of the line attenuation versus frequency which greatly simplifies the equalization problem.



N-System Method of 4-Wire Operation to Control Interaction and Near-End Crosstalk

Figure 5-7

In each repeater a modulator frogs the frequency of the low group to the high group, and the high group to the low group. Thus, for either direction of transmission, the repeater output is always in one frequency band and the input in the other. As a result, the interaction crosstalk path from a repeater output through a paralleling circuit to its own or another repeater input is blocked by the out-of-band characteristic of the input filter. This method removes one of the major crosstalk paths between systems and also eliminates a feedback path around each repeater, thus permitting higher gain to be used. Frequency frogging for interaction crosstalk control, together with a relatively simple transposition scheme to reduce far-end crosstalk, is used in the short-haul open-wire O system.*

Another interesting application of the frequency frogging technique is used in the type N short-haul cable system. Four-wire operation is used in this system, but this is combined with the high-group low-group principle of equivalent four-wire. The equivalent four-wire principle alone removes the problem of near-end crosstalk between the opposite directions of transmission of different systems using the same cable. In addition, by frequency frogging at each repeater, as shown in Figure 7, the interaction crosstalk path around a repeater is also eliminated, and it becomes possible to use the same cable for both directions

---

* Compandors, which will be discussed in detail in a later chapter, are also vitally important in the O, as well as the N and P systems, for meeting crosstalk requirements.

of transmission. Thus, this method achieves control over the same crosstalk paths that the two cable method of Figure 5 does.

From this discussion it should now be obvious that the method of operation of a carrier system is far from an arbitrary choice. Even in an isolated system (and a coaxial system can be considered isolated even though other coaxial systems may use the same cable) the choice is not always obvious. Four-wire operation requires an immediate use of twice as much copper, but in the long run a four-wire system will provide more than twice the number of message channels as two equivalent four-wire systems occupying the same bandwidth. Each amplifier in a four-wire repeater must then be designed to work over twice the bandwidth required for an amplifier in an equivalent four-wire system. However, this cost and complexity must be weighed against the corresponding costs, in the long run, perhaps, of two equivalent four-wire systems and, therefore, two sets of amplifiers and directional filters at each repeater location. Where crosstalk between systems on the same facility is a problem, the systems study must include the cost-performance relations of special line arrangements and treatment as compared to staggered frequency allocations and frequency frogging techniques and the additional terminal and repeater complexities involved. Since each new system is designed to meet a particular need, the variety of methods of operation which have been and will be used reflect the results of the study of all these factors in terms of the service to be provided and the state of the technology at the onset of system design.

Form of AM Signal

The factors which influence the choice of form of AM signal were discussed in detail in Chapter 4. To briefly review, the forms of AM signals which have found application in one or more carrier systems are double sideband with carrier, twin sideband, and single sideband suppressed carrier.*

Double sideband with carrier requires the simplest modulating and demodulating equipment. It is used, therefore, in relatively low capacity light-route systems where terminal costs must be held to a minimum.

Twin sideband permits associating two different sidebands with the same carrier. For the same bandwidth, then, a twin sideband system

---

* A fourth form, single sideband with carrier, has found use only in the now obsolete B carrier system.

can have twice the channel capacity of double sideband with carrier system. Although twin sideband increases the terminal complexity, its use is justified where the increased terminal costs are less than the cost of installing additional line facilities to handle the same number of channels on a double sideband with carrier basis.

Single sideband suppressed carrier is the best form of signal for a long-haul, multi-channel system. This method of AM transmission makes maximum use of the available bandwidth and minimizes the power handling capacity requirements of the repeaters since the carriers are not transmitted. It has the further advantage that intermodulation products produced in the system tend to be noise-like rather than intelligible in character. The terminal equipment required for single sideband suppressed carrier operation is, of course, relatively complex and costly.

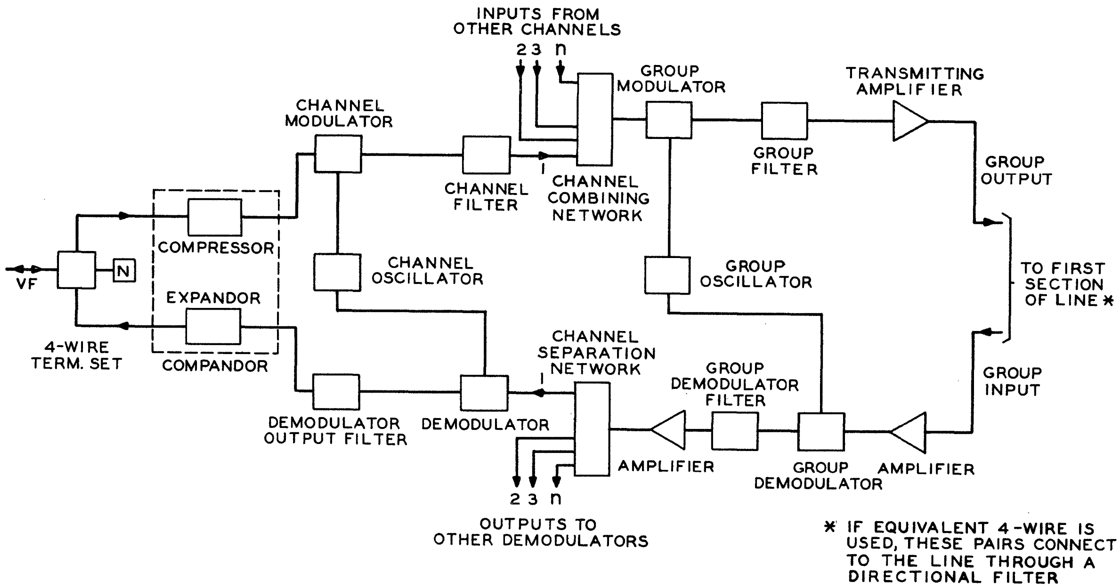
In summary, then, it follows that although single sideband suppressed carrier is probably the most desirable form of AM signal to use, it proves-in only on relatively large, long-haul type systems, such as coaxial or submarine cable systems, where line and repeater costs, rather than terminal costs, dominate. On light route, short-haul type systems, where terminal costs become more important, either the double sideband with carrier or twin sideband form of AM signal can be used. An economic study of the problems involved with either type of signal will generally dictate the choice for a particular system.

Terminals

The complexity and, therefore, cost of the system terminals depends on the form of AM signal used and the number of channels handled by the system. The major components of the transmitting and receiving terminals of a system handling the order of a dozen single sideband suppressed carrier channels are illustrated, in block schematic form, in Figure 8. The terminal arrangement of a larger capacity system is very similar, the primary difference being that a greater number of channel modulators and group modulators are used, and generally more steps of group modulation are involved.*

It is assumed, of course, that either four-wire or equivalent four-wire operation is used. The transmitting terminal begins, therefore, with a voice-frequency hybrid or, as it is more generally known, a

* Frequency multiplexing is discussed in detail in Chapter 4.



Components of Carrier System Terminals

Figure 5-8

four-wire terminating set, for each channel.* This network splits the voice-frequency path into the two directions of transmission. The transmitting side of the four-wire set is connected either to a compressor, if compandors are used in the system, or directly to the channel modulator and its associated output filter. For the case illustrated in Figure 8, the outputs of all the channel filters are combined in a channel combining network to form a frequency multiplex signal occupying the frequency range of the channel filters. This signal is then fed into a single step of group modulation which shifts the array to the line transmission frequency. The output of the group modulator is fed through a group filter, which selects the required sideband from the modulator and suppresses any other out-of-band products. The signal then goes into the transmitting amplifier. This amplifier raises the signal level for transmission over the first section of the line, which in many cases, may be the nominal line length between repeaters. In some cases a special transmitting amplifier may be designed for the terminal application, but frequently the same (or approximately the same) amplifier

---

*This assumes two-wire switching. At a four-wire switching center, the system terminals begin and end with what in Figure 8 would be the four-wire sides of the terminating set. When, at a four-wire switching center we connect to a two-wire facility (e.g., a toll connecting trunk), the four-wire set is associated with the two-wire facility.

that is used in each repeater can be used as the transmitting amplifier. It may, however, provide some special shaping of the signal (pre-emphasis) to partially equalize the system signal-to-noise performance vs frequency.

From the output of the transmitting amplifier the group passes through the line connecting network to the line. In four-wire operation a repeat-coil is generally used to match the amplifier impedance to the line (and is usually made part of the transmitting amplifier). In equivalent four-wire operation the connecting network must, of course, include the directional filter.

The receiving terminal generally goes through the same steps as the transmitting terminal, but with the order of operation reversed. For this reason many of the components, such as the group and channel filters, are the same in both the transmitting and receiving terminals. There is, however, considerable variation in the receiving terminal layouts of the various systems, and Figure 8 is intended to illustrate only the basic operations performed. Note that if the system uses compandors, the expandor follows each channel demodulator; otherwise the output of the demodulator circuit is connected directly to the four-wire set. Also, instead of using a single receiving amplifier, separate amplifiers are generally used to optimize the signal levels ahead of the group modulators and channel demodulators.

The method used for obtaining the oscillator frequencies required in the system terminals depends principally on the form of AM signal used, the type of signal (telephony, TV, program, etc.) transmitted, and the number of terminals which are expected to be connected in tandem. Up to this point in the discussion we have considered the terminals as the end points in a carrier system. A long-haul system will generally have a number of intermediate terminals in addition to those at the ends of the over-all system. This permits some channels (one or more supergroups in the L system, for example) to be dropped and others to be added at offices located along the route. Those channels which are not dropped are, of course, not brought down to voice frequency, but are wired directly to a transmitting terminal and sent on, either to another dropping point or to the end of the system.

It follows that as the number of terminals in tandem increases, the requirements on both the absolute and relative frequency of the carriers supplied at the terminals also increases in order to avoid frequency shifts in the signals transmitted by the system. In a single sideband system involving only a pair of terminals, an error of ten to twenty cycles in carrier frequency still provides satisfactory telephone message service. For program transmission the error should not exceed about two cycles. When we increase the number of terminals, however, the carrier frequency errors at each terminal must be held considerably below these values.

One method which is used in the long-haul single sideband systems is to derive all the necessary carrier frequencies from a precision crystal controlled oscillator. In the L system the relative frequency error between terminals is also reduced by designating the crystal oscillator at one end of the system as a master oscillator and using its output to synchronize the frequencies of all the other oscillators in the system.

When carrier power is transmitted, the demodulation problem at the receiving terminal is considerably eased.* Thus, in a double sideband system, for example, simple crystal controlled oscillators are generally used to supply the necessary carrier frequencies.

In addition to supplying carrier power to the modulators, oscillators are frequently used in the transmitting terminals (particularly in single sideband systems) to provide pilot tones which operate the regulator circuits in the repeaters. The details of the regulation problem will be discussed more fully in a later section. Suffice it to say here that the oscillators used for this purpose must be accurately controlled in frequency to permit sharp separation filters in the repeaters to separate the pilots from the channel groups. As a result the pilot tones are frequently derived from the primary oscillator from which the carrier frequencies are obtained. In some cases, instead of generating new frequencies, certain carriers can be transmitted to act as the pilot tones. Of course, when the system normally transmits the carriers, as in a double sideband system, the carriers themselves can generally be used by the regulator circuits and no special pilot tones are needed.

The discussion so far has made no mention of signaling. At the time such systems as the J and K were designed, signaling was considered a function entirely separate from the rest of the system. As a result it was necessary at the toll offices to install racks of signaling equipment along with the carrier terminals in order to translate the local (dc) signaling into a form which the carrier system could transmit. The signaling information was then fed into the terminals on the two-wire side of the terminating sets and was handled by the system just as if it were a voice-frequency signal. More recently there has been an increased emphasis on the integration of all aspects of the transmission system.

---

* In addition, since message frequencies are equal to the difference between carrier and sideband frequencies, errors in group carrier or frogging oscillator frequencies do not degrade the received message.

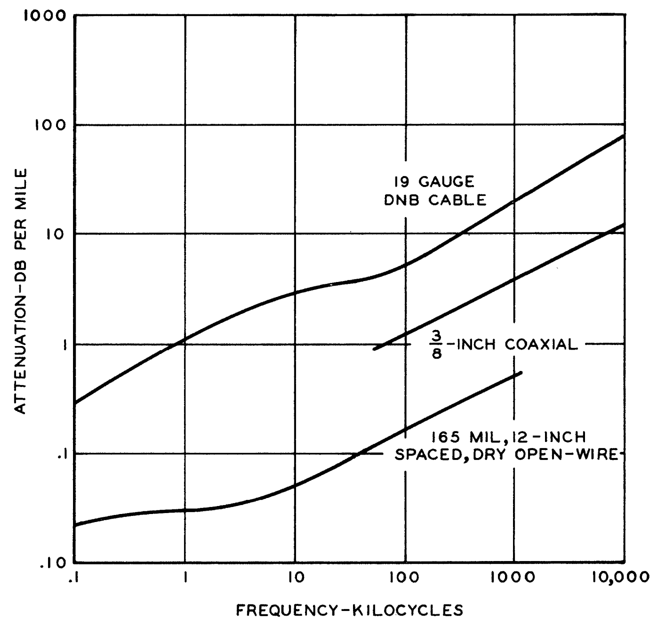
This has led to a closer coordination between the signaling needs and the terminal layout. Thus in the N and O systems, for example, a 3700 cps oscillator is included in each voice-frequency circuit to supply signaling tones. The circuit can be connected so that incoming dial pulses or supervisory signals in the voice-frequency channel will pulse this oscillator on and off. These pulses are then applied to the channel modulator and transmitted to the receiving terminal where they are recovered by a channel signaling receiver and restored to their original form. These systems will, of course, transmit the conventional multi-frequency key pulses and some of the other forms of signaling which are transmitted within the voice band.

This completes our discussion of the carrier system terminals. The next section takes up the problem of the high frequency line over which the signals are transmitted.

Lines and Repeaters

A specific wire facility on which the carrier system is to operate is frequently picked during the system design. This is especially true for long-haul cable and coaxial systems, where the transmission requirements are usually best met by integrating the repeater design and method of operation with a particular type and size line. In other cases, the system may be designed for application to a variety of lines. This is particularly desirable for a short-haul type system which is intended to provide additional message channels on existing wire plant.

With either type of system design, the wire medium offers a number of problems which the repeaters must be equipped to solve. One of these is the nominal line attenuation versus frequency characteristic, which is illustrated in Figure 9 for a typical size cable, coaxial, and open-wire line. For all types of lines, the attenuation increases with frequency, with the attenuation at high frequencies increasing as the square root of frequency. This normal attenuation characteristic necessitates the use of some method of equalization to make the via net loss of the system independent of the carrier frequency used for transmission of the signal. It will be recalled from Chapter 3 that, in voice frequency toll systems, some of the equalization in cable systems is provided by means of loading coils. Extension of the loading principle into the carrier frequency region would require the use of relatively close spaced, low inductance coils to move the cut-off frequency of the loaded line above the transmission band. The use of lower inductance, however, would reduce the effectiveness of the loading in minimizing the attenuation distortion of the line. Considerations such as these lead



Attenuation vs Frequency Characteristics of Various Types of Wire Facilities

Figure 5-9

to the conclusion that this method of line treatment generally is not practical at carrier frequencies. As a result, system equalization is normally achieved by suitable gain-frequency shaping techniques at the repeaters and terminals.* Various equalization methods are discussed in a later section.

Another line characteristic which is frequently important is the delay-frequency characteristic. Delay distortion (i.e., distortion of the signal which occurs when the delay is a function of frequency) is generally not a problem in telephone message circuits, since it produces negligible impairment. For other types of signals - data, monochrome and color television, and program, for example - delay as well as attenuation

- -

*One exception to this occurs in the case of toll entrance and intermediate cable links used in the C and J open-wire systems. Toll entrance cables are cable sections used to extend an open-wire line into a city. Additional cable links are frequently required to carry the open-wire circuit across river beds or around other obstacles. The lengths of these toll entrance and intermediate cable links vary from a few hundred feet to several miles, and each link must, of course, be worked into the over-all system equalization plan. Very light loading (together with special low-capacitance cable in the J system) is used in C and J carrier to provide some attenuation equalization.

equalization must be provided in the system. Whereas attenuation equalization is provided at every repeater, delay equalization in AM carrier systems is generally inserted only at the system terminals, and here it may be applied only to the signal (or signals) being dropped at that location. An exception to this occurs in a system such as the L which is equipped with protection switching (i.e., an automatic switching arrangement which permits substitution of a standby section of repeatered line for a working section which, for example, may fail or require a maintenance check). In this case, delay equalization is inserted at the ends of each switching section so that when the signal is switched from the active to the standby section, the over-all delay characteristic of the system is left unchanged.

So far we have discussed problems associated with the nominal line characteristics. An equally important problem is the change in line attenuation with changes in ambient conditions. Aerial cable, for example, has a mean temperature of 55°F in the northern United States. However, the annual variation in temperature is expected to cause the cable temperature to vary ± 55° or more. These temperature changes cause the cable resistance to change, which, in turn, changes the attenuation by the order of ± 5%. (This is approximate - the actual change is a function of frequency.) Smaller changes, of course, occur during the period of a day. Buried cable is somewhat more stable, the expected annual variation in temperature being ± 20°F. Submarine cable systems, once on the ocean floor, should see temperature variations only of the order of ± 1°F, corresponding to an attenuation change of ± 0.12%. The importance of even so small a change as this can be more fully appreciated when it is recognized that the total cable attenuation may, at high frequencies, amount to 10,000 db or more on a long system, so a 0.1% change is 10 db!

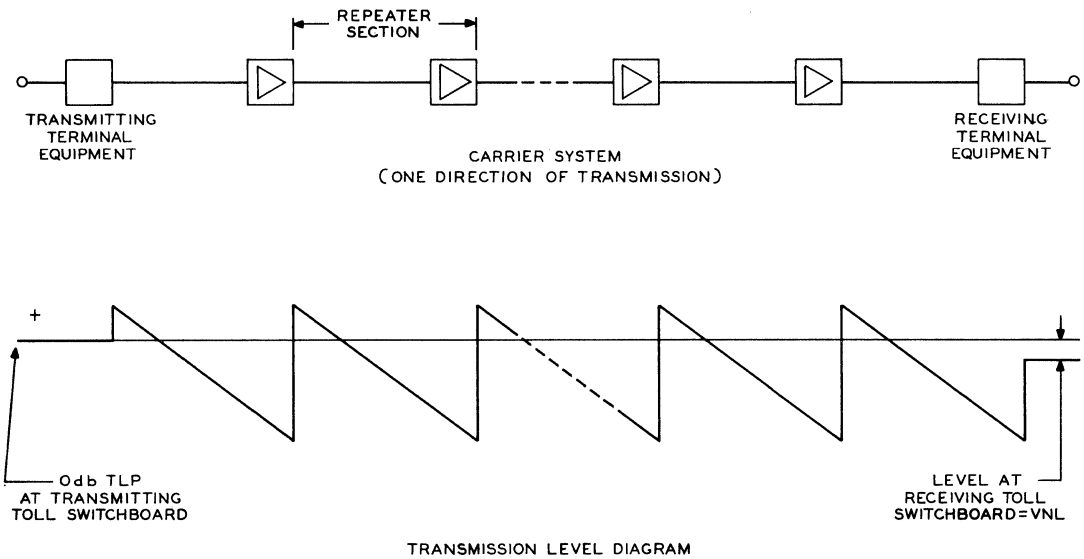
The attenuation changes of open-wire line are influenced primarily by moisture conditions, which cause variations in the shunt conductance of the pairs. Although at voice frequency the change in line loss with weather conditions, even on long systems, is usually of no importance, at carrier frequencies increases of the order of 35% in the attenuation can be expected during rainy weather. Sleet and ice loading produce even larger changes, of two to four times (or more), over the normal dry attenuation of the line.

This problem of attenuation change with changes in the ambient conditions requires that some means of automatic regulation be provided to maintain the via net loss of the system at (or close to) its assigned value. Several methods of regulation are described briefly in a later section. In general, regulation in carrier systems is provided at each repeater, so that by one means or another the repeater gain is made to change in accordance with changes in the attenuation of the preceding section of line.

Changes in cable temperature (or other ambient effects) will also cause changes in the delay characteristic of the system. Fortunately, the changes in delay are very closely given by the well-known relationship between gain (or attenuation) and phase - the so-called minimum phase law. This being so, delay distortion effects due to changes in ambient conditions tend to be automatically eliminated by correcting the gain deviation. To do this job adequately in systems where delay distortion is particularly important (e.g., L-3 coaxial), it is necessary to design the gain-correcting networks to follow the gain deviation curve to a frequency about half an octave above the top transmitted frequency.

From this discussion we see that the basic function of the carrier system repeaters is to make up for both the nominal loss of the preceding section of line as well as to compensate for changes in that loss. We have also seen that the two directions of transmission in a carrier system are separated, either physically in four-wire operation, or by frequency in equivalent four-wire. In either case this permits us to examine either direction of transmission as though it were a separate and distinct system. Figure 10 illustrates a typical level diagram for one direction of transmission in a multi-repeater system. The system begins with the terminal equipment at the transmitting toll switchboard. The 0 db transmission level point is located on the 2-wire side of the voice-frequency input to the terminal. The amplifier in the transmitting terminal raises the level for transmission over the first section of line. The amount by which levels are thus raised is a point to be discussed later. Then follows a succession of alternate line sections and repeaters until the receiving terminal is reached. Ideally, each repeater provides just enough gain to restore the signal to the level it had at the output of the transmitting terminal. The gain at the receiving terminal is then adjusted to deliver the signal at the receiving toll switchboard at the proper level as determined by the via net loss at which the system is to operate.

How far apart can the repeaters be spaced? (We assume, in asking this question, that if we change repeater spacing we will make the appropriate change in amplifier gain so that repeater gain will always



Transmission Level Diagram for One Direction of Transmission in a Multi-Repeater System

Figure 5-10

just compensate for cable loss.) This is a vital question, for the repeater spacing and, therefore, the number of repeaters required for a particular installation and their complexity, will directly affect the total cost of a carrier system facility. The repeater spacing question is clearly tied to the question of the levels to which signals are raised by the transmitting amplifier before launching onto the high frequency line. We find also that it involves the question of bandwidth - how many channels the system can carry. The next nine chapters of this volume consider these inter-related problems in detail. Obviously, then, all we can do in this chapter is to consider them in a very broad-brush sort of way.

With this restriction in mind, let us now look at some of the factors which the system designer, and, in some cases, the engineers concerned with laying out a new system, must consider in determining the spacing of the repeaters in a particular system.

Repeater Spacing and Transmitted Levels

In an AM carrier system, as in any type of multi-repeater system, each section of line, each repeater, and each terminal installation contribute to the total impairment of the message. Frequently, the impairment contribution of only one or two of these components will dominate in a particular system. Terminals, for example, generally add only a small amount of noise and other disturbance in a long, multi-channel carrier system. In all systems, however, allocations must be made to the magnitudes of various impairments which each of the system elements is allowed to introduce, and the system must then be designed and engineered to meet these allocations. To derive these allocations, it is necessary to translate channel objectives, such as those discussed in Chapter 2, into detailed requirements on each source of impairment. As we shall see, although the requirements allocated to the repeaters do not uniquely determine the repeater gain and, therefore, the spacing, they do act as a base for determining various combinations of spacing and gain from which the system designer must select what appears to be the optimum.

There are a number of types of impairment in an AM carrier system. Insofar as repeater spacing and level are concerned, however, they may usually be classified into one of the following categories: noise, crosstalk, modulation distortion, and overload.* The contribution of each of these factors to the total transmission impairment will increase as the system length and, therefore, the number of sources of impairment, increase. Let us consider these items to see how they affect repeater spacing (therefore, repeater gain) and transmitted levels.

Among the various types of sources of noise which are important at carrier frequencies are thermal noise, tube noise, battery noise, contact noise, static, and corona. Most of these have been described in Chapter 3. Some of them, like thermal and tube noise, are built into the system. Others, like static, come from external sources. Thermal and tube noise happen also to be amenable to calculation, for which reason subsequent chapters use them for illustrative purposes. The other noise sources mentioned cannot be computed; we rely on measurements for all our knowledge of them (and hope that the measurements are adequate and typical).

Whatever its nature and source, noise will be introduced in the repeatered line, and result in some particular signal-to-noise ratio at the input of the receiving terminal. Leaving compandors out of the discussion for the moment, the best we can do from the receiving terminal to the subset is to preserve this S/N ratio without further degradation.

---

*In addition to these problems, consideration must also be given to the problem of singing, or near singing, caused by unwanted coupling between the output and input of each repeater. As we have seen, the method of operation frequently eliminates the major coupling paths, but other paths that remain may assume importance if repeater gains are set too high.

In general, the magnitude of the noise at receiving terminal input will be independent of the signal magnitude. We can, therefore, improve the S/N ratio by a), launching signals from the transmitting terminal at higher levels, and b), shortening the distance between repeaters so that signals never sink too near to noise magnitudes. Conversely, as we lengthen repeater spacing or decrease signal levels, we worsen the S/N ratio.

All that has been said with respect to the noise within a system applies equally well to crosstalk from other systems. These two impairments can be thought of as setting a lower limit, below which our signals in a system must not sink; they represent the floors of a tunnel through which our signals must pass.

Intermodulation, crosstalk into other systems, and overload represent the ceilings. They prevent us from transmitting from the originating terminal (and from succeeding repeaters) at very high level. They thus set an upper limit on the output level of each line repeater. Taken together, the lowest ceiling and the highest floor govern in any particular case - one setting the maximum level at repeater output, the other setting the minimum level at repeater input. These two limitations determine the maximum loss, or cable length, between adjacent repeaters, and so govern the repeater spacing.

In concept, all this is simple enough. It is when we want to be numerically precise in any particular case that we find ourselves, of necessity, developing the sort of definitions and equations which are to be found in the following chapters.

Effect of Compandors

How do compandors affect our conclusions? The basic idea behind the compandor can be stated simply enough. It is a device which, at the receiving terminal, reduces the transmission of a channel through the receiving terminal when no signal is present on that channel. Since it is the noise or crosstalk between signal syllables which annoys the subscriber, the apparent performance of the system is greatly improved. When signal is present, the channel equipment at the terminal provides approximately normal transmission. The receiving terminal device is called an expandor since it increases the difference between signal and noise (or between strong signals and weak signals). To make over-all transmission of the signals themselves satisfactory, a corresponding compressor is added at the transmitting terminal.* By compressing the volume range at

---

*The word "compandor" is merely a contraction of "compressor and expandor", of course.

the transmitting end, we also make the signals from all talkers more nearly equal during their transmission over the trunk. As a result, loud-talker crosstalk into other systems is reduced, and weak talkers are transmitted satisfactorily. When there is no compressor at the transmitting terminal, weak talkers suffer more noise impairment during their transmission over the high frequency line than strong ones do; compressors decrease this relative disadvantage.

All this is very advantageous, and permits the operation of systems which would otherwise be non-commercial. The essential nature of our problem is unchanged, however. That is, we reduce but do not eliminate the effects which limited us before we added compandors, and so merely move the limits further out.

Misalignment

Throughout the above discussion, and in some of the following chapters, it is assumed that the repeater gains just compensate for the loss of the preceding section of cable. This design objective can be met only approximately in practice. Let us consider briefly the effect of departure from this objective.

Misalignment is defined as the cumulative departure from the ideal transmission in successive repeater sections (i.e., departure from the repeater gain equalling cable loss). The effects of misalignment can be appreciated from the following definitions:

a) Negative misalignment: Negative misalignment is present when the cable loss in each section exceeds the repeater gain, so that the signals become progressively smaller as they are transmitted from repeater to repeater. As a result, the effects of random noise (thermal, tube, etc.) become increasingly important. On the other hand, modulation distortion and crosstalk into other systems is reduced since each repeater handles a smaller signal.

b) Positive misalignment: Positive misalignment is present when repeater gains exceed cable losses, so that the signals become progressively stronger as they are transmitted from repeater to repeater. Modulation distortion and crosstalk into other systems therefore increases, and the signals more nearly approach the system overload point. The ratio of signal-to-random noise is, of course, improved.

These definitions illustrate rather clearly the concept of the "floor" and "ceiling" limits between which the signals must be carried.

It can be shown that for either type of misalignment the net effect is likely to be a penalty in the signal-to-interference ratio. Generally in a string of repeater sections, one tries to make the misalignment random. If it is, the penalty to the system will be smaller than if the misalignment is systematic. Methods for computing the effects of misalignment are discussed in Chapter 11. The amount of misalignment in a system will depend on the equalization and regulation performance. Let us take a look at these problems now.

Equalization

The basic problem of equalization is to compensate for the loss versus frequency characteristic of the line. The requirement on "flatness" of transmission depends on the amount of misalignment penalty the system can stand. Low transmission levels increase the degradation produced by thermal noise and crosstalk, high levels increase modulation noise, crosstalk into lower level channels, and may overload repeaters. In addition to these factors which require that the transmission level be kept within certain limits everywhere along the line, the overall transmission from terminal to terminal must be kept within certain limits to meet transmission requirements such as via net loss, delay distortion, and flatness.

Equalization methods vary from system to system. In general, loss equalization is provided at each repeater. One method, used in the K and J systems, for example, is to use an amplifier which has flat gain across the band and to insert shaping networks ahead of the amplifier so that the over-all transmission compensates for the line loss characteristic. Another method, as used in the L3 system, is to provide part of the equalization by shaping the gain-frequency characteristic of the repeater amplifier. Either method will generally impose signal-to-noise penalties on at least some of the channels, so that part of the problem of system design is to chose an equalization plan that appears best suited to the particular system.

In addition to basic equalization at each repeater, most systems include "mop-up" equalizers located at some of the repeater points along the line (usually at main stations, or dropping points). Mop-up equalizers are versatile, manually adjusted equalizers which permit compensation for imperfect equalization (i.e., misalignment) that has occurred previously in the system. Among other things, mop-up equalization corrects for manufacturing deviations from the design value for the wire circuit and the repeaters.

The over-all equalization problem of any system involves such a large number of variables that no unified approach has been devised for its solution. Again, experience with previous systems frequently serves as a valuable starting point. Chapter 14 discusses the problem of equalization in more detail, as well as our next topic, regulation.

Regulation

Regulation compensates for changes in the loss of the transmission system with time. It can be thought of as a kind of automatic system equalization. There are two basic types of regulators:

a) Non-feedback regulators: In this type of regulator, something is measured which serves as an indication of system gain (loss), and a gain regulating network is adjusted accordingly. An example of such a regulator is the thermistor regulator used at some of the L3 repeaters. The thermistor is buried in the ground adjacent to the repeater. Measurement of its resistance change with temperature serves as an accurate measure of the corresponding change in attenuation of the preceeding section of cable. The thermistor, as part of a simple resistance network, can therefore be used to adjust the gain of the repeater as the ground temperature varies. The non-feedback regulator has the advantage of being simple, reliable, and relatively cheap. Its biggest disadvantage is that it is not self-checking. Whatever residual error it leaves is passed along to the next repeater. If the error is systematic it will accumulate in direct proportion to the number of repeater sections traversed. Herein lies the advantage of the second type of regulator, the feedback regulator:

b) Feedback regulators: In a feedback regulator the transmission is measured continuously by comparing a pilot signal with a reference. The difference, or error, is used to actuate a gain control in the repeater until the system is in a state of equilibrium as determined by the feedback around the control loop. An example is in the K2 system in which a pilot tone is used to control the current through a thermistor in a regulating network, which in turn adjusts the gain of the repeater. An example of how this type of regulator circuit prevents errors from accumulating in the system is given in Chapter 14. A problem associated with this type of regulator circuit is that of gain enhancement. Gain enhancement refers to the ratio of output pilot variation to input pilot variation, and frequently becomes a problem when, for example, the input pilot amplitude is varying at some frequency at which the feedback loop provides a small gain, rather than loss. The variation in amplitude is then enhanced as the pilot is transmitted through the system. This problem is also discussed in detail in Chapter 14.

The type of regulation system chosen for the system will depend largely on the type of signals to be transmitted and general economic considerations. In general, we can expect to use relatively simple equalization and regulation methods on a short-haul telephone message type system. As the system length increases, and the problem of misalignment becomes more serious, the complexity of the equalization and regulation plan will increase. Additional requirements are imposed when we then attempt transmitting TV, data, and other signals more sensitive to impairment than message.

Conclusions

We have ranged over a large number of subjects in this chapter. As we have seen, there are generally a number of ways to solve the various problems we encounter. We can transmit single or double sideband (etc.); we can reduce the crosstalk by using two cables, or by using compandors (etc.); we can use physical or equivalent four-wire transmission - at every point we have multiple options, and we can find an example of each option being used somewhere in the plant today.

When all the choices have somehow been made, however, we find that our floor-ceiling limitations begin to bind as soon as we try to get the most we can for our money. The basic problem is obvious in one sense but perhaps somewhat vague in another. To make it less vague, and to be specific as engineers, we ought to consider quantitatively some of the problems we have talked about so far. If we are to stay within reasonable bounds, we had better limit our problem, however. With this in mind, we consider in the next chapters an example which meets these specifications - optimizing a long-haul single sideband system which is free from the crosstalk restraint.

Bibliography

1. C. H. Elmendorf, et. el., The L3 Coaxial System; Bell System Technical Journal, Vol. 32, July 1953. (Monograph 2090).
2. R. S. Caruthers, et. el., The N1 Carrier Telephone System, Bell System Technical Journal, Vol. 30, January 1951, Pages 1-32, and April 1951, Pages 418 - 446. (Monograph 1859).
3. P. G. Edwards, et. el., The Type O Carrier System, Bell System Technical Journal, Vol. 31, July 1952, Pages 688 - 723 (Monograph 2004).

4. R. C. Boyd, et.el., A New Carrier System for Rural Service, Bell System Technical Journal, Vol. 36, March 1957, Pages 349 - 390 (P-Carrier System).

5. Transatlantic Submarine Telephone Cable System, Bell System Technical Journal, Vol. 36, January 1957, Pages 1 - 326 (Monograph 2710).

6. "Engineering Considerations" are given in the following Bell System Practices.

C System	AB25.130
J System	AB25.140.1
K1 System	AB25.150
K2 System	AB25.151
M1 System	AB25.171
M1A System	AB25.172
N1 System	AB25.190
O System	AB25.200
P1 System	AB25.250.0

TABLE I

MAJOR BELL SYSTEM AMPLITUDE MODULATED CARRIER SYSTEMS

	Long-Haul				Short-Haul						
	J2	K2	L1	L3	C5	M1A	N	ON1	ON2	O	P
Line Facility	O.W.	Cable	Coax.	Coax.	O.W.	O.W.		Cable		O.W.	O.W. or Cable
Method of Operation	(1)	(2)	(2)	(2)	(1)	(1)		(2)		(1)	(1)
Terminals	A	A	A	A	C5	M	N	O	O	O	P
Channels	12	12	600	1860	3	5	12	20	24	16	4
Sidebands	1	1	1	1	1	2	2	1	1	1	2
Transmitted Carrier	No	No	No	No	No	Yes	Yes	Reduced	Reduced	Reduced	Yes
Frequency Allocation (kc)											
Lowest Trans. Freq.	36	12	64	308	5	150	36	40	36	2	9
Highest Trans. Freq.	140	60	3096	8320	30	420	268	264	268	156	99
System Length (miles)											
Minimum	125	75	75	75	60	10		15		15	
Maximum	4000	4000	4000	4000	1000	50		200		150	
Approx. Repeater Spacing (miles)(3)	30	17	8	4	150	(6)		8		50	(7)
Approx. Repeater Gain (db)(4)											
Lowest Freq.	10	43	8	8	12	(6)		46		4	(7)
Highest Freq.	20	64	56	45	22			48		17	
Frogging	No	No	No	800 miles	No	No		Each Repeater		Each Repeater	No
Compandors	No(5)	No	No	No	No	No		Yes		Yes	Yes
Nominal Transmission Level (db) (at Repeater Output)	+17	+9	-10		+18						
Pilots	Yes	Yes	Yes	Yes	Yes	No		No		No	Yes(8)
Approx. Regulation Range (db) at each Repeater	50	16	26	8	32			10		45	12

NOTES:

(1) Equivalent 4-wire.
(2) Physical 4-wire.
(3) Sleet-area spacing given for O.W. systems.
(4) Repeater gain for O.W. systems assumes dry 104 mil 8-inch spaced line.
(5) Compandors are sometimes added for crosstalk and noise control, but are not part of the system terminals.
(6) System consists of terminals only; line loss between terminals is not to exceed 47 db.
(7) Line loss between repeaters not to exceed 25 db; maximum of 4 repeaters in tandem. (Tentative)
(8) Pilot transmitted from remote terminal.

Entries which are omitted call for more detailed consideration than can be given here.

Chapter 6

SYSTEM LAYOUT TERMINOLOGY

Definitions are given for terminology to be used in subsequent discussions of single-sideband amplitude-modulation telephone message systems.

In earlier chapters, we examined the environment in which transmission systems must operate, and the message channel objectives which in large part govern their design. In the immediately preceding chapter, we discussed, in broad general terms, some of the AM system problems which arise when we try to meet these objectives. In the following eight chapters, we shall discuss, quantitatively and in detail, the factors affecting the design and performance of long cable systems carrying single sideband, AM, telephone message signals. We shall confine the discussion to systems such as coaxial or submarine cable in which crosstalk (except that caused by intermodulation) can be ignored. We do this for reasons of simplicity - we are trying, in this text, to illustrate a method of approach which is applicable to all transmission system problems, and a simple illustration will best serve our purposes.

The first step in designing any transmission system is to lay out a mathematical framework which relates the performance of the components, and the ways in which imperfections accumulate in long systems, to the signal-to-noise objectives of the system. It should be emphasized that there is no unique system for a particular job, even when we have narrowed our inquiry down to ssb AM message systems. For example, one can trade repeater spacing for bandwidth, or one can allow greater or less margin against aging or performance uncertainties. Initially, therefore, preliminary calculations are made on a number of systems that seem likely to meet the requirements. Extrapolation from some earlier system having a basic design similar to the one under consideration may be helpful in selecting a good starting point. As the study continues, the field narrows and uncertainties are "firmed up". Finally, a system evolves that appears to be the best compromise between a large number of factors, including economics, relative feasibility and schedules.

The purposes of the preliminary calculations are several:

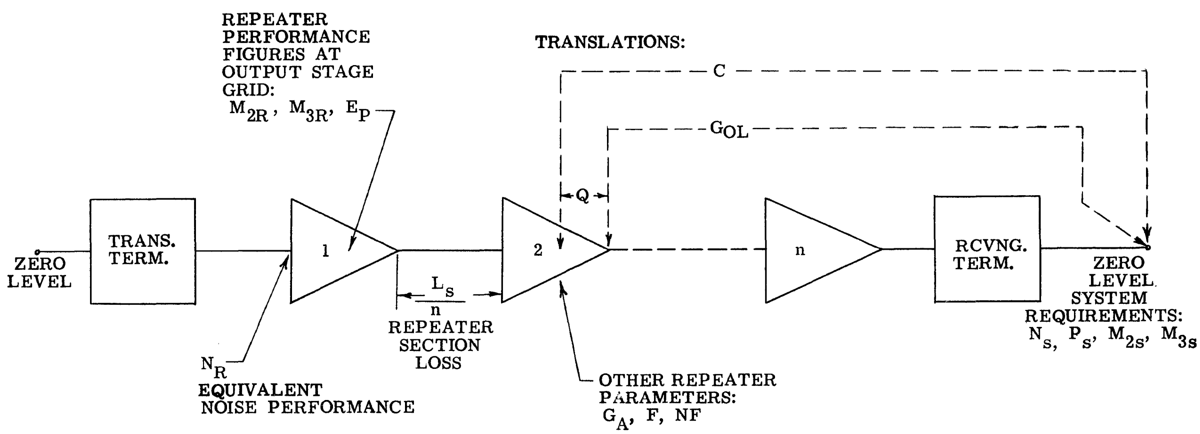
a) They provide a means for comparing the relative merits and costs of different system layouts.
b) They provide a quick answer to overall feasibility questions.
c) They point out areas where uncertain estimates must be "firmed up" by further study, and thus guide the allocation of development effort.
d) They set up a framework wherein design changes or refinements may be introduced or evaluated.

One usually starts the study by making an estimate of the important performance parameters. These are used to compute maximum allowable repeater spacings for bandwidths within the range of interest. In such preliminary computations one usually makes only very rough approximations for the effects of terminals, equalizing equipment, and other more or less subsidiary transmission equipment, which effects are small compared to the effect of the long repeatered line itself. (It must not be forgotten, however, that the form of the signal, i.e., single sideband AM, and the level at which it appears on the line is determined by the hardware in the terminals.) As the system design and development proceeds, the computations are refined, both by the inclusion of effects ignored during the first round, and by the use of more accurate values for performance parameters as they become available. In fact, the whole framework may have to be shifted because it is found that an important source of imperfection has been overlooked, and another, which was thought to be governing, has turned out to be negligible.

Layout and Terminology

To facilitate computations, a hierarchy of terminology has become established. It has grown through the years and, although it is the result of considerable thought and experience, it is not a unique terminology nor necessarily the best. As a result, it changes as old terms are deposed or modified and new ones are developed. The remainder of this chapter will be devoted to definitions of this terminology, with particular emphasis on the single side-band AM cable type of telephone system, although some of the quantities defined will be used in later discussion of FM radio relay systems. The meaning of these definitions will become more clear as they are used in subsequent chapters.

Figure 6-1 illustrates some of the symbols which have been found useful in analyzing one direction of transmission in a carrier transmission system.



Identification of Transmission System Design and Performance Parameters

Figure 6-1

In Figure 6-1 the signals appear at zero level as physically-separated telephone messages. Each message is modulated up into the carrier frequency spectrum by an AM modulator. We shall assume the use of single sideband suppressed carrier modulators, although some systems use other modulation arrangements. The signals are stacked frequency-wise and combined to produce a high-frequency signal composed of message sidebands at 4 kc intervals. This signal is transmitted through the cable and the repeaters and appears at the output. Here each signal is separated by frequency-selective filters and demodulated to produce physically separated messages at zero level. The filters, modulators, and demodulators are part of the terminal equipment and have been discussed in previous chapters. We will confine our discussion here to the high frequency line, ignoring any contribution of the terminals to the system noise performance. However, certain relationships between the voice signal (at zero level) and its carrier frequency sideband, as well as other signal properties, must be known. These properties and relationships will be covered in Chapter 12 where it will be seen that most of the statistical properties of the voice signal will apply directly to the carrier frequency sideband.

The important terms and symbols used to determine the system performance are listed below.

n: As shown in Figure 6-1, "n" is the number of repeaters operating in one direction of the system. This is one of the important parameters in the system design since the number of repeaters, which implies the repeater spacing, has an important effect on the signal-to-interference performance and on cost.

db: The original formal definition of the decibel is "10 log p_1/p_2", where p_1 and p_2 are in watts. It follows that the relationship can be used to compare voltages across, or currents through, equal impedances by taking "20 log" of the voltage or current ratio. It is also a convenient general practice to speak of voltage or current gains in "db", ignoring the fact that the voltages being compared may exist across different impedances. In this course we shall go further (though not beyond the bounds of common practice) and use the "db" as a mere computational device, like logarithms themselves. However, each term measured in db will be multiplied or divided by suitable constants in order to make it dimensionless.

dbv: The term dbv is defined by the equation

$$\text{dbv} = 20 \log_{10} \frac{e\ (\text{volts})}{\bar{e}\ (\text{volts})}$$

where $\bar{e}$ = 1 volt. The dbv, therefore, is a dimentionless way of specifying voltage relative to a reference voltage of one volt (rms, peak, or peak-to-peak are all used, in various contexts).

dbm: Similarly, as shown below, dbm is a dimensionless way of expressing power relative to one milliwatt.

$$\text{dbm} = 10 \log \frac{p\ (\text{milliwatts})}{\bar{p}\ (\text{milliwatts})}$$

where $\bar{p}$ = 1 mw.

dba: These have been defined in the preceding chapter.
level: It is suggested that they be reviewed.

L_S: The total attenuation in decibels between the end-points of the system is represented by "L_S". In preliminary calculations, the top frequency loss (being the highest loss) is frequently taken as the value of L_S. (In the final design the losses of filters and equalizers are included in L_S; and the loss-frequency characteristics of the filters, the equalizers, and the medium must be taken into account.) For example, in the frequency range of interest, the loss in decibels of coaxial cable varies very nearly as the square root of frequency. If the loss is 2 db per mile at one megacycle, it will be 4 db per mile at four megacycles. The characteristic impedance of this type of cable is almost purely resistive in the band of interest.

G_A: The insertion gain of the amplifier in db, measured between impedances equal to the cable characteristic impedance, is represented by "G_A". Ambiguity often exists in the use of the terms "amplifier" and "repeater". A repeater includes power separation filters and sometimes equalizers. The amplifier gain must compensate for the losses of these passive components of a repeater as well as the loss of the cable. For the purposes of computing G_A, the filters and equalizers are assumed not to modify the cable impedances. The amplifier and repeater gains are frequently taken as equal in early computations because passive component losses are temporarily neglected.

N_R: This is a term used to specify the random noise generated at one repeater. It is the noise power which one would have to apply at the input of a noisless amplifier of equal gain to get the noise observed at the output of the real amplifier when it is isolated from the rest of the system (but properly terminated). It is expressed in terms of noise power (in dbm) in a 3 kc band centered around the frequency of interest, "f", and is equal to the noise which the repeater delivers to the following cable section (in dbm in a 3 kc band centered at f) minus the insertion gain of the amplifier (in db at f). The determination of N_R is covered in Chapter 7.

$\underline{N_S}$: The total random noise allowed at the output of the system is designated "N_S". It is expressed in dbm in a 3 kc band at zero transmission level. Here by random noise we mean the noise which would be measured at system output with no signals applied to the system - i.e., N_S does not include "noise" arising from intermodulation.

$\underline{C}$: One of the important parameters in signal-to-interference analyses is the magnitude of signals transmitted. The magnitudes of telephone signals are usually defined (in statistical terms) at the transmitting toll switchboard, zero level. For reasons to be discussed in detail in a later chapter, it is often convenient to analyze system performance in terms of voltages at the grid of the output stage of the amplifier of a typical line repeater. "C" is a dimensionless conversion factor to tie these two points together, output grid to zero level. It may be defined as follows: Given a one volt rms sine wave from grid to cathode of the output stage of a line amplifier, the corresponding signal power at the zero db transmission level point is by definition C dbm. That is,

$$\text{dbv (grid)} + C = \text{dbm (zero level)}. \tag{6-1}$$

C may be a function of carrier frequency; in the initial stages of design, it is often assumed to be flat with frequency. It may be carefully shaped in the later, more refined stages.

If dimensions bother one, one may make C dimensionless (without changing anything we have said) by writing it:

$$C = 10 \log \frac{p_o/\overline{p}_o}{(e_g/\overline{e}_g)^2} = 10 \log \frac{\dfrac{e_o^{\,2}}{R_o} \Big/ \dfrac{\hat{e}_o^{\,2}}{R_o}}{(e_g/\overline{e}_g)^2}$$

where $p_o = \frac{e_o^{\,2}}{R_o}$ is the power in mw at zero level, e_g is the rms voltage, at the output grid,

$$p_o = \frac{\hat{e}_o^{\,2}}{R_o} = 1 \text{ mw.},$$

and $\bar{e}_g = 1$ v.

Therefore, C may be written

$$C = 20 \log \frac{(e_o \; \bar{e}_g)}{(e_g \; \hat{e}_o)}$$

This form of C is useful for computing distortion effects.

G_{OL}: While signal-to-interference analyses for an electron tube system are most readily made in terms of output grid voltages, transmission measurements in an operating system are most conveniently made at the output of a repeater. For such measurements to be meaningful, the gain or loss between repeater output and zero level must be known. This gain or loss is, in general, a variable with frequency and is a design parameter closely related to "C" defined above. The power gain in db from repeater output to zero level is "G_{OL}" as shown in Figure 6-1. If the system is to operate between zero level points, as in Figure 6-1, a loss of "G_{OL}" db must precede the high frequency line, and the transmission level at repeater output is $-G_{OL}$ db.

Q: The factor "Q" is the conversion of voltage at the output grid to power at repeater output. Given one volt rms from output grid to cathode of a repeater, the corresponding power delivered to the following cable section is, by definition, "Q" dbm. Therefore, in db:

$$C = Q + G_{OL} \tag{6-2}$$

E_P: This is equal to the maximum rms sine wave voltage that can be impressed between the output grid and cathode without causing the amplifier to overload, and is expressed in db with respect to one rms volt. What constitutes overload is discussed in Chapter 9.

P_S: When a system is designed to transmit a given number of channels, the statistical properties of the signal at zero level can be used to compute a peak signal voltage that will be exceeded only a small percentage of the time. For practical purposes this is the maximum voltage we must be prepared to transmit. It can be related to the magnitude of an equivalent sine wave signal at zero level. These relationships will be discussed in Chapter 12 The equivalent single frequency signal power which the system must be designed to carry without overloading is designated by "P_S". It is specified in db with respect to one milliwatt (dbm) at zero level.

M_2, M_3: These quantities express the non-linearity of the circuit of a single repeater, without feedback. They are defined in terms of the last tube of the repeater.

$$M_2 = 20 \log \frac{e_{2f}/\bar{e}}{e_f^{\,2}/\bar{e}^{\,2}} = 20 \log \frac{(e_{2f}\bar{e})}{(e_f^{\,2})} \qquad (6\text{-}3)$$

where $\bar{e}$ = 1 volt is used to make the expression dimensionless; and "e_f" is the rms value, in volts, of the sine wave signal, of frequency f, measured from grid to cathode of the output stage of the repeater. The term "e_{2f}" is the rms value, in volts, of the second harmonic grid-to-cathode signal which, applied to a distortionless but otherwise identical output stage, would produce the observed (or calculated) second harmonic distortion at repeater output. With corresponding definitions of terms,

$$M_3 = 20 \log \frac{(e_{3f}\ \bar{e}^2)}{e_f^{\,3}} \qquad (6\text{-}4)$$

Below overload, M_2 and M_3 are not functions of the signal level for devices which obey the power series law.

F: The feedback $|1-\mu\beta|$ in db. One of the principal reasons for using negative feedback is that unwanted modulation products due to non-linear input-output characteristics are suppressed. The amount of suppression for second order modulation products is given by F. Third order modulation products are not suppressed quite as much (see K_F, below). In a given design, "F" may be essentially flat or it may have a frequency characteristic; however, unless otherwise stated, "F" will be assumed to be flat.

K_F: Third order modulation is produced in a feedback amplifier even if the tubes' input-output characteristics have a third order term whose coefficient is zero. This is because fedback second order modulation combines with the fundamentals to produce third order products. The effective reduction of third order modulation is less, therefore, than the full amount of the feedback by some number of db which we shall define as K_F. An evaluation of "K_F" is given in Chapter 13.

M_{2R}, M_{3R}: From the above definitions it follows that, in terms of equivalent voltages at the grid of the output stage, the modulation coefficients of a single feedback repeater are

$$M_{2R} = M_2 - F \text{ db} \tag{6-5}$$

$$M_{3R} = M_3 - F + K_F \text{ db}$$

M_{2S}, M_{3S}: These terms represent the system requirements on second and third order modulation. They are defined, respectively, as the maximum allowable ratios (in db) of second and the third harmonic power to fundamental power at zero level when the fundamental is one milliwatt at zero level.*

H_x: System modulation performance, defined in Chapter 8 in the same way as the corresponding requirements (e.g., M_{2S}).

A_p, A_N, A_{2M}, A_{3M}: These will be used to signify margins on load carrying capacity, random noise, and modulation - the number of db by which the system could be degraded from assumed performance and still meet requirements. The proper choice of margin magnitudes for critical performance characteristics is one of the most difficult decisions the system designer is faced with be considered in more detail later.

It is important to distinguish N_S, P_S, M_{2S}, M_{3S}, and the margins from the rest of the symbols. The terms N_S, P_S, M_{2S}, and M_{3S} are requirements placed on the transmission, as measured at a zero level point at system output. The margins are safety factors added to the requirements to allow for variations in transmission with temperature, age, etc. The remainder of the terms describe the performance characteristics of the particular transmission system. The quality of transmission as computed from these terms should meet or exceed the requirements plus the margins.

These definitions will be used in the following chapters in which the performance to requirement relationships needed in laying out a system will be developed. In developing these relationships, we shall assume that both repeater performance and system requirements are known. Subsequent chapters will, in turn, discuss how system requirements are derived, and then how estimates of repeater performance can be made.

---

*Third harmonic being evaluated as if it added in phase from repeater to repeater in the same way that dominant third order products do, as discussed in Chapter 8.

Chapter 7

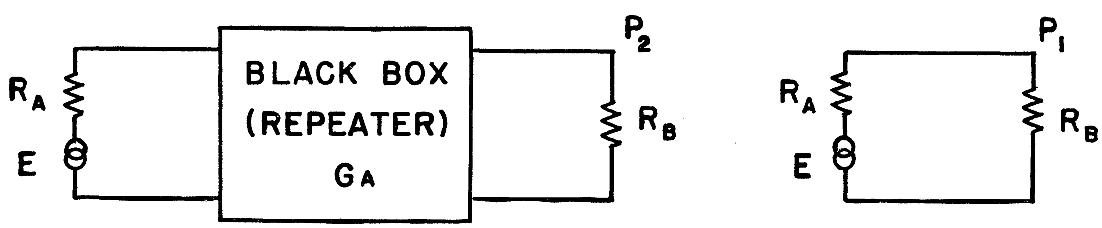
RANDOM NOISE

In theory, every element, every piece of wire, indeed, every electron, is a source of random noise. Fortunately, the problem of computing random noise is made manageable because most noise sources are negligible in themselves or because they are small compared to others in the environment in which they are located. The well known importance of tube noise in the first stage of an amplifier is a good example. Even though the later stages are equally noisy in an absolute sense, they are usually negligible compared to first stage noise because of the amplification of first stage noise and signal before the noises from the other stages are added.

Among the different types of sources of noise are thermal agitation, tube noise, battery noise, contact noise, corona effects, and certain atmospheric and switching disturbances. Thermal and tube noise sources are frequently of great importance and, furthermore, can be evaluated fairly precisely by analytic means. Some detailed consideration of these is therefore warranted.

Thermal Noise From Repeaters

Consider the circuit of Figure 7-1 in terms of the components of a high-frequency line as discussed in the previous chapter. Let the black box be a line repeater, of unspecified input impedance, and let R_A be the resistive termination which the repeater sees as it looks back into the preceding cable section, assumed to be infinitely long. Let R_B be the load presented to the repeater by the following cable section. Normally, for all practical purposes, $R_A = R_B$.



Thermal Noise in a Resistor

Figure 7-1

Before considering noise, let us pause for a moment to review the definition of insertion gain. Suppose that there is a voltage E in series with R_A, and that the black box is removed so that R_A and R_B are strapped together. Let the power delivered to R_B under these circumstances be P_1. Now with the black box in circuit, let the power delivered to R_B be P_2, caused by the signal source E. The ratio of P_2 to P_1, in db, is the insertion gain (call it G_A) of the black box. Note that we have not assumed that the input or output impedances seen looking into the black box have any particular value. They may be open or short circuits, or equal to the cable impedances - their value is immaterial to the definition of insertion gain.

Consider now R_A as a noise source. Theory tells us that it will produce noise as if it had in series with it an open circuit noise voltage

$$E_N = \sqrt{4KTBR_A} \text{ volts, rms} \tag{7-1}$$

where "K" is Boltzmann's constant, 1.38×10^{-23} joules per degree Kelvin, "T" is the temperature in degrees Kelvin, "B" is the bandwidth in cycles per second and "R_A" is in ohms.

If R_A were connected to an equal but noiseless resistor (e.g., R_B) it would therefore deliver a power of

$$P_N = (1/2\ E_N)^2 / R_B = KTB \text{ watts} \tag{7-2}$$

Since our main concern is with the noise generated in a telephone message band occupying a narrow (3 kc) slot in a wide carrier frequency spectrum, let us solve for P_N in terms of a 3 kc band; we find

$$P_N = 1.24 \times 10^{-17} \text{ watts} = -139 \text{ dbm, in 3000 cycles, } 27°C \tag{7-3}$$

For the condition that $R_A = R_B$, (but note we still place no restrictions on the input and output impedances of the black box) we see that P_N is analogous to the power P_1 defined in our discussion of insertion gain above. The noise power delivered to the following cable section by a single repeater must therefore be $(P_N + G_A)$ **dbm** assuming that the repeater contains no noise sources. In the terminology of the preceding chapter, P_N is the minimum possible value of N_R.

It might be pointed out that we do not know the noise power delivered by R_A to the unknown input impedance, Z_{in} of the black box. In the special case $Z_{in} = R_A$, the power delivered is, of course, P_N. In practice this may often be the case, but it is not one of our assumptions.

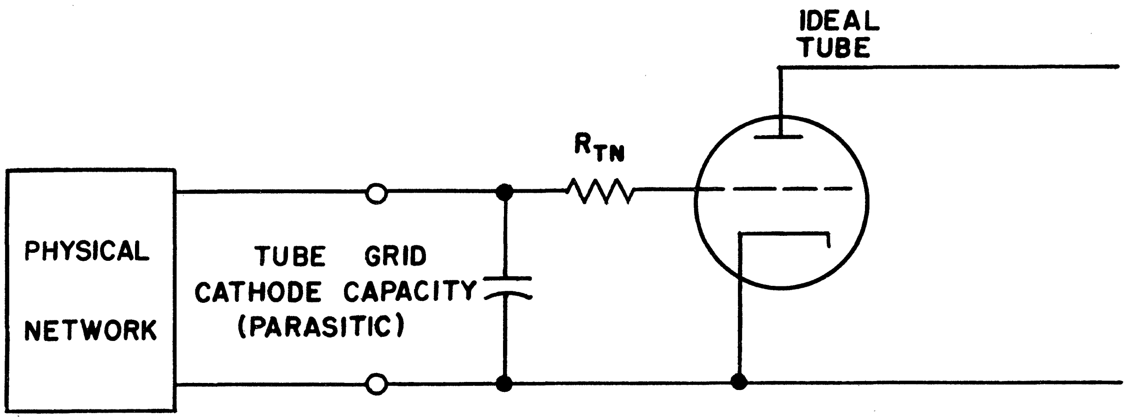
This method of specifying noise produced by a single repeater, in terms of an input generator and the insertion gain, is convenient when we want to vary the amplifier gain, as we shall when we vary repeater spacing. Usually, of course, the repeater itself will contain noise sources - for example, tube noise. We can take such internal sources into account by using the concept of noise figure.

<u>Noise Figure</u>

The concept of noise figure is essentially a very simple one: the noise figure of any black box is the difference, in db, between two powers - one, the noise power which the black box would deliver to the load impedance if the only noise source were the resistive component of the generator impedance; the other, the noise power actually delivered to the load impedance. For the general case, when the load and generator impedances are not equal, and other definitions of gain are substituted for the "insertion gain" concept, careful definitions of such quantities as "available noise power" are needed. In the simple cases which we shall consider, these complications do not appear.

Noise figure will, in general, be a function of frequency. Suppose, for example, that we have a wide band amplifier of flat insertion gain - say 30 db at all in-band frequencies - and that in a 3 kc slot at mid-band we find the noise delivered to the load to be -107 dbm. The noise figure at this frequency is then 2 db, since if the generator were the only source we would expect -109 dbm. At some other frequency, say upper band edge, we might find -105 dbm per 3 kc, corresponding to a noise figure of 4 db. Such variations of noise figure with frequency will normally occur because of the fact that even flat amplifiers will, in general, contain frequency sensitive networks. The term "spot noise figure" is sometimes employed to emphasize that the evaluation of noise figure must be made in a narrow slot around the frequency of interest.

The effect of internal noise sources within our black box can therefore be taken into account by saying that the noise delivered to the load will be $N_R + G_A$, where $N_R = -139$ dbm + NF (for the usual 3 kc band), the noise figure. N_R is thus a fictitious input noise source, which would result in the actual noise output if the repeater or black box were noiseless.



Equivalent Tube Noise Resistance

Figure 7-2

<u>Tube Noise</u>

Electron tubes, particularly those used in the first stages of the amplifiers, are a second important source of random noise. It is not possible to predict or compute tube noise with nearly the accuracy that is obtained in the case of thermal noise computations. Where accuracy is required, measurements must be used. However, when comparisons of one tube type with another are desired or when only approximate answers are required, computations are adequate. We find that the noise component of the plate current has the magnitude which we would compute for the circuit of Figure 7-2, where R_{TN} is a fictitious resistor, producing a noise voltage as given by Equation 7-1. The value of R_{TN} in ohms is given by the formula

$$R_{TN} = \frac{I_P}{I_C}\left[\frac{2.5}{g_m} + 20\,\frac{I_{SC}}{g_m^{\,2}}\right], \qquad (7\text{-}4)$$

where,

I = Current, in amperes (plate, screen or cathode according to subscript.)

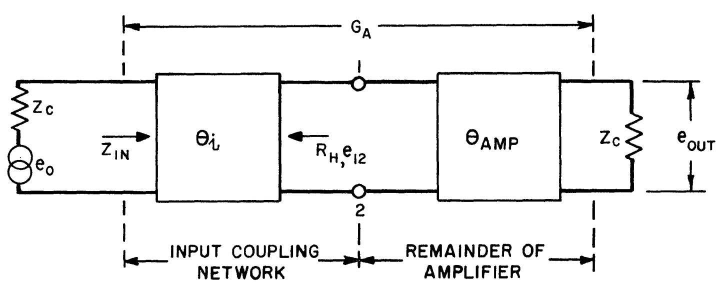
g_m = Transconductance in mhos.

The preceding expression for tube noise is fairly accurate for carrier frequencies where the little-understood 1/F noise, predominant at very low frequencies, has become negligible.

The importance of tube noise in the over-all random noise performance of an amplifier is dependent on the relative magnitudes of tube and thermal noise measured at a common reference point. This relationship is dependent on the type of tube used and the amount of gain between the point where thermal noise is generated (the input to the repeater) and the first stage. In many cases where the gain of the input network varies markedly over the band of interest, thermal noise will dominate in one part of the band and tube noise in another. The computation of the consequent net increase in N_R is discussed in Chapter 13.

Effect of Terminations

In order to avoid interaction effects between cable impedance and amplifiers, it is often desirable to design the amplifier so that it matches the cable impedances. (Such interaction effects would result in a ripply gain-frequency characteristic). The value of N_R will be affected differently depending on the method used to get the desired input impedance.



Effect of Repeater Termination on Noise Performance

Figure 7-3

Consider Figure 7-3. The Z_c impedances represent the preceding and following cable. The amplifier is divided into two parts at the input to the first stage, so that the θ_i network is the input coupling network, (usually a step-up transformer) and θ_{AMP} the remainder of the amplifier. The insertion gain of the amplifier, G_A, is to be kept constant. The resistive component of the high-side impedance is given by R_H, and e_{12} is the open circuit high side voltage. If e_o is a signal voltage, then, in db:

$$\theta_i = 20 \log \frac{e_{12}}{e_o} \tag{7-5}$$

$$G_A = 20 \log \frac{e_{out}}{1/2\ e_o} \tag{7-6}$$

Suppose that in the reference case, with no termination,* we have designed the transformer or input coupling network for high gain, thus making R_H as high as physical limitations permit, and that we have two important noise sources: a cable noise generator at the e_o position, and a tube noise generator which we can assume to be on the amplifier side of terminal 1. The cable noise can equally well be considered as originating in R_H, since for our present purposes R_H and e_{12} constitute a valid equivalent circuit for the combination of input coupling network and cable. The question is how the output noise $(N_R + G_A)$ will be affected by modifying the input network to make Z_{in} approximately equal to Z_c. (Since Z_c is predominately resistive, it will be satisfactory to consider that we want to make $R_{in} = R_c$).

Illustrative Example - Brute Force Termination

One method is to connect a resistor R_c across the input, or a resistor R_H on the high-side, of the input coupling network. If we put such a termination within the θ_i network, we find that:

a) θ_i is decreased by 6 db. Since the coupling network was already designed for maximum gain, we cannot get this gain back by increasing turns ratio or any other change in the network.**

b) θ_{AMP} must therefore be increased by 6 db to maintain G_A constant.

c) R_H will have half the value it had in the unterminated case, and hence the equivalent noise generator associated with it will be 3 db less than in the unterminated case.

d) At the output, the noise originally present in the unterminated case will be increased as follows: noise due to the tube will be 6 db greater, since the equivalent generator

---

*Since the input impedance of the θ_{AMP} box will usually be high, this means that Z_{in} will be a high impedance.

**Limitations of this sort on designs of physical networks are familiar from network design courses; they are briefly reviewed in Chapter 13.

is unchanged but θ_{AMP} is 6 db greater. "Cable noise" will be 3 db greater, since the generator (R_H) is decreased 3 db and the following gain (θ_A) is increased 6 db.

e) The spot noise figure has therefore been degraded somewhere between 3 and 6 db depending on whether cable or tube noise dominates at the frequency in question.

Hybrid Termination

Another method of termination is to connect the low-side of the coupling network as a hybrid; this decreases θ_i by 3 db, and leaves R_H unchanged, again taking the unterminated case as a reference. θ_{AMP} must then be increased by only 3 db, and since neither R_H or the tube noise generator is changed, the spot noise figure at all frequencies is degraded by 3 db, regardless of which source is dominant.

Other Termination Methods

The above two methods of getting good input impedance do not exhaust the list, of course. Some of the other alternatives, such as using shunt feedback to obtain the terminating impedance, are usually of academic interest in carrier systems where feedback design is dictated by the necessity to suppress intermodulation. Others, such as bridge structures, are useful in certain cases; hybrid feedback connections, in particular, appear to offer signal-to-noise advantages over purely passive-structure methods. A complete examination of these alternatives would be outside our scope.

Since providing terminations degrades the noise figure, each system design must include an evaluation of the need for terminations, weighing the degradation involved against the penalties associated with the interaction effects which would otherwise be present.

Other Sources of Noise

When noise sources other than thermal or tube noise are believed to be important, allowances must be made for them in one way or another. Power hum, battery noise, and contact noise are seldom important or controlling in transmission system design although occasionally they must be considered in connection with specific problems.

Atmospheric noise is frequently important in the design of open wire systems. Methods of analyzing such phenomena and making appropriate allowances will not be reviewed here since such noise is seldom controlling in broad band coaxial systems, our primary interest here.

In the design of systems where high voltages are involved, as in a coaxial cable system, the possibility of corona must be guarded against. The noise associated with corona can be so disturbing that its complete prevention is necessary.

Frequency Distribution of Random Noise

Random noise of the type under discussion here is usually "white" noise, noise which is distributed evenly over the whole frequency spectrum at the point where it is generated. The characteristics of transmission systems modify this flat distribution to give distributions of noise which are other than flat when considered with respect to the signals transmitted. Since we are concerned with signal-to-noise ratios, points in the system where the signals are flat with frequency are convenient reference points for measurement or computation. Hence, in telephone systems the transmitting toll switchboard is frequently made the reference point. We shall see later that the output grid of a line repeater is also often a convenient reference point for signal-to-noise calculations. As the amplifier circuit is developed and its output circuit characteristics become known, the noise distribution may be computed in terms of output grid values as

$$E_{RN} = N_R + G_A - Q \text{ dbv}, \tag{7-7}$$

where G_A and Q are defined in Chapter 6. E_{RN} is the voltage from grid to cathode of the output stage (expressed in db with respect to one rms volt) due to random noise in a 3000 cycle band at any frequency.

Addition of Noise, Multi-Repeatered System

In preliminary calculations leading to the design of a system, we assume the system to be ideal in the sense that the gains of all repeaters are identical and that these gains exactly compensate for the losses of the immediately preceding cable section. Hence the gain of each repeater section is unity or zero db. This is equally true of transmission from the output grid of one repeater to the output grid of the next. Therefore, the grid voltage due to noise generated in one repeater will be transmitted to the grid of the next repeater where its value will be the same as that at the first repeater. At the last repeater, there will be "n" such voltages, one from each of the "n" repeaters. These voltages, corresponding to noise power, a random phase phenomenon, will add at random. Hence, the total noise voltage will be $\sqrt{n}$ times larger than the voltage from one repeater. The total random noise may be expressed as a voltage at the grid of the last repeater as

$$E_{RNT} = N_R + G_A - Q + 10 \log n \quad \text{dbv}, \qquad (7\text{-}8)$$

and as a power at zero level,

$$P_{RNT} = N_R + G_A - Q + 10 \log n + C \quad \text{dbm} \qquad (7\text{-}9)$$

where "C" is defined in Chapter 6.

We may now write the first of a series of important system equations. The total random noise accumulated in the system "P_{RNT}", increased by an amount "A_N" db (margin for contingencies as defined in Chapter 6) must be equal to or less than the allowable random noise at zero level, "N_S" dbm. That is,

$$N_R + G_A - Q + 10 \log n + C + A_N \leqq N_S. \qquad (7\text{-}10)$$

Methods of combining this relationship with similar relationships for modulation, power, and gain-loss equivalences will be given in Chapter 10.

Illustrative Example - System Analysis

Problem: A repeater has been designed with the following parameters at the top frequency:

Noise Figure	7 db
Gain - Input to 3rd grid (G_A-Q)	64 db
C	9 db
Number of repeaters in the system	51
Length of the system in miles	2000

Find the system noise performance in dba at zero level.

Solution: Since we are attempting to determine system performance, neither requirements nor the margin allowed in designing the system enter into the problem. Using Equation 7-9:

$$P_{RNT} = (-139 + 7) + (G_A\text{-}Q{=}64) + 10 \log 51 + 9$$
$$= -42 \text{ dbm}$$

adding 82 db to get dba (since -82 dbm of thermal noise is 0 dba)

Noise, at zero level, = 40 dba

Bibliography

1 - Thermal Noise

S. Goldman; Frequency Analysis, Modulation and Noise. pp. 211-223, 381-403.

A. van der Ziel, Noise. pp. 8-17.

2 - Noise Figure

H. T. Friis. "Noise Figure of Radio Receivers", Proceedings of the I.R.E., July, 1944 - p. 419.

3 - Noise in Vacuum Tubes

S. Goldman, op. cit. pp. 223-233.

A. van der Ziel op. cit. pp. 88-119.

Note: These two references sum up the work of a number of authors notable among whom are B. J. Thompson, D. O. North, and W. A. Harris whose work is reported in the R.C.A. Review for January, April, and October, 1940, and January, April and July, 1941.

4 - Noise in Transistors

H. C. Montgomery, "Transistor Noise in Circuit Applications". Proceedings of the I.R.E., November, 1952, pp. 1461 - 1471, also Bell System Monograph 2063.

P. M. Bargellini and M. B. Herscher "Investigations of Noise in Audio Frequency Amplifiers Using Junction Transistors". Proceedings of the I.R.E., February, 1955, pp. 217-226.

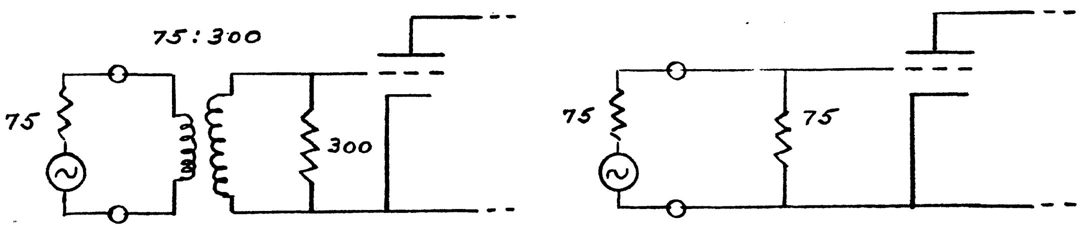
E. G. Nielsen, "Behavior of Noise Figure in Junction Transistors" - Bell System Monograph 2889.

Compare the relative noise outputs of a non-feedback amplifier for the conditions listed below. Find the noise figure for each condition. In each case the insertion gain is the same.

	Case 1	Case 2	Case 3
Signal source impedance	75 ohms	75 ohms	75 ohms
Input transformer ratio	75:300	none	none
Grid-to-ground termination	300 ohms	75 ohms	75 ohms
Input tube	tetrode	tetrode	triode
gm	.03 mho	.03 mho	.05 mho
Ip	20 ma	20 ma	36.5 ma
Ic	28 ma	28 ma	36.5 ma

Note: The material of Chapter 13, referred to in Chapter 7, is not required for the solution of this problem.

Case 1 Case 2, 3



Solution:

In each of the circuits shown, the signal source is a 75 ohm generator. The remaining elements, to the right of the terminals shown, are components of the repeater. The insertion gain would be measured between these terminals on the diagram and the load impedance at repeater output (not shown).

We are asked to find "relative noise outputs" and the noise figure for each circuit. The noise figure in each case can be found as follows:

Recall that the noise figure for a single black box is the ratio (expressed in db) of two powers: a) the actual noise output, b) the noise output one would obtain if the generator were the only source of noise. For any one of the circuits shown, this ratio is equal to the ratio, in db, of a) the actual noise voltage at the grid of the tube (including noise from the "equivalent tube noise resistance") and b) the noise voltage at the grid when the generator is the only resistance contributing noise.

Case 1

The generator and transformer, seen from the right, can be replaced by a 300 ohm resistance.

The equivalent tube noise resistance, from the formula given in the text, is 187 ohms. The total resistan ce seen by the grid is the parallel combination of two 300 ohm resistances in series with 187 ohms, or 337 ohms. The rms noise voltage associated with a 337 ohm resistance would be

$$E_T = \sqrt{4\ KTB\ R} = k\sqrt{337}$$

If the generator were the only source, we would have, in series with the 300 ohm resistance which is equivalent to the transformer and 75 ohm generator, a voltage of

$$E_G = k\ \sqrt{300}$$

The shunting effect of the 300 ohm termination causes only one-half this voltage to appear from grid to cathode. Comparing the actual and ideal noise voltages, we have

$$NF = 20\ \log \frac{E_T}{1/2\ E_G} = 6\ \text{db} + 10\ \log \frac{337}{300}$$

$$= 6.5\ \text{db}$$

Case 2

Using the same method, the actual noise voltage at grid is

$$E_T = k\ \sqrt{224.5}$$

and

$$E_G = k\ \sqrt{75}$$

Whence, again observing the shunting effect of the termination on the transfer of E_G voltage to the tube grid,

$$NF = 20\ \log \frac{E_T}{1/2\ E_g} = 6\ \text{db} + 10\ \log \frac{224.5}{75}$$

$$= 10.76\ \text{db}$$

<u>Case 3</u>

Same as Case 2, but now equivalent grid resistance is 50 ohms, and

$$E_T = k\sqrt{87.5}$$

$$NF = 6.67 \text{ db}$$

Since insertion gains are given as equal in all three cases, the noise outputs relative to Case 1 are merely the differences in noise figures, thus

Case	Relative Noise Output
1	0 db (reference)
2	+4.26 db
3	+0.17 db

The three cases have some reality - they represent possible input circuits for a video amplifier, using 436A tetrodes, with and without a step-up transformer, or a 437A triode (similar grid-cathode spacing and cathode structure) without a transformer.

Chapter 8

MODULATION DISTORTION

The interference commonly known as modulation noise arises from the non-linear input-output characteristics of transmission devices. Every element is non-linear, but in practice only coils, transformers and the active elements - tubes and transistors - tend to be important in this regard. Usually, the active devices dominate completely and only occasionally is it necessary to allow for the non-linear effects of coils or transformers.

This is fortunate, because the non-linear characteristics of coils and transformers are difficult to express mathematically and are not well controlled in practice. The non-linear effects in electron tubes, on the other hand, are well represented by a power series.* Although the effects are difficult to control in manufacture, they are at least fairly reproducible so that results can be predicted with fair accuracy.

The amount of modulation noise is directly related to the magnitude of the signals at the non-linear element. Since signals are, in general, highest in the output stage of an amplifier this is the source of most modulation noise. Other stages are usually neglected until the design is fairly well crystallized, at which time check computations are made to verify the earlier assumptions. If the earlier tube stages are significant contributors, correction factors can usually be applied.

Power Series: Three Frequency Input

Let us first consider electron tube non-linearity as described by the equation

$$i_p = a_o e_g^o + a_1 e_g + a_2 e_g^2 + a_3 e_g^3 \ldots \qquad (8\text{-}1)$$

This relates the instantaneous grid-cathode voltage "e_g" to the instantaneous plate current "i_p". The wanted signal is

$$i_p = a_1 e_g \qquad (8\text{-}2)$$

---

*Non-linear effects in transistors are not yet so well understood. It appears that whereas in electron tubes the transfer function is the only important contributor to non-linearity, in transistors the non-linearity of input and output impedances is important. Furthermore, the transfer function in transistors is inherently a function of frequency. These effects give rise to a more complex, frequency dependent non-linear behaviour which at present (mid-1956) has not been fully analyzed from the transmission system standpoint.

The higher order terms represent distortion due to the non-linear properties of the device. The values of the "a" coefficients will depend on the tube's characteristic and the operating point used. For a given tube at a given operating point, the coefficients are constants, provided the plate load impedance is constant vs frequency, or small compared to the tube's output impedance. The latter condition is usually met in systems of the type under discussion.

The output current components obtained from a device represented by Equation (8-1) when

$$e_g = A \cos \alpha + B \cos \beta + C \cos \gamma \qquad (8\text{-}3)$$

are given on Table 8-1. Attention should be called to the behavior of modulation product amplitudes as signals are raised or lowered in magnitude - for a one db increase in all three fundamentals, there will be a two db increase in the magnitude of all second order products, and a three db increase in all third order products. Relative to the fundamental, the second order products increase one db, the third order products by two db.

The relative magnitudes of the various products should also be noted; these are summarized in the lower left corner of the table. These relationships permit us to determine the magnitudes of important second and third order modulation products in a single repeater from the magnitudes of the harmonics.

Relationship of M_2, M_3 to a_2, a_3

For a repeater in which the output stage is the only significant modulation contributor, the modulation indices M_2 and M_3 are related to the power series coefficients as follows.*

$$M_2 = 20 \log \frac{a_2}{a_1\sqrt{2}} \qquad (8\text{-}4)$$

$$M_3 = 20 \log \frac{a_3}{2\,a_1} \qquad (8\text{-}5)$$

Since the power series coefficients are usually unknown, and can only be obtained from measurements which give the M's more easily than they give the a's, the relations are usually only of academic interest.

---

*In deriving this relationship, the compression term $\frac{3}{4}\,a_3\,A^3 \cos \alpha$ is ignored. The $\sqrt{2}$ and 2 factors result from converting peak to rms volts for the cosine wave signals assumed.

EXPANSION OF POWER SERIES FOR THREE TONE INPUT

APPLIED SIGNAL: $e_g = A\cos\alpha + B\cos\beta + C\cos\gamma$

OUTPUT: $i_p = a_0 e_g^0 + a_1 e_g^1 + a_2 e_g^2 + a_3 e_g^3$

TABLE, SHOWING FREQUENCIES TO BE FOUND IN OUTPUT, AND RELATIVE MAGNITUDES

TERM	DC	FIRST ORDER	SECOND ORDER	THIRD ORDER
0:	a_0			
1:		$a_1 A\cos\alpha + a_1 B\cos\beta + a_1 C\cos\gamma$		
2:	$\frac{1}{2}a_2(A^2+B^2+C^2)$		$\frac{1}{2}a_2\left[A^2\cos 2\alpha + B^2\cos 2\beta + C^2\cos 2\gamma\right] +$ $a_2 AB\left[\cos(\alpha+\beta)+\cos(\alpha-\beta)\right] +$ $a_2 BC\left[\cos(\beta+\gamma)+\cos(\beta-\gamma)\right] +$ $a_2 AC\left[\cos(\alpha+\gamma)+\cos(\alpha-\gamma)\right] +$	
3:		$\frac{3}{4}a_3 A(A^2+2B^2+2C^2)\cos\alpha +$ $\frac{3}{4}a_3 B(B^2+2C^2+2A^2)\cos\beta +$ $\frac{3}{4}a_3 C(C^2+2A^2+2B^2)\cos\gamma$		$\frac{1}{4}a_3\left[A^3\cos 3\alpha + B^3\cos 3\beta + C^3\cos 3\gamma\right]$ $+\frac{3}{4}a_3\left\{\begin{bmatrix} A^2B(\cos(2\alpha+\beta)+\cos(2\alpha-\beta)) \\ A^2C(\cos(2\alpha+\gamma)+\cos(2\alpha-\gamma)) \\ B^2A(\cos(2\beta+\alpha)+\cos(2\beta-\alpha)) \\ B^2C(\cos(2\beta+\gamma)+\cos(2\beta-\gamma)) \\ C^2A(\cos(2\gamma+\alpha)+\cos(2\gamma-\alpha)) \\ C^2B(\cos(2\gamma+\beta)+\cos(2\gamma-\beta)) \end{bmatrix}\right\}$ $+\frac{3}{2}a_3 ABC\left[\cos(\alpha+\beta+\gamma)+\cos(\alpha+\beta-\gamma) + \cos(\alpha-\beta+\gamma)+\cos(\alpha-\beta-\gamma)\right]$

NOTE:
OBSERVE THAT IF A=B, THEN $\alpha+\beta$ AND $\alpha-\beta$ PRODUCTS ARE 6DB GREATER THAN 2α, ALSO THAT $2\alpha-\beta$ (AND SIMILAR TERMS) ARE 9.6 DB GREATER THAN 3α. IF A=B=C THEN THE $\alpha+\beta-\gamma$ TERM, AND SIMILAR TERMS (BUT DO NOT CONFUSE WITH $2\alpha-\beta$ TYPE) ARE 15.6 DB GREATER THAN 3α. THE COMPRESSION, OR α COMPONENT ARISING FROM THE e_g^3 TERM, IS AT LEAST 9.6 DB GREATER THAN 3α AND MAY BE MUCH GREATER, DEPENDING ON THE NUMBER OF SIGNALS APPLIED; FOR THE 3 FREQUENCY INPUT GIVEN ABOVE, IT IS 23.5 DB GREATER.

Table 8-1

The following presentation, however, may help to clarify the M_2 and M_3 concepts, which will be used frequently in subsequent discussion.

Let the signal be: $e_g = A \cos \omega t$ (8-6)

For the tube characteristic, write: (ignore dc term)

$$i_p = a_1 e_g + a_2 e_g^{\,2} + a_3 e_g^{\,3} \tag{8-7}$$

Then (ignoring dc term arising from second order, and fundamental term arising from third order):

$$i_p = a_1 A \cos \omega t + \frac{1}{2} a_2 A^2 \cos 2\omega t + \frac{1}{4} a_3 A^3 \cos 3\omega t \tag{8-8}$$

$$= a_1 \left[A \cos \omega t + \frac{1}{2} \frac{a_2}{a_1} A^2 \cos 2\omega t + \frac{1}{4} \frac{a_3}{a_1} A^3 \cos 3\omega t \right] \tag{8-9}$$

The bracketed quantity can be considered an "equivalent" grid voltage -- i.e., that voltage which, applied to a distortionless, tube, would result in the observed output current. In Figure 1 this equivalent grid voltage is represented as arising from three generators at f, 2f, and 3f cps. Associated with each generator is a peak and rms value.

Peak Equivalent Grid Voltage	RMS Equivalent Grid Voltage, Expressed in Various Forms
A	$\frac{A}{\sqrt{2}} \qquad = e_f$
$\frac{1}{2}\frac{a_2}{a_1} A^2$	$\frac{1}{2\sqrt{2}} \frac{a_2}{a_1} A^2 = \frac{1}{\sqrt{2}} \frac{a_2}{a_1} \left(\frac{A}{\sqrt{2}}\right)^2 = \frac{1}{\sqrt{2}} \frac{a_2}{a_1} e_f^{\,2} \equiv m_2 e_f^{\,2} = e_{2f}$
$\frac{1}{4}\frac{a_3}{a_1} A^3$	$\frac{1}{4\sqrt{2}} \frac{a_3}{a_1} A^3 = \frac{1}{2} \frac{a_3}{a_1} \left(\frac{A}{\sqrt{2}}\right)^3 = \frac{1}{2} \frac{a_3}{a_1} e_f^{\,3} \equiv m_3 e_f^{\,3} = e_{3f}$

Fundamental, Second, and Third Harmonic Grid-to-Cathode Voltages for an Equivalent but Otherwise Distortionless Tube

Figure 8-1

We see, from the last two columns of the figure, that the definition of M_2 and M_3 given in Chapter 6 come about quite logically:

$$m_2 = \frac{e_{2f}}{e_f^{\,2}}\ , \text{ then let } M_2 \equiv 20 \log m_2 = 20 \log \frac{e_{2f}}{e_f^{\,2}} \qquad (8\text{-}10)$$

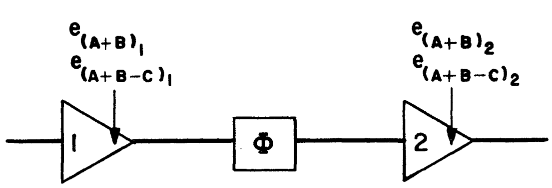
$$m_3 = \frac{e_{3f}}{e_f^{\,3}}\ , \text{ then let } M_3 \equiv 20 \log m_3 = 20 \log \frac{e_{3f}}{e_f^{\,3}} \qquad (8\text{-}11)$$

where the factors required to make M_2 and M_3 dimensionless can now be added, as in Chapter 6. It is important to remember when the above forms are used that the e's must be in RMS volts.

At this point we pause to point out two special cases; thinking about them now, while this is still fresh in our minds, will help us later. First, suppose the signal is one volt rms of fundamental. (By definition, this means the signal power at zero level is C dbm of fundamental.) In this case the equivalent grid voltages of second and third harmonic are m_2 and m_3 volts rms. The resulting powers of harmonic (ignoring contributions from other repeaters in the system) are $[C+M_2]$ dbm and $[C+M_3]$ dbm. Next, suppose the signal is lowered C db to be one milliwatt at zero level. The second harmonic power at zero level will drop 2C db to become $[M_2-C]$ dbm; third will drop 3C to $[M_2-2C]$ dbm.

System Modulation: Laws of Addition

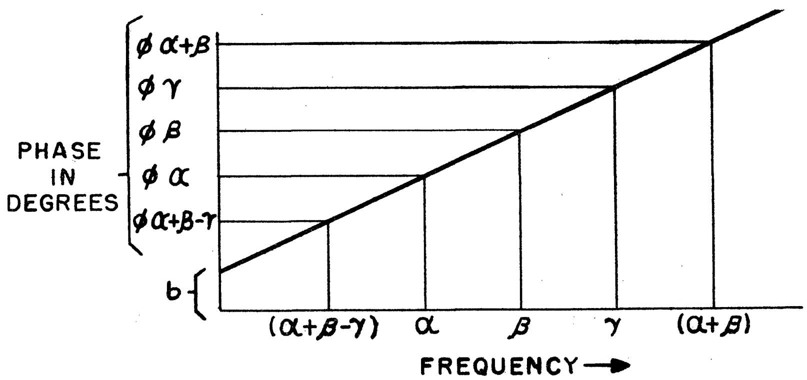
Thus far we have considered only the modulation arising in a single repeater. Let us now calculate how modulation products add from repeater to repeater in a system. In so doing, we must distinguish between the various types of product that can be generated; the law of addition is different for different types. Consider a circuit like that in Figure 2. Two repeaters are shown with a section of cable between them. The voltages are the "equivalent" modulation



Addition Of Modulation Products In Successive Repeaters

Figure 8-2

product voltages at output stage grids arising from non-linearity in that particular repeater. It is assumed that the phase-frequency characteristic from the output of the first repeater to the output of the second is linear as shown on Figure 3. In general this assumption is justified except near the edges of the transmitted band.



An Assumed Phase - Frequency Characteristic

Figure 8-3

To illustrate the laws of addition for modulation products, two types will be considered, "$\alpha+\beta$" and "$\alpha+\beta-\gamma$". These products will be generated in the first repeater where they are designated $(A+B)_1$ and $(A+B-C)_1$ and are transmitted to the second repeater.* The second repeater will generate similar products, $(A+B)_2$ and $(A+B-C)_2$. The objective is to compare the phases of $(A+B)_1$ and $(A+B-C)_1$ with the phases of $(A+B)_2$ and $(A+B-C)_2$ at the output of the second repeater where the products appear together and will add. These phases will determine the laws of addition.

For this analysis we may ignore magnitudes altogether. We may assume then that at the output of the first repeater, we have the signals

$$A_1 = \cos\alpha \qquad (8\text{-}12)$$

$$B_1 = \cos\beta \qquad (8\text{-}13)$$

$$C_1 = \cos\gamma \qquad (8\text{-}14)$$

and the modulation products

$$(A+B)_1 = \cos(\alpha+\beta) \qquad (8\text{-}15)$$

$$(A+B-C)_1 = \cos(\alpha+\beta-\gamma)$$

- -

*Writing $e_{(A+B)_1}$ would be more consistent, but it is easier to abbreviate this to $(A+B)_1$.

These signals and modulation products are transmitted to the output of the second repeater in Figure 2 where their phases have been shifted due to the transmission characteristic of the section of cable and one repeater. They appear at the output of the second repeater as:

$$A_{1\to2} = \cos\ (\alpha+\Phi_\alpha) \qquad (8\text{-}17)$$

$$B_{1\to2} = \cos\ (\beta+\Phi_\beta) \qquad (8\text{-}18)$$

$$C_{1\to2} = \cos\ (\gamma+\Phi_\gamma) \qquad (8\text{-}19)$$

$$(A+B)_{1\to2} = \cos\ (\alpha+\beta+\Phi_{\alpha+\beta}) \qquad (8\text{-}20)$$

$$(A+B-C)_{1\to2} = \cos\ (\alpha+\beta-\gamma+\Phi_{\alpha+\beta-\gamma}) \qquad (8\text{-}21)$$

Products $(A+B)_2$ and $(A+B-C)_2$ will be generated in the second repeater by the above A, B, and C signals. These products will be

$$(A+B)_2 = \cos\ (\alpha+\beta+\Phi_\alpha+\Phi_\beta) \qquad (8\text{-}22)$$

$$(A+B-C)_2 = \cos\ (\alpha+\beta-\gamma+\Phi_\alpha+\Phi_\beta-\Phi_\gamma) \qquad (8\text{-}23)$$

Now, let us examine the phases of $(A+B)_{1\to2}$ and $(A+B)_2$ given by Equations (20) and (22). By reference to Figure 3 we may write

$$\Phi_{\alpha+\beta} = m(\alpha+\beta) + b \qquad (8\text{-}24)$$

$$\Phi_\alpha = m\alpha + b \qquad (8\text{-}25)$$

$$\Phi_\beta = m\beta + b \qquad (8\text{-}26)$$

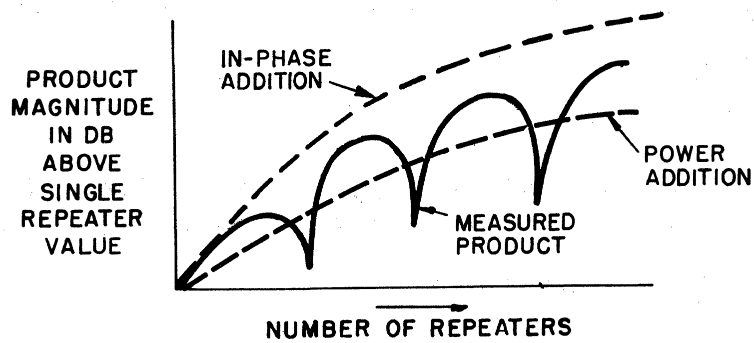
where m is the slope of the phase curve and b is its zero frequency intercept.

Substituting these values of Φ in Equations (20) and (22) we have at the output of the second repeater:

$$(A+B)_{1\to2} = \cos\ [\alpha + \beta + m(\alpha+\beta) + b] \qquad (8\text{-}27)$$

$$(A+B)_2 = \cos\ [\alpha + \beta + m(\alpha+\beta) + 2b] \qquad (8\text{-}28)$$

We see that these two products will be in phase with each other only when b = 0 or a multiple of 2π. This is seldom the case and, hence, in-phase addition of such products is not to be expected. When the phase characteristic is reproducible from section to section, the law of addition will be systematic. The vectors representing the modulation product voltages generated in the various repeaters will add by a systematic law so that the total magnitude of the product will vary considerably between limits somewhat greater than power addition and much lower than power addition. This variation will be observed for product magnitude versus frequency for a given number of repeaters* and versus number of repeaters for a given frequency of a modulation product. A possible addition phenomenon for the latter case is illustrated in Figure 4. (A plot of product magnitude versus frequency for a given number of repeaters would look very much like Figure 4 also.) Because of the averaging effects of large numbers of contributors and deviations from the ideal laws described above, and especially because there is no other general way of estimating the accumulation of modulation products along the line, power addition is usually assumed for "A+B" products.



Second Order Modulation Product Variation

Figure 8-4

Now, let us examine "A+B-C" products in a similar manner. Again, from Figure 3 we may write

$$\Phi_{\alpha+\beta-C} = m(\alpha+\beta-\gamma) + b \qquad (8\text{-}29)$$

$$\Phi_{\alpha} = m\alpha + b \qquad (8\text{-}30)$$

$$\Phi_{\beta} = m\beta + b \qquad (8\text{-}31)$$

$$\Phi_{\gamma} = m\gamma + b \qquad (8\text{-}32)$$

Substituting these values of phase in equations (21) and (23) we have

$$(A+B-C)_{1\to 2} = \cos[\alpha + \beta - \gamma + m(\alpha + \beta - \gamma) + b] \qquad (8\text{-}33)$$

$$(A+B-C)_{2} = \cos[\alpha + \beta - \gamma + m(\alpha + \beta - \gamma) + b] \qquad (8\text{-}34)$$

---

*Because the linearity of the curve of Figure 3 is only approximate, and slight deviations from linearity vs frequency will lead to large apparent changes in "b".

Here, the phases of the products are identical; in-phase or voltage addition is to be expected. It can be shown by these methods that only those products for which the coefficients of A, B, C, D... add algebraically to +1 will add in phase. Thus, if we can consider products higher than third order to be negligible, the only products which add in phase are:* A+B-C (in which 1+1-1 = +1) and 2A-B (in which 2-1 = +1).

System Modulation Performance

With the above relationships in mind, we are now ready to discuss the modulation behavior of a system composed of "n" identical repeaters in tandem, each of which exactly compensates for the loss of the preceding cable section at the frequencies of interest. It will be convenient to define an index of system modulation "H_x", where "x" is the product under consideration - thus "x" may be "2A", "A+B-C", etc., as follows:

> H_x: The power of the "x" type product, measured at the zero transmission level point at the output of a system, and formed by the intermodulation of fundamentals each of which is zero dbm at the zero level point, is H_x dbm.

Suppose we apply to the system a sine wave which results in a one volt rms signal from output grid to cathode of each repeater.** This is our fundamental. Then at the output of the first repeater we will have, by definition, that power of second harmonic which would result from a second harmonic signal of amplitude M_{2R} (in db relative to one volt rms) applied from the output grid to cathode of a perfectly linear but otherwise identical repeater. At the n-th repeater the corresponding second harmonic grid-cathode voltage will be higher by "10 log n" db, because of power addition thru the n repeaters. At the zero level point, by definition of "C", the power of the second harmonic will be

$$(M_{2R} + 10 \log n + C) \text{ dbm.} \tag{8-35}$$

Since the power of the fundamental signal was such as to produce one volt rms at the output grid, the power of the fundamental

---

*The in-phase addition of these important third order products will not, in general, hold for a 4000 mile telephone system, since usually the carrier signals will be frogged at intermediate toll offices - i.e., brought down to voice and then retransmitted in different carrier frequency slots in the broad-band spectrum. Frogging at 400-800 mile intervals is common.

**Except for the fact that we are now talking about a feedback repeater and therefore write M_{2R} instead of M_2, this discussion parallels the special cases we discussed in considering the rationale of M_2 and M_3.

at the zero level point is, by definition, "C" dbm. Assuming for convenience that "C" is positive, let us now decrease the applied signal by "C" db, so that the fundamental is now zero dbm at the zero level point. We thus decrease by "C" db the magnitude of the fundamental at each tube in the system. For every db decrease in fundamental amplitude, the second harmonic amplitude or power will decrease by two db. The new power of second harmonic at the output of the system will therefore be 2C db less than the original power, or

$$(M_{2R} + 10 \log n + C - 2C) \text{ dbm.} \tag{8-36}$$

Then, since we now meet the conditions for which H_x is defined,

$$H_{2A} = M_{2R} + 10 \log n - C \text{ dbm} \tag{8-37}$$

Had we considered an "A+B" product, recalling that for fundamentals of equal magnitudes such a product in a single repeater will be 6 db higher than a 2A product, that it varies two db per db variation in both fundamentals, and adds as power vs number of repeaters, we would have found, for the same system,

$$H_{A+B} = M_{2R} + 10 \log n - C + 6 \text{ dbm} \tag{8-38}$$

Similarly,

$$H_{3A} = M_{3R} + 10 \log n - 2C \text{ dbm} \tag{8-39}$$

$$H_{A+B-C} = M_{3R} + 20 \log n - 2C + 15.6 \text{ dbm} \tag{8-40}$$

Here we find "2C" instead of "C" because third order products vary three db if we change each fundamental one db; we find "20 log n" for products that add in phase vs number of repeaters, as against "10 log n" for products that add as power.

Output Grid as a "Flat-Level Point"

In general, for electron tubes, the coefficients of the power series will not be functions of frequency in the band of interest. If "F" and "K_F" are flat over the transmitted band, and "C" is also constant, then "H_x" will not be a function of frequency. Except where otherwise specified, we shall assume in the next few chapters that these conditions are met, in order to avoid complicating the analysis. The condition that "C" be constant in-band implies that for a 1000 cycle zero dbm signal applied at the transmitting toll test board, the output grid-cathode voltage at a line repeater will be the same regardless of the channel selected. By stretching the definition of "Level", we can then speak of the output grid as a "flat voltage-level point".

Performance-to-Requirement Relationships

Assuming that we have determined a requirement for second order modulation noise, and that this requirement is expressed in terms of permissible second harmonic of a single frequency, we can state that for a satisfactory system,

$$H_{2A} + A_{2M} \leqq M_{2S} \tag{8-41}$$

where A_{2M} and M_{2S} are defined as in Chapter 6.

Strictly speaking, the corresponding condition for satisfactory third order modulation performance should be stated in terms of a 2A - B or A+B-C product, since by virtue of their voltage addition these dominate the third order noise performance. Since this would lead to a cumbersome notation, we adopt the fiction that third harmonic adds in phase from repeater to repeater - that is,

$$H'_{3A} = M_{3R} + 20 \log n - 2C \text{ dbm} \tag{8-42}$$

and write

$$H'_{3A} + A_{3M} \leqq M_{3S} \tag{8-43}$$

The conditions of (41) and (43) may be restated in terms of repeater and system parameters as

$$M_2 - F + 10 \log n - C + A_{2M} \leqq M_{2S} \tag{8-44}$$

$$M_3 - F + K_F + 20 \log n - 2C + A_{3M} \leqq M_{3S} \tag{8-45}$$

Bibliography

1 - F. B. Llewellyn, Operation of Thermionic Vacuum Tube Circuits, B.S.T.J., Vol. 5, pp. 433-462, July, 1926.

2 - For some material on transistor non-linearity see: Design Principles for Junction Transistor Audio Power Amplifiers by D. R. Fewer, I.R.E. Trans. on Audio, pp. 183-201, Nov.-Dec., 1955.

Chapter 9

LOAD CAPACITY; GAINS AND LOSSES

Introduction: Load Capacity

The difficulties of defining load capacity, overload, and overload requirements are many. Overload mechanisms are at best only partly understood. The statistical properties of various types of signals are frequently complex. Hence, relationships between signal characteristics at overload and the performance characteristics of systems at overload have not been established in a completely satisfactory manner.

Our objectives in this chapter will be (1) to present descriptions of overload from several points of view and to describe circumstances under which each of the points of view is most useful, (2) to show how performance and requirements are expressed in terms of single frequency signal magnitudes, and (3) to develop a system equation which relates required load capacity to available system load capacity. A more detailed treatment of telephone signal characteristics and their relation to overload phenomena will be given in Chapter 12.

Overload in Feedback Amplifiers

To illustrate the overload phenomenon in feedback amplifiers, let us consider the second harmonic generated when a single frequency signal is transmitted. The second harmonic for small signals may be computed by

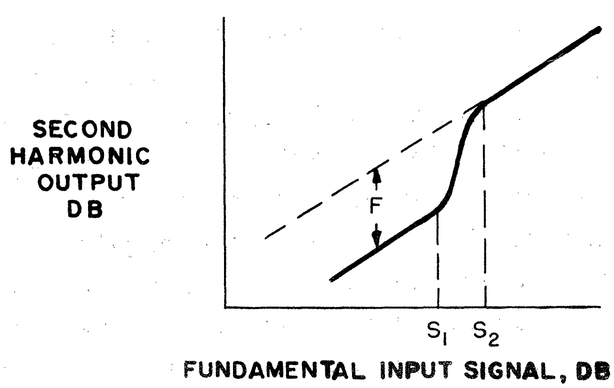
$$20 \log \frac{e_{2f}}{\bar{e}} = M_{2R} + 20 \log \frac{e_f^2}{\bar{e}^2} \qquad (9\text{-}1)$$

where e_f and e_{2f} are as defined in Chapter 6 and $\bar{e} = 1$ volt. This relationship is illustrated on Figure 9-1. The second harmonic increases 2 db for each db change in fundamental magnitude up to some value S_1, the upper boundary of the "small signal region". Between S_1 and S_2 is a transition region in which the amplifier appears to lose feedback. Above S_2 the curve again follows the 2:1 slope, the magnitude of the harmonic being approximately that which would be obtained for a non-feedback amplifier employing the same tubes. When there is a very large amount of feedback, say 60 db, the curve is very steep in the transition region between S_1 and S_2, i.e.: for a small increase in fundamental signal magnitude, there is a very large increase in second harmonics and, in fact, all modulation noise. Observation of this phenomenon has led to a "stone wall" concept of overload wherein the overload point is defined as a signal level below which the amplifier follows a power series law and performs satisfactorily, and above which the amplifier is completely inoperative.

The transition region can be interpreted as the point where the amount of feedback begins to decrease rapidly toward zero. Our previous equations indicate that for each db decrease in F we can expect approximately a db increase in modulation superimposed upon the normal 2:1 rise associated with increasing input. This explains the steeper slope in the transition region. The decrease in feedback is caused by the decrease in the through gain, μ, which occurrs for small signals. In terms of our power series model of modulation, the decrease in μ is caused by the fact that the compression term, instead of being negligible, has grown to the same order of magnitude as the linear one.

When there is a more modest amount of feedback, say 40 db or less, the departure from power series operation is found to be more gradual. The "stone wall" concept is hardly applicable when this is true. As signal levels are raised, modulation noise increases more rapidly than would be predicted by the power series, but there is no point at which the amplifier "cracks" and becomes suddenly inoperative.

This observation, that the modulation noise increases faster than the power series predicts but in a gradual fashion, leads to a definition of amplifier overload in terms of a departure of the modulation coefficients from constant values. The increasing noise may be expressed by increasing values of M_{2R} (or M_{3R} for third order modulation) with increasing signal magnitudes. The overload point is then defined as the point at which M_{2R} and M_{3R} depart substantially from constant values - say by one to three db.



Modulation Through Overload

Figure 9-1

A third criterion of amplifier overload is the change in gain that may be observed as signal levels are increased. The gain of the amplifier has been defined as "G_A" db. This is true only below overload. As signal magnitudes are increased through the overload region the gain will be measured as "$G_A - \Delta$" db where "Δ" increases as the signal magnitudes increase. Some value of signal magnitude corresponding to an arbitrarily small value of "Δ" may be defined as the overload point.

These three definitions or criteria of overload are all related. The "stone wall" and "change in modulation coefficient" concepts are both related to the apparent loss of feedback and differ only in the rapidity with which degradation takes place as signal magnitudes are increased. The "change in gain" criterion may be explained by increases in the compression terms of the power series representation of amplifier modulation. The increases in these terms are accelerated as signal magnitudes are raised and feedback is lost.

Another criterion of overload is that of drawing grid current. This criterion is sometimes used because there is some evidence (not well documented) that the reliability of tubes is affected by drawing grid current. Where reliability is of great importance, as in submarine cable systems, this criterion is used, since it appears to result in a somewhat more stringent requirement than results from the criteria described above, which are associated with loss of feedback.

Load Capacity Requirements

Just as the modulation performance of a telephone system is most conveniently described in terms of its single frequency performance, so is its overload performance described most conveniently in terms of single frequency loading. The details of how the varying telephone speech load can be related to a single frequency tone are given in a monograph by B. D. Holbrook and J. T. Dixon.* Their analysis will be reviewed in some detail in Chapter 12.

The way in which we apply Holbrook's and Dixon's results will depend on the repeater design. We may be designing a system whose amplifiers have very high values of feedback, thus making the "stone wall" concept of overload a valid one. Under these circumstances, one would be led to select from the load distribution curve a value which would be exceeded only some very small per cent of the time, say 0.01%, and specify that the system must be capable of transmitting a tone having a peak voltage equal to the 0.01% value of the telephone load voltage. Such a requirement would necessarily be based on the judgment that the system, thus operated, would be commercially satisfactory.

---

*Monograph B1183 - "Load Rating Theory For Multi-Channel Amplifiers".

However, if the system under study had a value of feedback such that the "increase in modulation coefficient" concept were valid, less conservative engineering of the overload characteristics would probably represent good design practice.

Other Overload Considerations

Sometimes a system is designed to carry signals which are more easily defined with respect to overload than are speech signals. The load capacity requirements in a television transmission system are generally more easily defined than in a telephone system. In the course of design, care must always be exercised so that some relatively obscure design change does not radically change system performance. Predistortion of television signals, for example, may make radical changes in the transmitted waveform; the peak factor may easily be increased enough to exceed the load capacity of the system.

It is frequently necessary to take into account, when considering overload phenomena, the necessity for transmitting tones which may contribute markedly to system load. Pilots and carriers are obvious examples.

System Performance-to-Requirement Relationship

Assume we have determined that to be satisfactory for our purposes, the system must be able to deliver at zero level a single frequency tone of "P_S" dbm without overloading, and that the amplifiers will just carry, without overloading, a sine-wave signal* which is "E_p" db with respect to one volt rms at the output grid. Then we may relate performance to requirements by writing

$$E_p + C - A_p \geq P_S \qquad (9\text{-}2)$$

where the quantities have the definitions given in Chapter 6.

Gains and Losses

For flexibility, and to achieve optimum signal-to-interference and overload performance, the gain of each repeater must be designed to equal the loss of the preceding section of cable as closely as practicable.

- -

*Here again we observe that the analysis is greatly simplified because we assume the output grid-cathode to be a "flat voltage-level point".

Since the loss-frequency characteristic of the transmission medium is not flat, the repeater gain-frequency characteristic must be shaped to match it if the loss of the section is to be zero db at all frequencies. This may be accomplished by shaping the gain-frequency characteristic of the amplifier itself, or by building out with equalizers the loss of the cable section to be flat with frequency or by some combination of these techniques. Over-all signal-to-interference performance suffers least when such equalizer losses are held to a minimum.

Perfect compensation at each repeater for the loss of the cable cannot be achieved, however. Deviations from the ideal must be allowed to accumulate through a number of repeaters, to be corrected for at terminals or at equalizing repeaters. The optimum spacing of equalizing repeaters is an important system design problem, which will be discussed in more detail later.

If we ignore these imperfections, however, we can say that since each repeater section is to be designed as a zero db loss section, the total loss of the cable, which we may define here as "L_C" db, must be equal to the total gain of "n" repeaters or "nG_R" db. The gain of the repeater is made up of the insertion gain of the amplifier "G_A" minus the insertion losses of whatever passive elements are associated with the repeater such as power separation filters, equalizers, etc. Let these losses equal "L_E" db and to simplify the analysis, assume that "L_E" db is the same at each repeater. Then, we may write

$$n(G_A - L_E) = L_C \qquad (9\text{-}3)$$

$$\text{or, } nG_A = L_C + nL_E.$$

The right side of this equation represents the total loss of the system, "L_S" db. Hence, we may write

$$nG_A = L_S \text{ db} \qquad (9\text{-}4)$$

This is the last of the important system equations needed for the design and analysis problems considered in the next chapter.

ILLUSTRATIVE EXAMPLES - CHAPTER 9

For certain feedback repeater, we find the following

a) The transmission from output grid to load is:

freq. kc	Q db
200	10
300	12
600	16

b) Feedback is flat vs frequency

c) For 0 dbm of a 300 kc fundamental delivered to the load we find the second harmonic power in the load is -60 dbm. For 0 dbm of 200 kc fundamental, third harmonic power is -80 dbm.

d) The repeater overloads when delivering to the load a power of +15 dbm of 600 kc (which is the top transmitted frequency).

Find M_{2R}, M_{3R} and E_p.

Solution:

a) -60 dbm second harmonic load power at 600 kc would correspond to an equivalent grid-cathode 600 kc signal at the output tube of $(-60-Q_{600}) = -60 -16 = -76$ dbv.
The fundamental associated with this second harmonic distortion is 0 dbm load power, or at the output grid, a signal of $(0-Q_{300}) = 0 -12 = -12$ dbv.
A convenient way to find M_{2R} is to increase the output grid fundamental signal to 0 dbv and inquire what the equivalent second harmonic signal at the grid would then be. Increasing the fundamental signal 12 db increases the second harmonic in the load (or in terms of equivalent grid voltage) by 24 db.

Hence: $M_{2R} = -76 + 24 = -52$ db

b) -80 dbm third harmonic load power at 600 kc corresponds to: $(-80 - Q_{600}) = -80 -16 = -96$ dbv equivalent third harmonic output grid-cathode signal.
The 200 kc fundamental associated with this third harmonic distortion is 0 dbm load power, or at the output grid, a signal of $(0 -Q_{200}) = 0 -10 = -10$ dbv.
To find M_{3R}, increase the output grid fundamental signal 10 db to 0 dbv.

Increasing the fundamental signal 10 db increases the third harmonic 30 db.

Hence: $M_{3R} = -96 + 30 = -66$ db

c) At overload:

$$E_p = 15 \text{ dbm} - Q_{600} \text{ db}$$

$$= 15 - 16 = -1 \text{ dbv}$$

Chapter 10

SYSTEM LAYOUT AND ANALYSIS

The basic relationships developed in the preceding chapters may be used (a) in laying out a new system, and (b) in analyzing system performance.

New System Layout

It has been shown that in specifying the "transmission plan" of a new system, the following conditions (10-1) to (10-5) must be satisfied. Respectively, the first four formulae merely say that we must not have excessive random (thermal) noise, or excessive second order, or third order, modulation, and that the system must be able to carry the maximum signal without overloading. The terms on the left specify the performance of devices and the way in which imperfections add. The terms on the right are not performance, but system requirements to be met. The final formula - that the repeater gains must make up for the system losses - is an obviousity included for use in later manipulation of the formulae.

$$N_R + G_A - Q + 10 \log n + C + A_N \leqq N_S \qquad (10\text{-}1)$$

$$M_2 - F + 10 \log n - C + A_{2M} \leqq M_{2S} \qquad (10\text{-}2)$$

$$M_3 - F + K_F + 20 \log n - 2C + A_{3M} \leqq M_{3S} \qquad (10\text{-}3)$$

$$E_p + C - A_P \geqq P_S \qquad (10\text{-}4)$$

$$nG_A = L_S \qquad (10\text{-}5)$$

The conditions expressed by (10-1) to (10-5) with the A's set equal to zero, must all be satisfied over the transmission band throughout the life of the system. Since F and E_p will decrease as tubes age, and other effects (e.g., misalignment, discussed in the next chapter) will, in the course of time, cause system performance to be degraded, the A's must be non zero and positive for a new system.

The selection of proper values for A's is extremely difficult, but analysis and performance optimization on the basis of the above equation will not be more accurate than the precision with which the proper A's can be determined. The sort of considerations which lead to the choice of values for the margins are given in Example 2 at the end of this chapter.

Discussion of Terms

In multi-channel telephone systems, the intermodulation products formed by many fundamentals gives rise to a noise-like type of interference. In any given channel this modulation noise adds to the thermal noise on a power basis to give the total noise heard by the subscriber. Usually

it is this effect, rather than the production of single-tone interferences, which determines the requirement on permissible system non-linearity - i.e., which sets the values of M_{2S} and M_{3S}.
Obviously a one db increase in M_2, for example, will increase by one db the magnitude of every second order product, assuming that other terms remain unchanged.* This would increase by one db the modulation noise caused by second order products falling in a particular channel. Similarly any other change in the value of one of the terms of the left-hand side of (10-2) or (10-3) is a change in system performance that will affect the modulation noise, db for db.

The conversion factor C bears an inverse relationship to the magnitude of the signals transmitted in the system. When C is large, the voltages at the grid are small, and vice versa. Thus, when all other terms are known equation (10-1) establishes a maximum value of C which can be used (that is, a minimum magnitude of signal that can be transmitted) without exceeding random noise requirements. On the other hand, when all other terms are known, equations (10-2) and (10-3) establish a maximum signal magnitude (a minimum value of C) that can be used without exceeding modulation requirements. Equation (10-4) similarly sets a lower limit on the value of C - if C is too small, the maximum voltage that can be applied to the grid will not result in enough power at the zero level point to satisfy the P_S requirement.

The behavior of total system noise as a function of C is illustrated by Figures 10-1 and 10-2 Suppose that in a given system, third order modulation noise is negligible. This might be the case if the system is a narrow-band one so that not many third-order products fall in the channel of interest, or if the third-order modulation coefficient of the active devices is extremely small. In such a case, if second-order products are not negligible, we would find that the total system noise was given by the addition of thermal and second-order noise. Figure 10-1 shows these components and their sum as a function of C. For a given repeater design, Q is fixed; the only way we can change C is by changing the value of G_{OL} by adjustments in the terminals at each end of the high frequency line. Suppose we increase levels on the high frequency line by one db, by decreasing the loss of a pad in the transmitting terminal. Thermal noise at line repeater outputs will not be changed, of course, but we have to put in a db of pad at the receiving terminal to get back to zero level (or some other reference level) there - so the thermal noise at receiving terminal decreases one db. Similarly second order modulation products at line repeater output will increase 2 db, but after passing through the

---

*M_2, it will be recalled, measures device non-linearity in terms of a single-frequency modulation product.

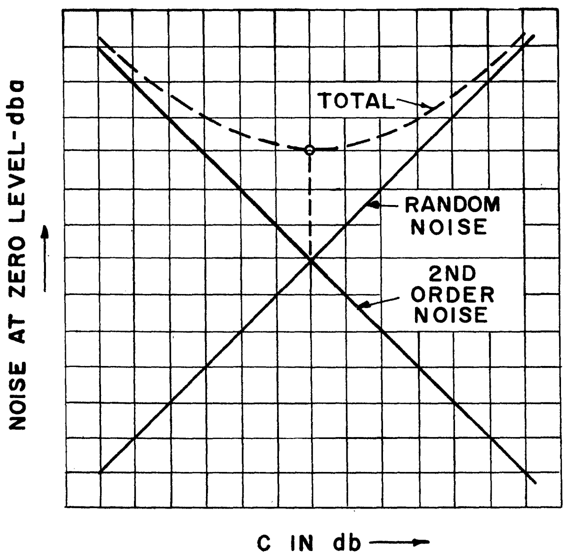


Figure 10-1

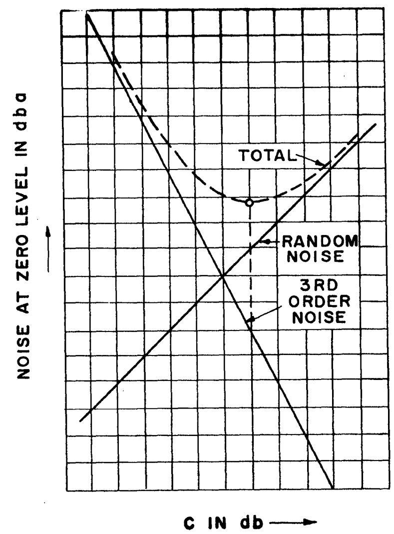


Figure 10-2

Optimum Noise Relationship
(Each Division is 1 db)

added one db of loss in the receiving terminal, appear at zero level only one db higher. The noise at zero level is minimum when the components are equal; the total is then 3 db greater than either component.

If, on the other hand, third-order modulation noise is important and second-order is negligible, the optimum S/N performance is obtained when thermal noise is 3 db above modulation noise, giving a total 1.8 db above thermal, as shown by Figure 10-2.

Before levels and repeater spacings have been set, we cannot know the relative importance of thermal and modulation noise sources. The relationships discussed above, however, are helpful in making initial estimates of the values to be assumed, as discussed below.*

Maximum Repeater Spacing and Optimum Levels

To illustrate one way in which the conditions given by (10-1) to (10-5) might be used, suppose that we have tentatively selected:

a) the components of the system - cable, repeater configuration, tubes, etc.

b) the bandwidth and number of channels to be transmitted, and hence the total top-frequency cable loss L_S for a system of given length.

---

*The additions of voltages and powers are among the most common operations that must be carried out in the process of system design. This would present no particular problem if it were not for the fact that the voltages and powers are usually expressed in some sort of logarithmic units - dbv, dbm, dba, etc. It is laborious to convert from db to volts or milliwatts, add, and reconvert to db. Adequate accuracy may usually be obtained by the use of curves which are given on Figures 10-5 and 10-6.

Assume also that we have estimated or computed, by methods to be discussed in subsequent chapters:

c) the repeater performance parameters: N_R, M_2, M_3, F, K_F, E_p, and Q;

d) the system requirements: N_S, M_{2S}, M_{3S} and P_S.

Under these circumstances, we can use formulae (10-1) to (10-5) to determine, for whatever margins we may choose, the maximum repeater spacing (minimum value of "n") for which we can satisfy the system requirements, and deduce the transmission level (value of C) which must be used.

We begin by solving (10-5) for G_A, and substituting this value for G_A in (10-1). We then solve (10-1) for C, and substitute this value of C in (10-2), (10-3), and (10-4). Rearranging terms, and writing M_{2R} and M_{3R} for (M_2-F) and (M_3-F+K_F), we obtain

$$\frac{L_S}{n} + 20 \log n \leqq (N_S - N_R) + (M_{2S} - M_{2R}) - (A_{2M} + A_N) + Q \qquad (10\text{-}6)$$

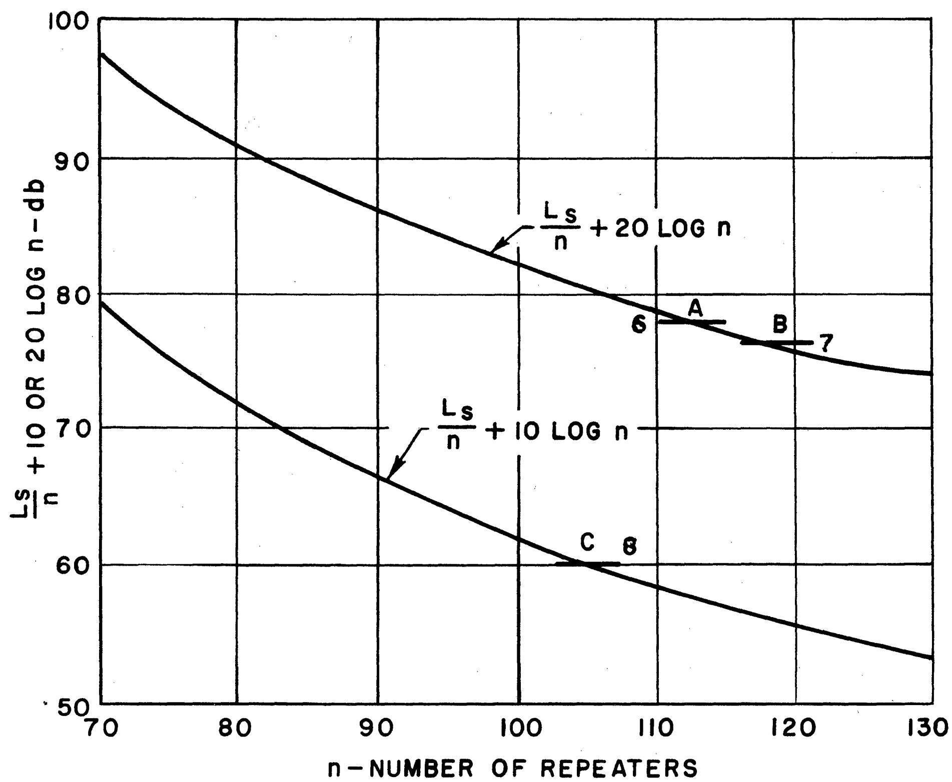
$$\frac{L_S}{n} + 20 \log n \leqq (N_S - N_R) + \left(\frac{M_{3S} - M_{3R}}{2}\right) - \left(\frac{A_{3M}}{2} + A_N\right) + Q \qquad (10\text{-}7)$$

$$\frac{L_S}{n} + 10 \log n \leqq (N_S - N_R) - (P_S - E_p) - (A_N + A_p) + Q \qquad (10\text{-}8)$$

We can think of these three formulae as representing three ceiling-floor limits - the common floor is random noise; the ceiling in (10-6) is second-order modulation, in (10-7) it is third-order modulation, and in (10-8), overload. Usually one of them will be of primary importance in determining the maximum repeater spacing that can be used, and the optimum signal levels.

Since in general it will be most difficult to meet requirements in the top transmitted channel, the repeater spacing is generally set by using the top-frequency value of L_S. The performance of the system at other frequencies will meet requirements with greater margins. Figure 10-3 illustrates the determination of "n", the number of repeaters, for a particular case. Here the abscissa is "n", and the ordinate is the value of the left-hand sides of (10-6), (10-7) and (10-8). The right-hand sides of three formulae are assumed constant versus "n", and their values are given as the horizontal lines A, B, and C respectively. In each case the intersection of the horizontal line with the appropriate curve gives the minimum number of repeaters for which the particular condition [e.g., n = 117 for (10-7)] is satisfied. Using a greater number of repeaters would then merely give greater margins.

In this case, then it is (10-7) which expresses the condition of primary importance in setting repeater spacing. If instead of n = 117 we used n = 113, we would indeed just meet the second-order modulation requirement, have more margin than we thought necessary on overload, but would fail to meet third-order modulation requirements. The largest value of "n" found must obviously be taken as governing. The only value of "C" which will result in the margin chosen, for the governing ceiling, can then be found by substituting this value of "n" in equation (10-1), since (10-1) is common to equations (10-6, 7, and 8). Thus the level as well as repeater spacing has been determined.



Determination of Maximum Repeater Spacing

Figure 10-3

Value of N_S

Ambiguity exists as to the value of N_S, the system random noise **requirement**, to be used in evaluating the right hand sides of formulae (10-6) to (10-8). If modulation noise is negligible, the first two formulae will turn out to be of no interest, and in evaluating (10-8) we would get the right answer by assigning all the noise allowance to thermal noise, and using the corresponding value of N_S. A little consideration of the process we are engaged in should persuade us that finding the minimum value of "n" will be expedited if in evaluating the various formulae we proceed as follows:

1) In (10-6), use a value of N_S which yields 3 db less than the total tolerable noise, and use a value of M_{2S} corresponding to second order modulation noise equal to thermal noise.
2) In (10-7), set N_S to correspond to 1.8 db below total tolerable noise, and M_{3S} to correspond to a third order modulation noise magnitude 4.8 db below the total.
3) In (10-8), allocate all the tolerable noise to thermal sources.

Methods of relating M_{2S} and M_{3S} to the corresponding permitted values of modulation noise in dba are discussed in Chapter 12.

If it turns out that only one of the three formulae is of real interest, giving a value of "n" for which the others are satisfied with large margins, we will have found the correct value of "n". If, on the other hand, two or even all three of them are significant - i.e., if both second and third order modulation noise are important, for example - then we must use judgment in selecting, with hind sight, a more sophisticated allocation of the total noise requirement.

Necessity for Successive Approximations

The only difficulty with this approach is that we have blithely assumed that we knew a number of quantities that we in fact do not know. Usually at the start of a new system design, we know (or can get someone's educated guesses about) the characteristics of the cables, electron tubes, and other component devices which are available or which could be developed if need can be shown. We do not usually know, however, what bandwidth we are to transmit - rather it is up to us to determine what bandwidth we can get for various repeater spacings (or vice versa) using available components. Since we do not know the bandwidth, we do not know the top-frequency cable loss, or the system modulation or load-carrying capacity requirements, these being functions of bandwidth. Similarly, as long as spacing and bandwidth are undetermined we do not know what feedback, transmission from output grid to cable, or noise figure we can obtain in the repeaters. Yet we need these quantities in order to determine repeater spacing. Finally, the margins we should use can hardly be well specified until we have analyzed a transmission plan and a repeater design for uncertainties and for sources of deviations from ideal performance.

Under these circumstances, it is necessary to use a trial-and-error approach. Guided by some knowledge of what has been achieved in other systems with other components, we arbitrarily choose a cable, bandwidth and repeater-spacing.* A preliminary repeater design can then be roughed out, and estimates made of repeater performance. Using arbitrary values for margins, we can then use equations (10-6, 7 and 8) to find the maximum repeater spacing. This may turn out to be substantially greater or less than the spacing we assumed, calling for a revision of repeater performance estimates. When our computations have converged on a solution, we can repeat the process for other bandwidths with less waste motion. On the basis of economics and other less tangible considerations, a choice of the optimum design can then be made.

The whole process, however, is one of continual re-examination and re-consideration. As the results become known, it may become obvious that development of new devices is needed; revisions of the original choices of margins may call for a revision of the previous spacing and bandwidth decisions. One can hardly over-emphasize the statement that system objectives and design, repeater design and requirements, and component device design and requirements are all inter-dependent.

Analysis of System Performance

Compared to the problem of designing unknown parts into a system which is to meet unspecified objectives, the analysis of the performance of a given system is straight-forward. For this purpose, equations similar to (10-1) to (10-4), but using some new terms defined below, can be written:

$$N_R + G_A - Q + 10 \log n + C + A'_N = N'_S \qquad (10\text{-}9)$$

$$M_2 - F + 10 \log n - C + A'_{2M} + 6 = H_{A+B} \qquad (10\text{-}10)$$

$$M_3 - F + K_F + 20 \log n - 2C + A'_{3M} + 15.6 = H_{A+B-C} \qquad (10\text{-}11)$$

$$E_p + C - A'_p = P'_S \qquad (10\text{-}12)$$

---

*Sometimes the cable is already determined.

Where:

A'_N is the random noise penalty due to misalignment.*

A'_{2M} is the second-order modulation penalty due to misalignment.

A'_{3M} is the third-order modulation penalty, due to misalignment, evaluated for in-phase addition.*

A'_p is the effect of misalignment in bringing the total grid swing nearer to the overload point, evaluated for the repeater having the highest positive misalignment. Exact computation of all these penalties is difficult except in the simplest cases, since for a broad-band signal we must take into account the fact that misalignment is usually a function of frequency.

P'_S is the power, in dbm at zero level, of the maximum amplitude sine wave which can be transmitted without overloading the system.

Equations similar to (10-10) and (10-11) can be written for A-B and 2A-B products, and would be used for a broad-band analysis of system performance.

It should be noted that here we are determining systems performance, without asking what requirements it should meet, so that whereas N_S was a requirement, N'_S is a performance value, and similarly for other terms.

From the values obtained for H_x, the modulation noise can be obtained for various channels and types of products using methods discussed in Chapter 12. These can be combined with random noise (translated into dba) to obtain a curve of signal-to-interference for the transmitted channels. Figure 10-4 illustrates the results for a particular system.

Output Grid as Reference Point

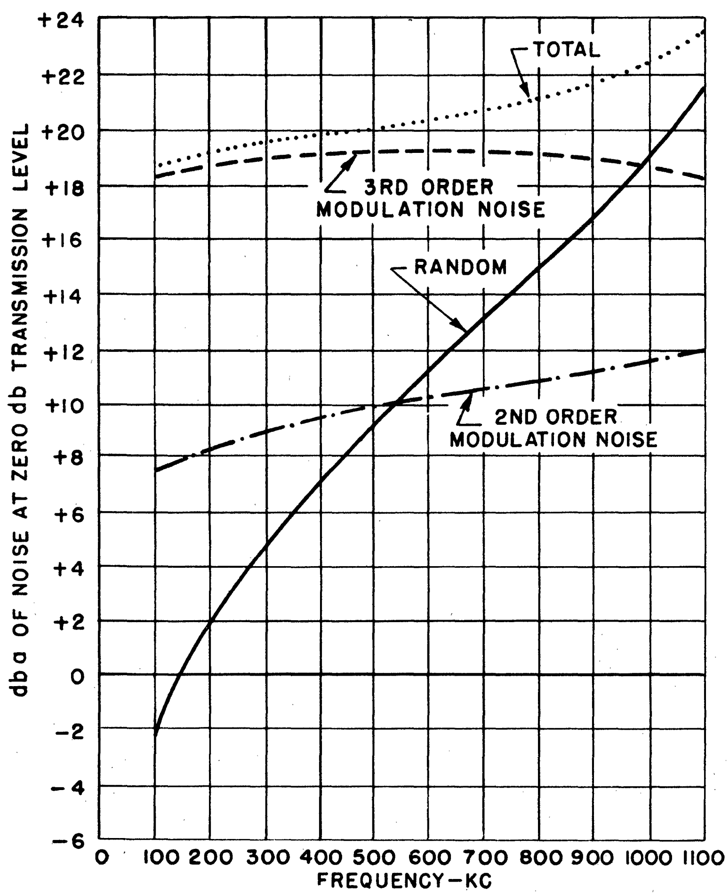
There are circumstances under which a system designer or analyst cannot conveniently work with results expressed in terms of noise at zero level. Suppose, for example, that we are going to transmit simultaneously a group of telephone channels and some other signal such as television. In such a case we need some common reference point.

---

*Or aging, or other degradation. The evaluation of misalignment penalties is discussed in Chapter 11. It might be mentioned that in general, if misalignment gives a positive random noise penalty, it will give a negative modulation noise penalty - i.e., an advantage, if the required equalization is done, as usual, at the far end of the line section.

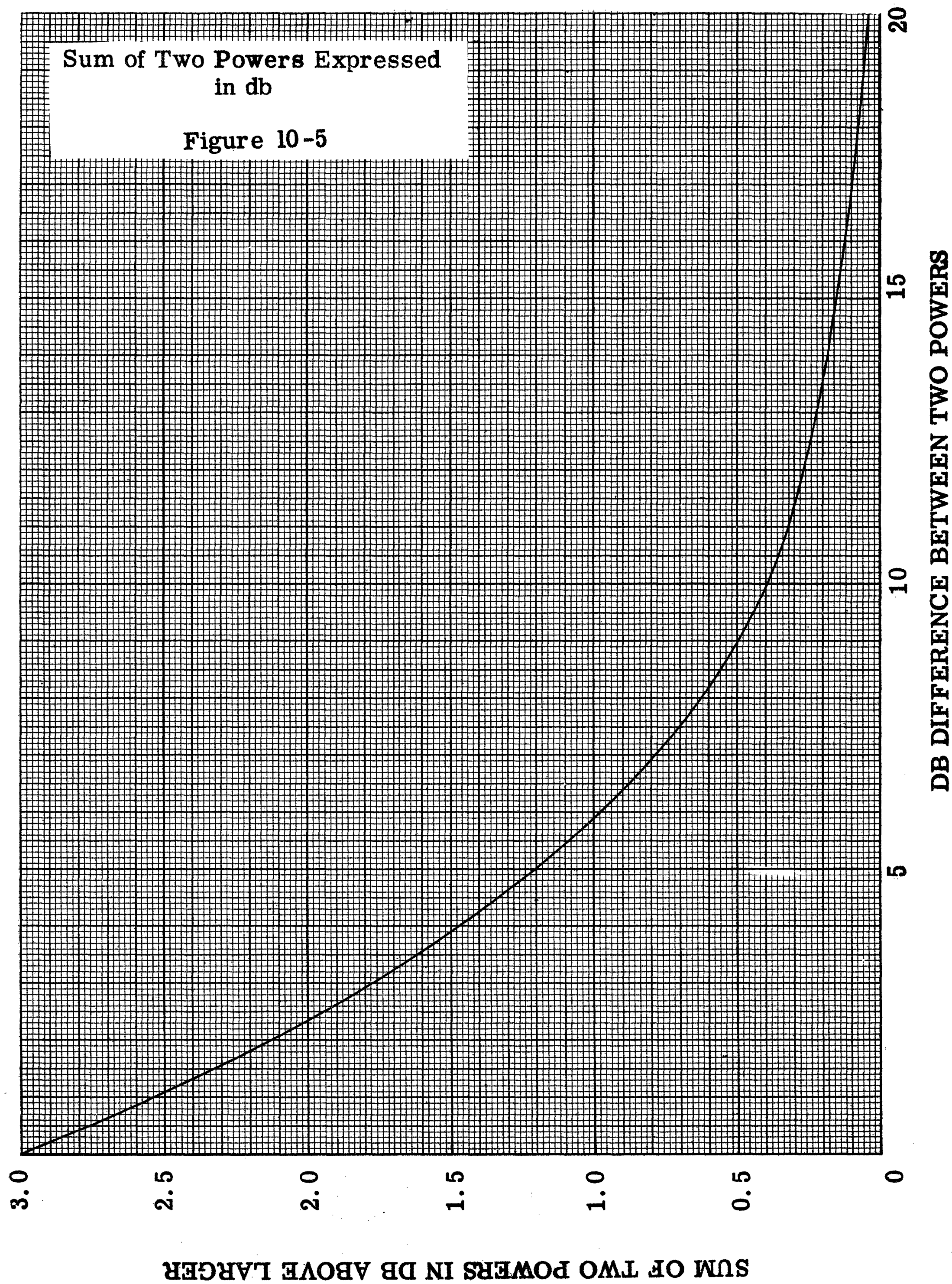
"Zero transmission level" will not serve - the magnitudes of the telephone and television signals are not fixed relative to each other at any common physical point in the plant. Indeed, usually the problem is to decide what their relative magnitudes in the repeaters should be, in order to get optimum system performance.

In such a case a convenient reference point is the output grid of the last line repeater. We can "translate" telephone message requirements and performance from zero level to output grid using C as a conversion factor. Cross-modulation effects between the two services, as a function of their relative levels at this point, can then be computed and compared with requirements. A full treatment of this process, and the definitions and terminology involved, is beyond our scope here.



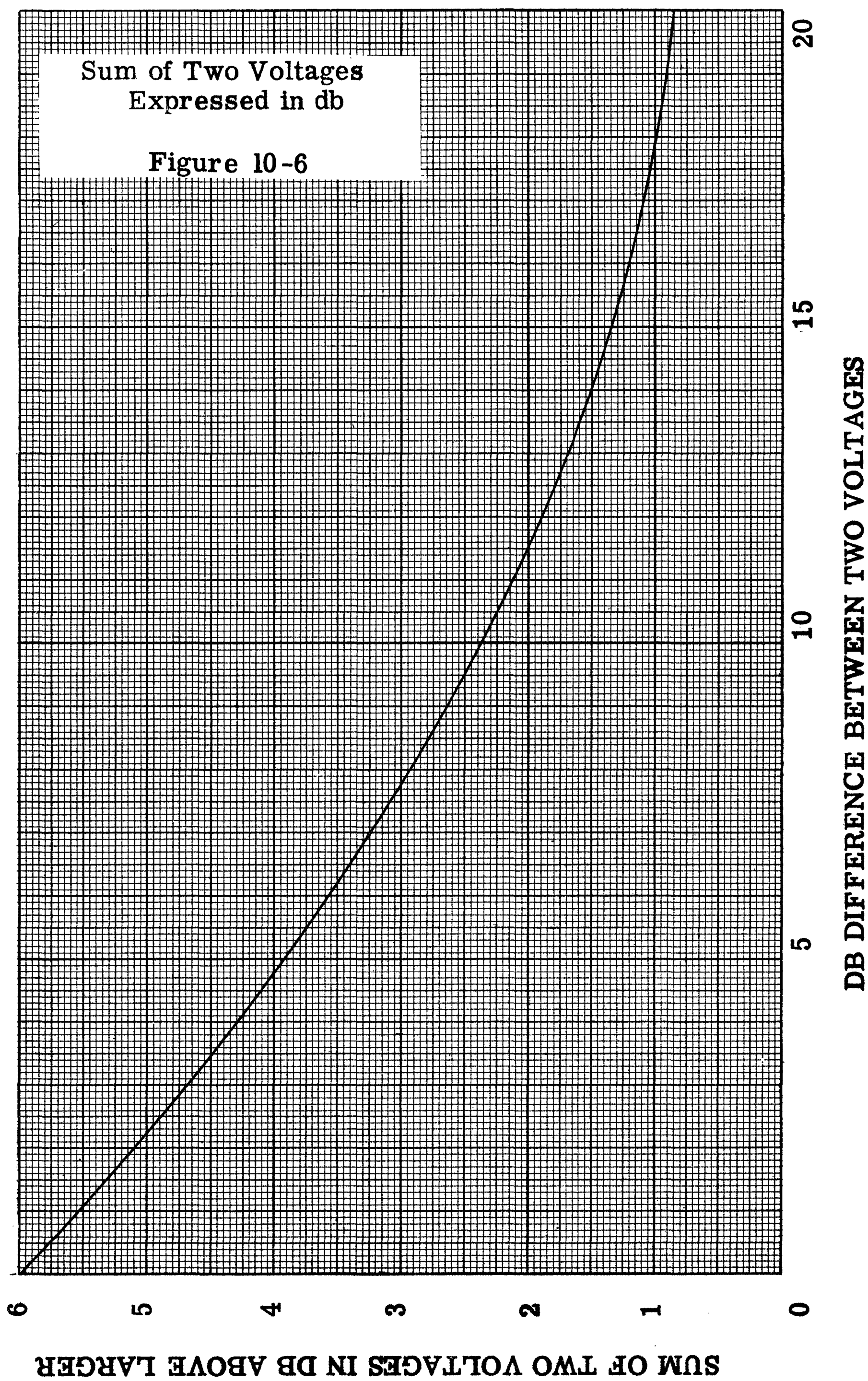
Noise vs. Frequency From Various Sources in a Particular 2,000 Mile System - Computed

Figure 10-4



Sum of Two Powers Expressed in db

Figure 10-5



Sum of Two Voltages
Expressed in db

Figure 10-6

ILLUSTRATIVE EXAMPLES - CHAPTER 10

The following conditions hold for a particular transmission system:

Levels on the high-frequency line have been set to optimize the ratio of signal to total noise. The total noise at zero level at receiving terminal is found to be +32 dba. This is made up of second order modulation noise and thermal noise. Third order modulation noise is negligible as a contributor to this total. The system is designed to operate well below overload. Noise and modulation in the terminals is negligible.

Now a change is made in repeater design which degrades the noise figure of each repeater by 4 db. Everything else in the system and repeater performance remains unchanged. Assuming that third order noise remains negligible,

a) What change would you make in levels on the high frequency line to re-optimize the signal to total noise?
b) What does the total noise at zero level then become?
c) What is the accompanying change in the negligible third order noise?
d) What is the minimum number of db by which second order non-linearity would have to be improved in order to permit achieving (by suitable re-adjustment of levels) the original system signal to total noise performance.

Solution:

a) If the total noise was +32 dba after S/N had been optimized, then thermal and second order noise must each have been +29 dba. The 4 db degradation in noise figure would make these contributors, before any change in levels,

Thermal	+33 dba
Second Order	+29 dba

To re-optimize, we should readjust levels to make these contributors equal, or

Thermal	+31 dba
Second Order	+31 dba

which can be done by increasing levels on the line by 2 db.

b) The total noise at zero level is then +34 dba.
c) Third order noise at zero level increases 4 db.

d) If levels on the line are raised another 2 db, we can make thermal noise at zero level +29 dba as it was originally. With no improvement in second order non-linearity, second order noise would then be +33 dba. A four db improvement in linearity would restore the original S/N performance (+29 dba for each source, +32 dba total). This is the minimum number of db improvement in linearity which will permit us to get a +32 dba total.

Illustrative Example #2

In the top frequency channel of a particular 2000 mile submarine cable system we find that the only important sources of modulation noise are the A+B and A+B-C products caused by the intermodulation of talker speech signals. We also find that the modulation noise caused by these products is related to H_x, the non-linearity performance of the system by the following equations. (For convenience, define W_x as the noise in dba at zero level at the receiving terminal caused by x-type intermodulation products).

$$W_{A+B} = H_{A+B} + 79 \text{ dba}$$

$$W_{A+B-C} = H_{A+B-C} + 74.4 \text{ dba}$$

The total noise in dba at zero level is to meet a requirement of +35 dba.

Find values of N_S, M_{2S} and M_{3S} to use in equations 10-6, 10-7, and 10-8 in the first round of the process of setting optimum signal levels and minimum repeater spacing.

(Hint: From a comparison of the definitions of M_{2S} and M_{2A}, it can be seen that if a system just meets second order modulation noise requirements with no margin, then H_{2A} and M_{2S} are equal by definition).

Solution:

The total noise is to be +35 dba. For equation 10-6 allocate:

+32 dba to random noise

+32 dba to second harmonic

$\therefore N_s = -82 + 32 = -50$ dbm

And for a system which just meets second order harmonic requirements:

$$32 \text{ dba} = H_{A+B} + 79 \text{ dba}$$

$$H_{A+B} = -47 \text{ db}$$

$$H_{2A} = -47 - 6 = -53 \text{ db}$$

$$= M_{2S}$$

For equation 10-7, allocate:

+33.2 dba to random noise

+30.2 dba to third harmonic noise

$\therefore N_s = -82 + 33.2 = -48.8$ dbm

And for a system which just meets third order harmonic requirements:

$$30.2 \text{ dba} = H_{A+B-C} + 74.4$$

$$H_{A+B-C} = -44.2 \text{ db}$$

$$H_{3A} = -44.2 - 15.6 = -59.8 \text{ db}$$

$$= M_{3S}$$

For equation 10-8 allocate:

Total noise to N_s

$$N_s = -82 + 35 = -47 \text{ dbm.}$$

Illustrative Example #3

In the example above, and in the example appended to Chapter 9 values were found for

M_{2R}	-52
M_{3R}	-66
E_p	- 1
N_S	-50, -48.8, -47 dbm
M_{2S}	-53
M_{3S}	-59.8

for use in Equations 10-6 10-7 and 10-8. Values were also given for Q and other parameters.

Given also the following: if it is not to overload excessively on peaks of the telephone speech load, the system must carry a sine wave test signal of +17 dbm at zero level. The noise figure of the repeaters is 3 db. The cable loss, which is 1.5 db per mile at 600 kc, can be taken as the only important loss in the system.

The above repeater performance values are assumed not to be functions of the insertion gain of the repeaters. They will, however, be affected by tube aging, new tubes are assumed above. Since we want the system to meet transmission objectives after the tubes have aged, we must put in margins for aging. The following values are tentatively selected:

A_N	3 db
A_{2M}	12 db
A_{3M}	12 db
A_p	3 db

Questions:

1. For a transmission plan which to a first approximation just meets these objectives, what is the transmission level in db at the output of each repeater?
2. For this transmission plan, what is the total noise in dba at zero level at system output when the system is new? (Ignore trivival contributors.)
3. What margin against overload is obtained? By how much can each noise source be increased (the others remaining fixed) without making total noise greater than +35 dba?
4. If the transconductance of each tube decreases 3 db because of aging, resulting in a 9 db decrease in feedback (assuming a single-loop three-stage repeater) and a 3 db decrease in transmission from output grid to load (no change in NF or G_A).
   a) what will be the noise in dba at zero level at system output?
   b) what will be the new answers to the questions of Part 3 above?
5. If repeaters cost $75,000 apiece, what is the dollar value per two way system of
   a) improving noise figure by 3 db?
   b) increasing transconductance of output stage by a factor of 1.4?
   c) obtaining 3 db more feedback?
   d) improving third order modulation of the repeaters by 3 db?
   e) decreasing cable loss to 1.4 db/mile.

   (Take these improvements one at a time, not cumulatively).

Solution:

Evaluating the right-hand sides of formulae 10-6, 10-7 and 10-8, we have

$N_S - N_R + M_{2S} - M_{2R} - A_{2M} - A_N + Q$ (10-6)

$-50 + 136 - 53 + 52 - 12 \quad - 3 \quad + 16$

$= 86$

$N_S - N_R + 1/2\ (M_{3S} - M_{3R}) - 1/2\ A_{3M} - A_N + Q$ (10-7)

$-48.8 + 136 + 1/2\ (-59.8 + 66) - 6 - 3 \quad + 16$

$= 97.3$

$N_S - N_R - P_S + E_p - A_N + A_p + Q$ (10-8)

$-47 + 136 - 17 - 1 \quad -3 \quad -3 + 16$

$= 81$

Comparing with the left-hand sides of the formulae, we find the minimum number of repeaters to be

(10-6)	Thermal and 2nd Order Noise	59-60 rptrs
(10-7)	Thermal and 3rd Order Noise	47 rptrs
(10-8)	Thermal noise and overload	46-47 rptrs

Thus the repeater spacing and levels must be set to meet thermal and second order modulation noise limits. Using 60 as the number of repeaters, we find

$$G_A = \frac{3000}{60}\ \text{db} = 50\ \text{db}$$

The value of C can be found from 10-1 or 10-2 since we chose to use an integral number of repeaters, the results will differ - thus,

$C = N_S - N_R - G_A + Q - 10 \log n - A_N$ (10-1)

$= -50 + 136 - 50 + 16 \quad -17.8 \qquad -3$

$= 31.2$

$C = -M_{2S} + M_{2R} + 10 \log n + A_{2M}$ (10-2)

$= +53 \quad -52 + 17.8 \qquad +12$

$= 30.8$

In solving these equations we use the -50 dbm value of N_S, since our solution of the previous formulae has told us that we have a system in which the total noise requirement should be equally divided between thermal and second order modulation noise.

Specific Questions

1. Level at Repeater Output

Using a value of 31 db for C, we find the top-channel level at repeater output (recall Q is 16 db) to be -15 db relative to transmitting switchboard.

2. Total Noise, before Tube Aging

To find the noise in dba at zero level for un-aged system:

Using (7-9)

$$P_{RNT} = N_R + G_A - Q + 10 \log n + C \text{ dbm}$$

$$= -136 + 50 - 16 + 17.8 + 31$$

$$= -53.2 \text{ dbm}$$

$$\approx 28.8 \text{ dba for thermal noise, 0 TLP}$$

Using (8-23)

$$H_{A+B} = M_{2R} + 10 \log n - C + 6$$

$$= -52 + 17.8 - 31 + 6$$

$$= -59.2$$

and since

$$W_{A+B} = H_{A+B} + 79 \text{ db}$$

W_{A+B} = 19.8 dba of second order modulation noise at 0 TLP

And similarly from (8-25) and the formula relating W and H for A+B-C products,

W_{A+B-C} = -2.4 dba of third order modulation noise at 0 TLP

These numbers should not be unexpected. Thermal noise is almost 32 dba minus the planned 3 db margin; second order noise should come out to +20 dba (32 dba with a 12 db margin) and is 19.8 dba. Third order noise is negligible - as we would guess from the fact that we are using 60 repeaters instead of 47.

The total noise at 0 db TLP is approximately +29.3 dba (sum of 28.8 dba and 19.8 dba).

3. Margins

The power that the system could cary is

$$P = E_p + C$$

$$= -1 + 31 \text{ or } 30 \text{ dbm at } 0 \text{ TLP}$$

This is 13 db more than required, so the A_p achieved is 13 db.

Approximately, the thermal noise could increase 6 db, or the second order modulation noise could increase to 34 dba, which is an increase of 14 db, before total noise exceeded +35 dba. Third order noise could increase by about 36 db before total noise exceeded +35 dba.

4. Effects of Tube Aging

Noise in dba at zero level after aging: thermal noise will not change, since G_A, NF and G_{OL} are assumed to be unchanged. (Actually the noise figure would change, since in general the equivalent tube noise resistance will increase as tubes age. The magnitude of this increase is not accurately known.)

By how much will modulation noise change? In answering this question, let us consider the effects one by one. First, the 9 db decrease in feedback will cause a 9 db increase in every modulation product, all else remaining the same.

What, however, will be the effect of the loss in transconductance in the output stage? The levels at the repeater output will remain the same, since insertion gain, to a first approximation, remains unchanged, and we have not specified that any changes are to be made at the terminals to change the levels of the signals delivered to the high frequency line. If we are to obtain the same fundamental power at repeater output with three db less transconductance in the output stage, the grid-cathode signal voltages must be three db greater than for new tubes.

Before we can say what effect this increased grid drive will have on modulation, we need information on the actual behaviour of tube modulation with age. Based on limited data, it is found that in old tubes, the second harmonic to fundamental power <u>ratio</u> in the load becomes somewhat better, and the third order ratio becomes somewhat worse, than for new tubes, <u>for the same fundamental grid-cathode voltage</u>. The results are spotty, and probably depend on the exact aging mechanism involved. If one must make some quantitative simple assumption, the one recommended is that for the same grid drive, the harmonic to fundamental ratios in the output do not change as tubes lose transconductance with age.

If we express this information in terms of the power series, we would say (somewhat uncertainly) that as a_1 decreases by 30%, a_2 and a_3 will also decrease by 30%. If, when we suffer a loss of 3 db in transconductance, due to aging, we increase the drive on the grids to get the same fundamental current as before,* the second order products delivered to the load will increase 3 db in power compared to their original pre-aging value, and the third order products by 6 db. In terms of the power series:

Unaged: $I_{p_o} = a_1\, eg_1 + a_2\, eg_1^{\,2} + a_3\, eg_1^{\,3}$

Aged, same grid drive: $I_{p_1} = \frac{a_1}{\sqrt{2}}\, eg_1 + \frac{a_2}{\sqrt{2}}\, eg_1^{\,2} + \frac{a_3}{\sqrt{2}}\, eg_1^{\,3}$

Aged, grid drive up 3 db: $I_{p_2} = \frac{a_1}{\sqrt{2}} \cdot \sqrt{2}\; eg_1 + \frac{a_2}{\sqrt{2}} \cdot 2\; eg_1^{\,2} + \frac{a_3}{\sqrt{2}} \cdot 2\sqrt{2}\; eg_1^{\,3}$

Since the gain from repeater output to zero level is unchanged, the increase in power of products at repeater output will directly affect the power of the products at zero level.

Summing up the effect of feedback and the change in grid cathode voltage, second order modulation noise will increase 12 db, and third order 15 db, at zero level.

The same results can be expressed in terms of some of the equations we have been using. Thus

$$M_2 = 20 \log \frac{a_2}{a_1\sqrt{2}} \quad \text{and} \quad M_3 = 20 \log \frac{a_3}{2a_1}$$

and we are told that these are constant, to a first approximation, as tubes age. Also, for the new system

- -

*which is what the feedback does in keeping the insertion gain and hence the repeater output power unchanged. This increase in drive would increase second order noise by 6 db if it were not for the fact that a_2 and a_3 change with age, as well as a_1.

$$W_{A+B_1} = H_{A+B} + 79.0 = H_{2A} + 85$$

$$= 85 + M_2 - F + 10 \log n - C$$

and for the old system

$$W_{A+B_2} = 85 + M_2 - F + 9 + 10 \log n - (C-3)$$

Whence

$$W_{A+B_2} - W_{A+B_1} = 12 \text{ db}$$

For the various contributors, therefore:

	New, dba	Change, db	Aged, dba
Thermal	28.8	0	28.8
2nd Order	19.8	12	31.8
3rd Order	- 2.4	15	12.6
Total	29.3	-	33.6

The overload margin will decrease 3 db to become 10 db. Third order noise is still negligible. The total margin against the noise requirement is 1.4 db. The thermal component could increase to 32.2 dba, or by 3.4 db without producing more than 35 dba total (assuming second order noise fixed at 31.8 dba). The second order noise could increase to 34.5 dba, or by 2.7 db, without causing more than 35 dba total, if thermal remains 28.8 dba.

5. Relative Value of Improving Parameters

a) When noise figure improves by 3 db, the right hand side of (10-6) goes from 86 db to 89 db. From the curve, the number of repeaters required would go from 60 to 56. For a two-way system at $75,000 per repeater, this is worth $600,000.

b) The intention here was to ask the value of increasing Q by 3 db. Thus interpreted, the answer is the same as for part a) above.

Actually one should also ask what the corresponding changes in feedback and in M_2 and M_3 would be. A three db increase in output stage transconductance would at best give about 1.5 db more feedback, for the same gain and phase margins. Probably less improvement than 1.5 db in feedback would really result, since capacitances would also increase as higher transconductance was obtained. (Chapter 13 discusses the mechansims involved). M_2 and M_3 might be expected not to change, but this is just a reasonable guess, on the assumption that the higher transconductance tube would have correspondingly higher components for all terms of the power series, including the d.c. term.

c) Again the same saving.

d) No benefit, since third order noise is negligible already.

e) Re-evaluating the quantity $\frac{L_s}{n}$ + 20 log n at n = 55 and L_s = 2800, we find 55 repeaters would satisfy equation (10-6). This saves ten repeaters (two-way) which is $750,000. As a matter of interest, this might be compared to the cable cost, which at the factory (not including laying expenses) is about $4000 per mile, or $16,000,000 for the two-way 2000 mile system under discussion.

Chapter 11

MISALIGNMENT

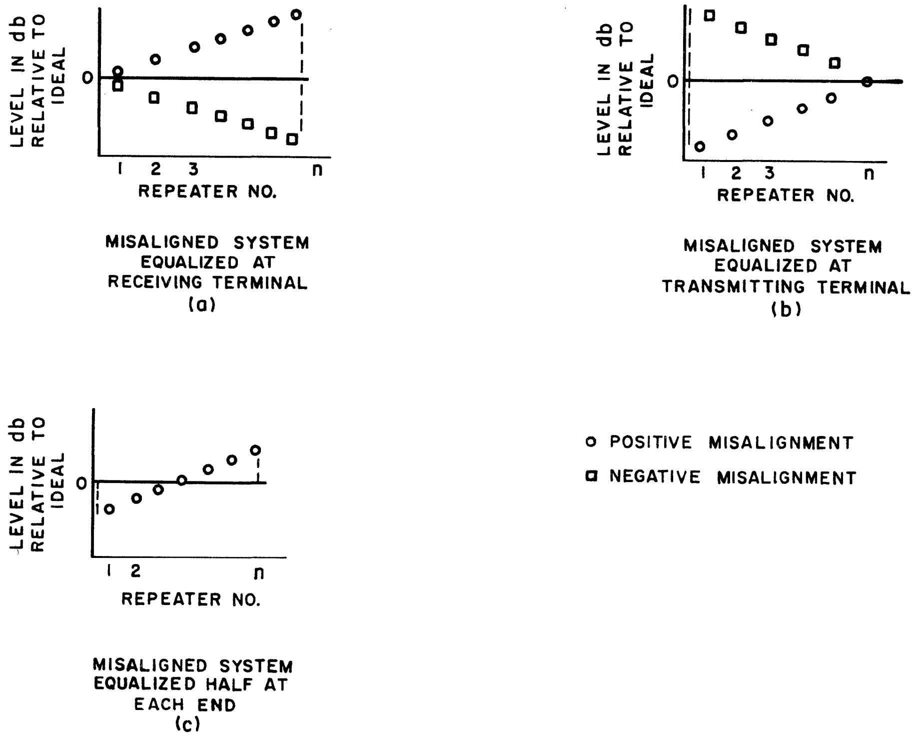
In SSB-AM transmission systems such as those we have been considering here, the gain of a repeater is designed to match the loss of the cable associated with that repeater. A cable plus repeater is termed a repeater section. It follows that the gain of such a repeater section is designed to be unity (zero db). This design objective can be met only approximately. The departures from the ideal are usually small for one repeater section. Since it is generally uneconomical, if not impossible, to correct the transmission deviations at the end of each repeater section, these departures from the ideal are allowed to accumulate until they are large enough to be accurately measured. Then, when possible, an "equalizer repeater" is inserted. Such a repeater contains additional equipment to correct the accumulated transmission deviation.

When the output of a repeater section differs from the input, the section is said to be misaligned. We can represent this condition by writing

$$E_{OUT} = \delta_x E_{IN} \tag{11-1}$$

In an ideal system δ_x would be 1. However, when misalignment occurs, δ_x can be either greater or less than 1. The misalignment per section in db is 20 log δ_x. Because misalignment is usually expressed in db, a repeater section which has a net gain (δ_x>1) is referred to as having positive misalignment. Correspondingly, repeater sections which have a net loss ($\delta_x < 1$) are denoted as having a negative misalignment. δ_x is in general a function of frequency and the subscript, x, indicates the frequency at which it applies.

In many cases the misalignment in each of a string of repeater sections will be almost the same and deviations will accumulate systematically from section to section. This will be true when the causes of misalignment in each section are the same, e.g., cable aging, tube aging, or temperature effects. In the following pages, expressions will be developed for the calculation of the effect of such a systematic misalignment on thermal and modulation noise. When the misalignment in successive sections is random, the signal to interference penalties are difficult to compute; for such cases, it will often be possible to use the expressions



Repeater Levels on a Misaligned System

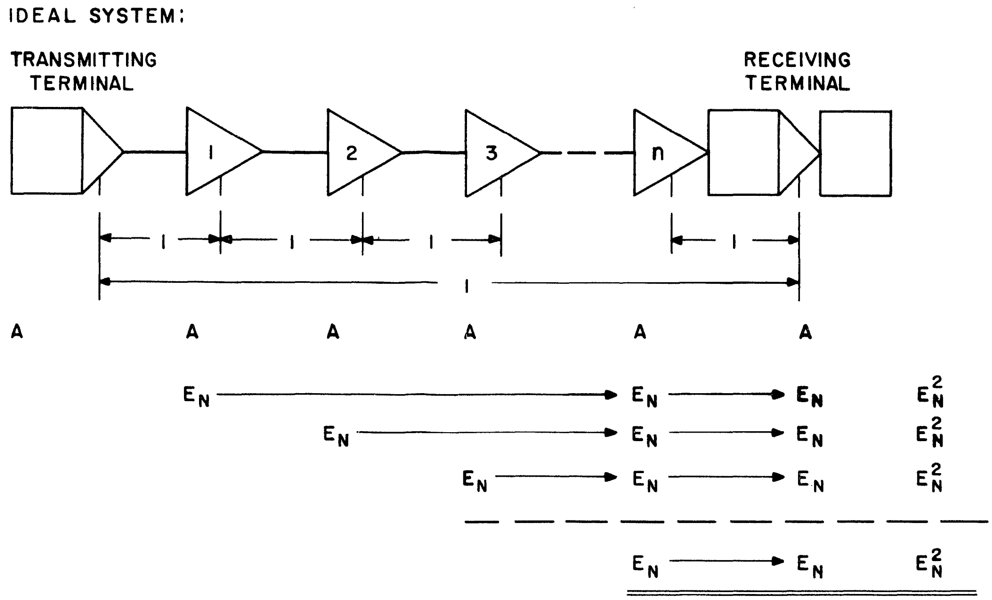
Figure 11-1

for the penalties associated with systematic misalignments to determine an upper limit for the penalties to be expected.

Misalignment Penalties in Uniformly Misaligned Systems

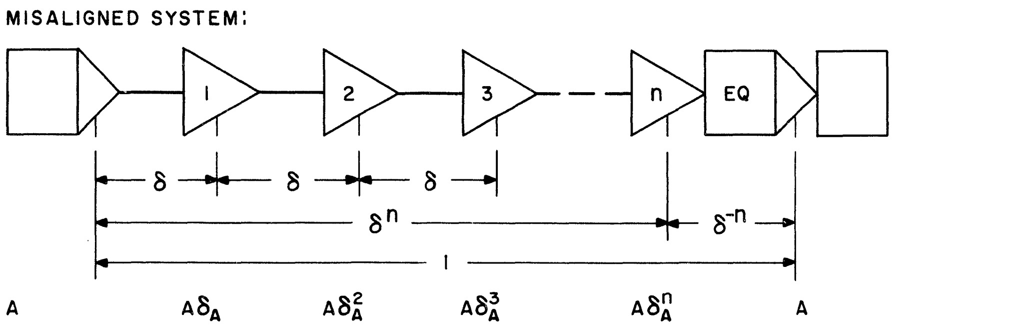
Figure 1 shows repeater levels for systematically misaligned systems and various types of equalization. We shall begin by investigating the effect of misalignment on thermal noise, with equalization at the receiving end as in Figure 1a. Then, having become familiar with the sort of results which can be expected, the somewhat more complex case of modulation noise will be examined. Finally the effect of other types of equalization will be discussed.

The analysis is based on Figure 2. A system of N identical repeaters designated 1, 2, ... n is assumed. Each repeater section generates the same random noise voltage as all the others. This voltage, referred to the output grid of the repeater, will be designated as E_N. Primarily because it leads to a neater notation, we will treat the receiving amplifier as the nth repeater and the transmitting amplifier as ideal - leaving its noise contribution to be incorporated in the more refined noise calculation which will also include the contributions of the terminal flat gain amplifiers mentioned in the final section of this chapter.



TOTAL RECEIVED NOISE POWER = nE_N^2

(a)



$$E_N \longrightarrow E_N\delta_N \longrightarrow E_N\delta_N^2 - - \rightarrow E_N\delta_N^{n-1} \longrightarrow E_N\left(\frac{1}{\delta_N}\right) \qquad E_N^2\left(\frac{1}{\delta_N}\right)^2$$

$$E_N \longrightarrow E_N\delta_N - - \rightarrow E_N\delta^{n-2} \longrightarrow E_N\left(\frac{1}{\delta_N}\right)^2 \qquad E_N^2\left(\frac{1}{\delta_N}\right)^4$$

$$E_N - - - \rightarrow E_N\delta^{n-3} \longrightarrow E_N\left(\frac{1}{\delta_N}\right)^3 \qquad E_N^2\left(\frac{1}{\delta_N}\right)^6$$

$$E_N\delta^{n-n} \longrightarrow E_N\left(\frac{1}{\delta_N}\right)^n \qquad E_N^2\left(\frac{1}{\delta}\right)^{2n}$$

LET $X = \frac{1}{\delta_N}$

THEN TOTAL RECEIVED NOISE POWER =

$$E_N^2\,X^2 + E_N^2\,X^4 + E_N^2\,X^6 + \ldots\ldots + E_N^2\,X^{2n} = E_N^2\sum_{n=1}^{n} X^{2n} = E_N^2\,\frac{X^2 - X^{2(n+1)}}{1-X^2}$$

(b)

Effect of Misalignment on Thermal Noise

Figure 11-2

In Figure 2a an ideal system with no misalignments is represented. "A" stands for the amplitude of a signal voltage transmitted through the system. It has the same magnitude at the output grid of each repeater and at a corresponding level point in the receiving terminal. At each repeater, a noise voltage E_N is generated, each of which also appears as E_N at the output grids of all succeeding repeaters and at the reference point in the receiving terminal. When they are all added up on a power basis, the total noise power at the reference point in the receiving terminal has been found.

$$\text{Noise power in an Ideal System} = 10 \log n \, E_N^2 \qquad (11\text{-}2)$$

In Figure 2b a misaligned system is represented. Before comparing the noise in this system to that in the ideal system, it is necessary to bring the noise to the same transmission level, i.e., to a point where the signals in the two systems have the same magnitude. This can be brought about by inserting an equalizer or regulator at the receiving end, which, at each frequency, just compensates for the overall transmission deviation produced by the cumulative effect of all the system misalignments. Such an equalizer is not only necessary for our theoretical calculations but is required in the physical system. That portion of the plant which is to receive the signals from the system under consideration has been designed to accept them at a particular level. If they do not arrive at this level because of misalignments, they should be adjusted before sending them on their way. In other words, our system should have the correct Via Net Loss; hence the requirement for an equalizer (which will include an amplifier when $\delta < 1$) with a shape corresponding to δ^{-n}.

In the misaligned system, each repeater still generates a thermal noise voltage E_N, but each of these voltages is now subject to a different amount of gain (or loss) on its way to the receiving terminal, depending on the repeater in which it is generated. In general, the noise voltage generated at the x'th repeater arrives at the receiving amplifier, n, with magnitude $E_N \delta_N^{(n-x)}$ since it will have gone through n-x repeater sections. After passing through the equalizer, it arrives at the reference point in the receiving terminal with a magnitude $E_N \delta_N^{-x} = E_N (\frac{1}{\delta_N})^x$. When we add up on a power basis all of the noise generated in the repeaters after it has arrived at the receiving terminal, the result is

Noise Power in Misaligned System

$$= 10 \log E_N^2 \sum_{n=1}^{n} X^{2n} = 10 \log E_N^2 \frac{X^2 - X^{2(n+1)}}{1-X^2} \qquad (11\text{-}3)$$

where $X = \frac{1}{\delta_N}$. Misalignment penalty is defined as the difference (in db) between the noise in the misaligned as compared to the ideal system. It should be emphasized that before comparing these two noise magnitudes we have brought the signals back to their correct levels. The misalignment penalty is a true measure of S/N degradation. Subtracting Equation 2 from Equation 3 we get

$$\text{Misalignment Penalty} = 10 \log \left\{ \frac{1}{n} \frac{X^2 - X^{2(n+1)}}{1-X^2} \right\} \qquad (11\text{-}4)$$

Note that the penalty will be positive when $X > 1$ (corresponding to negative misalignment) which means that the thermal noise is greater in the misaligned case. When $X < 1$, a negative penalty results, i.e., thermal noise at the output of the positively misaligned system (equalized at the receiving end) is less than in the corresponding ideal system. The significance of this will be examined later.

Figure 3 illustrates the effect of misalignment on modulation noise. We examine what happens when the three tones

$$E_A = A \cos 2\pi f_A t$$

$$E_B = B \cos 2\pi f_B t$$

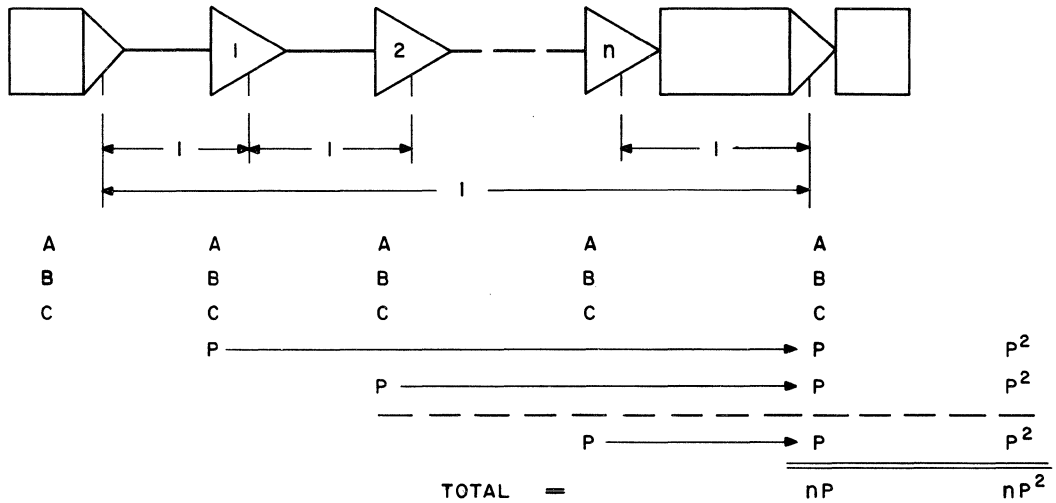
$$E_C = C \cos 2\pi f_C t$$

are fed into the system. All of the repeaters are assumed to have identical M_{2R} and M_{3R}. Figure 3a shows the ideal system without misaligned sections. Under these conditions, each repeater will generate products whose magnitudes are equal to the magnitudes of the products generated by all the other repeaters. For example, each repeater generates an A+B-C product equal in magnitude to

$$E_{A+B-C} = 6[\log^{-1} \frac{M_{3R}}{20}] E_A E_B E_C = K E_A E_B E_C \ .$$

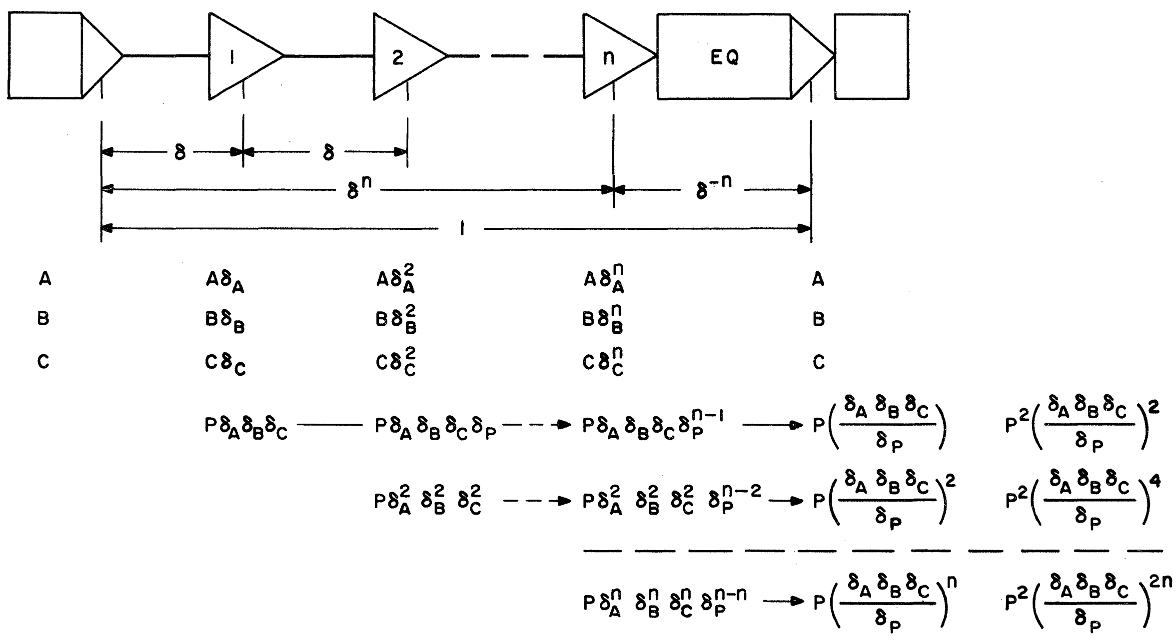
The modulation product, P, generated at the output grid of each repeater appears at the receiving terminal as well as at the output grids of all

IDEAL SYSTEM:



(a)

MISALIGNED SYSTEM:



$$\text{LET } X = \frac{\delta_A \delta_B \delta_C}{\delta_P}$$

TOTALS

$$\text{(1ST COLUMN)} = PX + PX^2 + PX^3 + \ldots\ldots + PX^n = P\sum_{n=1}^{n} X^n = P\,\frac{X - X^{n+1}}{1-X}$$

$$\text{(2ND COLUMN)} = P^2X^2 + P^2X^4 + P^2X^6 + \ldots\ldots + P^2X^{2n} = P^2\sum_{n=1}^{n} X^{2n} = P^2\,\frac{X^2 - X^{2(n+1)}}{1-X^2}$$

(b)

***The first Column Applies for Voltage Addition Products**
The Second Column Applies for Power Addition Products

Effect of Misalignment on Modulation Noise
(Receiving Equalizer Only)

Figure 11-3

intermediary repeaters, and always at the same value as that at which it was generated. Adding up these products (on a voltage or r.m.s. basis, depending on the type of product), there results

$$\text{Ideal Voltage Addition Product Power} = 20 \log n P \qquad (11\text{-}4)$$

$$\text{Ideal Power Addition Product power} = 10 \log n P^2 \qquad (11\text{-}5)$$

In the misaligned system, Figure 3b, the misalignments of each section at fundamental and product frequencies are designated as δ_A, δ_B, δ_C and δ_P. Again an equalizer of shape δ^{-n} is inserted at the receiving end to bring the signals back to their proper levels. As in the case of thermal noise, the modulation product generated at repeater X is subject to a gain (or loss) of $\delta_P{}^{n-x}$ before reaching the equalizer input. Unlike the thermal noise, the modulation product generated at one repeater in a misaligned system differs from that generated at another since it is a function of the magnitude of the incoming signal. For instance in the case of the A+B-C product generated at the x^{th} repeater, the product magnitude will be $P\delta_A^x \ \delta_B^x \ \delta_C^x$ where P is the value for the product magnitude in the unmisaligned system. After traversing the n-x repeater sections between its point of generation and the receiving amplifier, this product appears as $P\delta_A^x \ \delta_B^x \ \delta_C^x \ \delta_P{}^{n-x}$. It finally arrives at the reference point in the receiving terminal with a magnitude $P \left(\frac{\delta_A \delta_B \delta_C}{\delta_P}\right)^x$. Letting $X = \frac{\delta_A \delta_B \delta_C}{\delta_P}$ and adding up the contributions of each repeater on a voltage or power basis, depending on the nature of the product, we get:

$$\text{Misaligned Voltage Addition Product Power} = 20 \log P \frac{X-X^{n+1}}{1-X} \qquad (11\text{-}6)$$

$$\text{Misaligned Power Addition Product Power} = 10 \log P^2 \frac{X^2-X^{2(n+1)}}{1-X^2} \qquad (11\text{-}7)$$

Signal-to-interference penalties are obtained as in the case for thermal noise, by subtracting the corresponding ideal system interferences from these values.

$$\text{Misalignment Penalty}_{\substack{\text{VOLTAGE}\\ \text{ADDITION}}} = 20 \log \left(\frac{1}{n} \frac{X-X^{n+1}}{1-X}\right) \qquad (11\text{-}8)$$

$$\text{Misalignment Penalty}_{\substack{\text{POWER}\\ \text{ADDITION}}} = 10 \log \frac{1}{n} \frac{X^2-X^{2(n+1)}}{1-X^2} \qquad (11\text{-}9)$$

Comparison of Equations 11-9 and 11-4 will show that these two expressions are identical in form. The only difference between the penalty for power addition modulation penalty and thermal noise penalty is in the definition of the X's. Thus, positive misalignment, which gave an advantage in the case of thermal noise, gives a penalty in the modulation case. This last statement is strictly true only for the normalized misalignments which are the reciprocals of the X's given for various types of products in Table 1.

TABLE 11-1

Product Type	X
2A	δ_A^2/δ_P
3A	δ_A^3/δ_P
A+B	$\delta_A\delta_B/\delta_P$
A-B	$\delta_A\delta_B/\delta_P$
2A+B	$\delta_A^2\delta_B/\delta_P$
2A-B	$\delta_A^2\delta_B/\delta_P$
A-2B	$\delta_A\delta_B^2/\delta_P$
A+B-C	$\delta_A\delta_B\delta_C/\delta_P$
A+B+C	$\delta_A\delta_B\delta_C/\delta_P$
A-B-C	$\delta_A\delta_B\delta_C/\delta_P$
Random Noise	$1/\delta_P$

In the derivations up to this point, the equalizer in the misaligned systems has been located at the receiving end. This is arbitrary. The only requirement that must be met is that the over-all transmission from terminal to terminal remain unchanged when we go from the ideal to the misaligned system. This condition can equally well be met if the equalizer is located in the transmitting terminal. More generally yet - we can use two equalizers, one at each terminal, whose combined effect corresponds to the total correction necessary to compensate for the effect of misalignment. This last case can be represented by a transmitting equalizer of characteristic δ^a and a receiving equalizer of characteristic δ^{-a-n}. This case reduces to the previously solved case when a = 0. When a = -n, it corresponds to a system where all equalization is inserted at the transmitting end. By a line of reasoning which should be all too familiar by now, the misalignment penalty for the general case of two equalizers can be obtained. This is done in Figure 4 for modulation noise. As before, the expression for thermal noise is identical to that for power addition modulation products with $X = \frac{1}{\delta_N}$. The results obtainable from Figure 4 are:

$$\text{Misalignment Penalty}_{\substack{\text{VOLTAGE}\\ \text{ADDITION}}} = 20 \log \frac{X^a}{n} \frac{X-X^{n+1}}{1-X} \qquad (11\text{-}10)$$

$$\text{Misalignment Penalty}_{\substack{\text{POWER}\\ \text{ADDITION}}} = 10 \log \frac{X^{2a}}{n} \frac{X^2-X^{2(n+1)}}{1-X^2} \qquad (11\text{-}11)$$

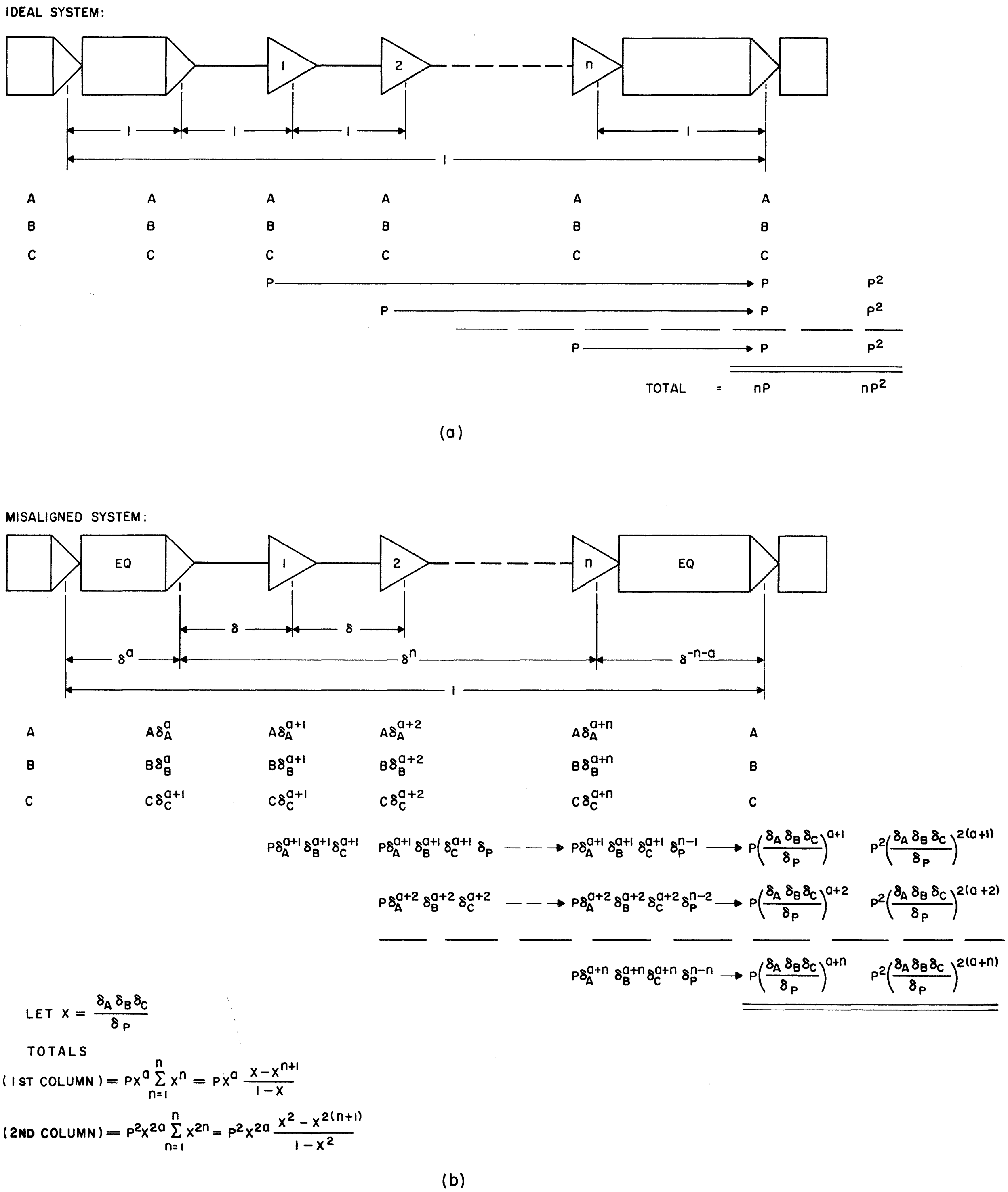
Type of Misalignment Equalization to be Used

Whether the equalizer is to be placed at the receiving or the transmitting end, or whether equalization at both ends is best, depends on the type of system and the nature of the misalignment.* In the discussion of this problem Figure 1 will be useful. The diagrams in that figure represent the levels of the output grids of the repeaters making up the system. In order to simplify our thinking we will assume that misalignment is constant with frequency. In cases where misalignment is a function of frequency, our considerations will still be valid, but must be applied at each frequency of interest.

From our previous work, we know that raising repeater levels on a system reduces the thermal noise at zero level at the receiving terminal, but increases modulation noise at that point. Examining Figure 1a it is evident that an increase in repeater levels occurs as a result of positive misalignment which is equalized at the receiving terminal. This would be expected to produce an increase in modulation noise and a decrease in thermal noise. If numbers are substituted into Equations 8 and 9, this is indeed found to be the case. Negative misalignment (coupled with receiving terminal equalization) on the other hand, will reduce levels with a resulting increase in thermal noise and decrease in modulation noise. Figure 1b shows that the situation is reversed for the case of transmitting terminal equalization. In this situation negative misalignment produces higher level repeaters whereas positive misalignment lowers the repeater levels with the usual effects on thermal and modulation noise.

---

*Difficulties in administration of a particular equalizing scheme must, of course, be weighed against advantages to be gained.



***The First Column Applies for Voltage Addition Products**
The Second Column Applies for Power Addition Products

Effect of Misalignment on Modulation Noise
(Two Equalizers)

Figure 11-4

In Chapter 10 we found that the minimum repeater level was set on the basis of maximum permitted thermal noise at zero level. Maximum repeater levels were determined by modulation noise requirements <u>or</u> overload requirements, depending on which condition was more stringent. If the ceiling on repeater levels is set by the maximum signal permitted on the output grid, the system is said to be overload limited.

Just how inflexible this ceiling is, depends on what phenomenon occurs when a repeater is overloaded. If we have a repeater with only a moderate amount of feedback, somewhat higher than ideal levels on a few of the repeaters may be permissible. On the other hand, if overload is defined by "cracking" in a repeater with a large amount of feedback (where the "stonewall" concept applies), or by reliability considerations, the ceiling will be relatively inflexible. In the extreme case,* it is not permitted to have <u>any</u> repeater at a higher than ideal level. If it is assumed that the original ideal system was designed to have all the repeaters operating at such a maximum level, care must be taken that no repeater in the misaligned system exceeds this level. In the light of this conclusion, examination of Figure 1a and b indicates that for an overload limited system, positive misalignment must be equalized at the transmitting end whereas negative misalignment must be equalized at the receiving terminal.

If modulation noise is controlling, the procedure of Chapter 10 provides for setting repeater levels of an ideal system so as to minimize the total noise composed of thermal and modulation noise. When misalignment occurs in such an optimized system, it shifts repeater levels which will produce either an increase in thermal noise and a decrease in modulation noise or vice versa. Since the system was originally set for an optimum value, such a shift can be expected to produce an increase in the total noise on the system. We would expect this increase to be minimum if the repeater levels of the misaligned system were held as near as possible to the ideal level. Figure 1c shows that this can be done if half the equalization is introduced in the transmitting end and the remaining half in the receiving terminal. This corresponds to two equalizers, each of which has the characteristic $\delta^{-n/2}$. If $a = -n/2$ is substituted into equations 10 or 11 it is seen that considerable reductions in penalty is possible (see Figure 5). It must of course be ascertained that the highest level repeater in a system such as depicted in 1c is not so much higher than the ideal, that it is subject to overload. Subject to this caution, modulation limited systems should be equalized at both ends.

- -

*One practical example where this extreme case applies is the present submarine cable system.

Approximate Expressions for Misalignment Penalties

The discussion of the effect of misalignment on system performance has been completed. However, the expressions for misalignment penalty so far derived are somewhat unwieldy. It is possible to derive considerably neater expressions which are still sufficiently accurate, especially in light of the fact that in no real situation will the assumptions, on which the previous deviations are based, be exactly satisfied. These approximate expressions will also make it possible to present the results of Equations 8 through 11 in a very compact fashion.

The simplification of the previous equations follows the same pattern for all the cases considered. The case for the thermal noise penalty for a negative misaligned system and receiving terminal equalization will be worked out as an example. The expression for this penalty is given by Equation 9 as

$$\text{Penalty} = 10 \log \frac{1}{n} \frac{X^2 - X^{2(n+1)}}{1-X^2}$$

Let $\Delta = \delta^n$ = total system misalignment, i.e., the over-all change in amplitude that a signal would experience in the absence of equalizers. $M = 10 \log \Delta^2$ is this total misalignment in db. Remembering that $X = \frac{1}{\delta}$ in the case of thermal noise we can rewrite the above equation as

$$\text{Penalty} = 10 \log \frac{1}{n} \frac{\Delta^{-\frac{2}{n}}(1-\Delta^{-2})}{1-\Delta^{-2/n}}$$

Expanding $\Delta^{\frac{-2}{n}}$ in a power series around $\frac{1}{n} = 0$, the result is

$$\Delta^{-2\frac{1}{n}} = 1 + \frac{1}{n} \ln \Delta^{-2} + \frac{1}{2} (\frac{1}{n} \ln \Delta^{-2})^2 + \dots$$

$$= 1 + \frac{1}{n} (\frac{-M}{4.34}) + \frac{1}{2} [\frac{1}{n} (\frac{-M}{4.34})]^2 + \dots$$

Now in almost all cases $|\frac{-M}{4.34n}| \ll 1$. Therefore, this result can be substituted in the expression for misalignment penalty and all but the first remaining terms in the numerator and denominator discarded. This gives

$$P = 10 \log \frac{4.34(1-\Delta^{-2})}{M}$$

Rewriting the result in terms of M, the expression becomes

$$\text{Penalty} = 10 \log \frac{1-\log_{10}^{-1} \left(-\frac{M}{10}\right)}{M} + 6.4 \text{ db}$$

We have made a big gain since the above expression is now independent of the number of repeaters and is a function of the total misalignment (in db) only. If we had carried out the above steps for a positively misaligned system with an equalizer at the transmitting end (as would be required for an overload limited system), the result would have been

$$\text{Penalty} = 10 \log \frac{\log_{10}^{-1} \frac{M}{10} - 1}{M} + 6.4$$

which gives the same penalty as the previous expression for negative misalignment when the M's are numerically equal (but opposite in sign). This is not surprising since there almost is a one-to-one correspondence in the repeater levels found in the two cases (repeater n-1 is at the same level as repeater 1 in the negative misalignment case, repeater n-2 is at the same level as 2, and so on). We can therefore rewrite the above results in one equation covering both cases. For positive misalignment and transmitting equalizer, or for negative misalignment and receiving equalizer, the following equation gives the thermal noise penalty:

$$\text{Penalty} = 10 \log_{10} \frac{\log_{10}^{-1} \left|\frac{M}{10}\right| - 1}{|M|} + 6.4$$

This result can also be used to calculate the modulation noise penalty for products that add on a power basis if $|M|$ is interpreted as the total normalized misalignment

$$\left| 10 \log \left(\frac{\delta_A \delta_B \delta_C}{\delta_P}\right)^{2n} \right| .$$

By calculations analogous to the one just carrier out, similar simplified expressions can be obtained for all the other cases of interest. These are summarized below and plotted on Figure 5. If misalignment is less than 1 db per section, the expressions will be in error by less than about 0.5 db. This condition is just a quantitative statement of the condition given on the facing page which must be satisfied for the approximation to hold.

I. Positive Misalignment and Transmitting Equalizer, or Negative Misalignment and Receiving Equalizer

A. Thermal noise, or power addition modulation products

$$\text{Penalty} = 10 \log_{10} \frac{\log_{10}^{-1} \left|\frac{M}{10}\right| - 1}{|M|} + 6.4 \qquad (11\text{-}12)$$

B. Voltage addition modulation products

$$\text{Penalty} = 20 \log_{10} \frac{\log_{10}^{-1} \left|\frac{M}{20}\right| - 1}{|M|} + 18.8 \qquad (11\text{-}13)$$

II. Identical Equalizers at Both Ends

A. Thermal noise, or power addition modulation products

$$\text{Penalty} = 10 \log_{10} \frac{1}{|M|}\left[\log_{10}^{-1}\left|\frac{M}{20}\right| - \frac{1}{\log_{10}^{-1}\left|\frac{M}{20}\right|}\right] + 6.4 \qquad (11\text{-}14)$$

B. Voltage addition modulation products

$$\text{Penalty} = 20 \log_{10} \frac{1}{|M|}\left[\log_{10}^{-1}\left|\frac{M}{40}\right| - \frac{1}{\log_{10}^{-1}\left|\frac{M}{40}\right|}\right] + 18.8 \qquad (11\text{-}15)$$

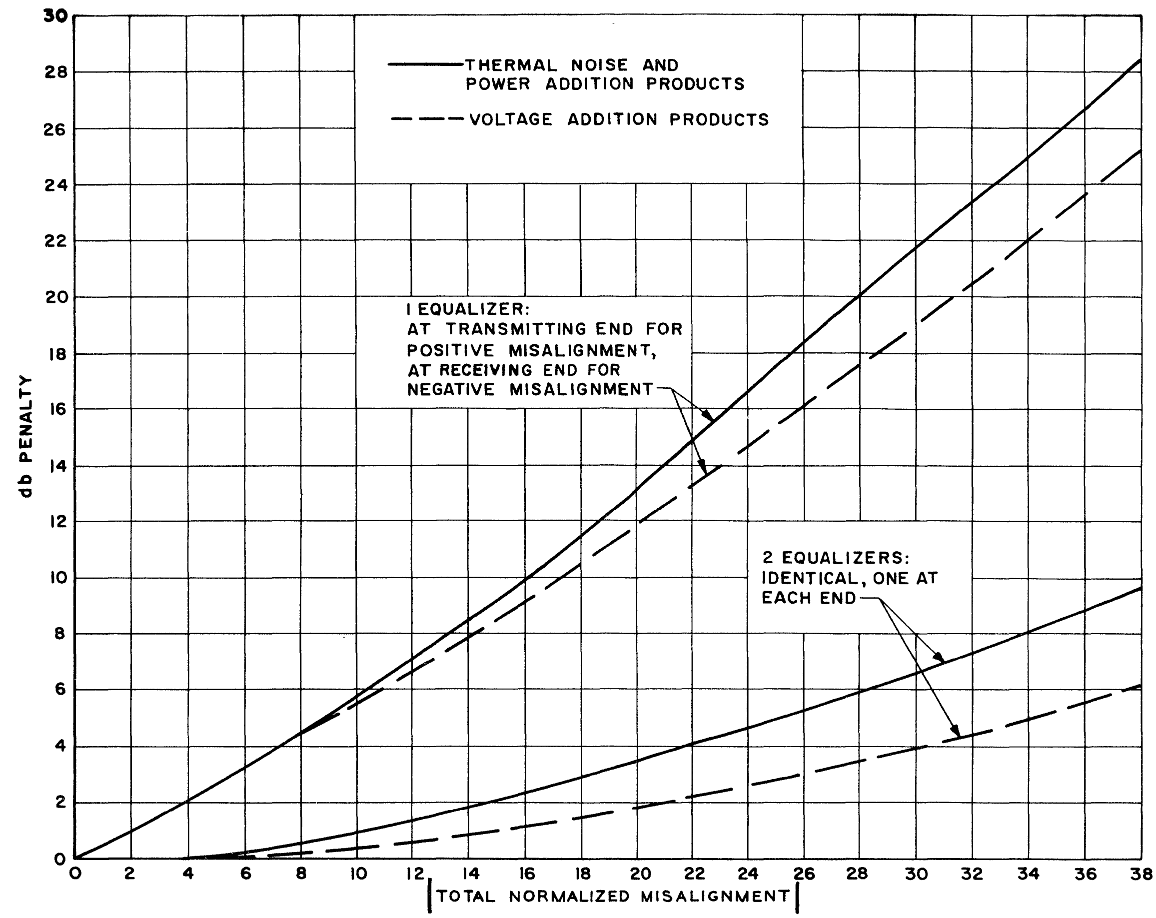
S/N Penalties due to Equalizer Loss

Up to this point, we have ignored the signal-to-noise penalties introduced by the equalizers themselves. Actually these penalties are an important part of the total system penalties associated with misalignment.

In a subsequent chapter, it will be seen that to fulfill their basic functions, equalizers must be complex networks with a considerable adjustment range. Being passive, they will have rather large flat losses when set at the middle of their adjustment range.* The full complement of equalizers in a system may comprise as much as 10% of the total loss of the high frequency line at the top transmitted frequency, and a much greater percentage at lower frequencies. Flat gain amplifiers are commonly associated with each equalizer (or combination of equalizers at a particular office) to compensate for these losses.

- -

*Many so-called "active equalizers" have been proposed in which the multi-shape equalizers act as interstages or beta networks in amplifiers. Most, if not all, of these designs have had to be abandoned because of inherently poor signal-to-noise performance.



Misalignment Penalty*

Figure 11-5

The flat-gain amplifiers will be modulation and random noise contributors. In a system having a large amount of slope over the transmitted band, and where a large amount of equalizer range is needed (hence a large amount of flat loss), the input of the flat-gain amplifiers at equalizing points may represent the lowest level point in the system at low frequencies. In consequence the noise contribution of the flat-gain amplifiers may dominate the system random noise at low frequencies. The permitted magnitude of this contribution will determine the maximum permissible loss of an individual equalizer - and hence its maximum adjustment range. While the modulation contribution of the flat-gain

---

* This figure cannot be used for positive misalignment coupled with receiving equalization or for negative misalignment coupled with transmitting equalizations. For these cases similar figures can be obtained by an analogous method, but these cases are not often of interest.

amplifiers must also be allowed for, it is in general not a critical parameter - if these amplifiers are similar, modulation-wise, to the line amplifiers they will merely increase "n", the number of repeaters in our ideal system computations, since signal levels at the output grid will be very nearly the same as the line amplifier signal levels are in the "ideal" system.

With proper system design, the line amplifiers of a cable system usually dominate its signal-to-interference performance at the higher transmitted frequencies. Early signal-to-noise analyses are therefore usually carried out without regard to the effects of equalization gear. However, equalizer noise penalties must be included in more refined final analyses.

Chapter 12

OVERLOAD AND MODULATION REQUIREMENTS

In this chapter we shall discuss the nature of the telephone speech load applied to multi-channel systems and the derivation of the requirements for load carrying capacity and modulation. The material is presented in some detail, not only because of its fundamental importance in Bell System work, but also because it illustrates the type of knowledge, about a signal and its sensitivity to interferences, which is necessary if we are to design a system to efficiently carry any particular service.

Ideally, we should have at our command all the field data that we require, measured in the terms which will make it most usable for our purposes. Actually, however, gathering data on the telephone speech load and subscriber reactions is a long and expensive process, and we are forced to use what data are available. This involves extrapolation, conversion from one type of measurement-unit to another, and other sources of uncertainty. Again, this is likely to be true for other signals, too.

Volume

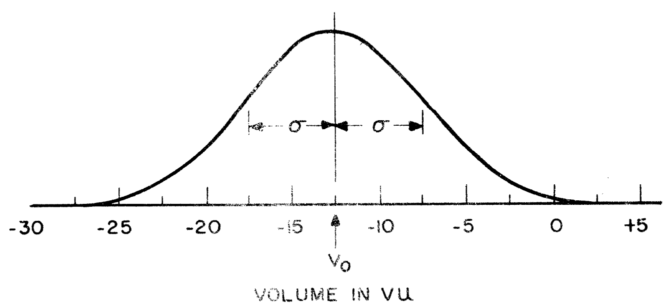
In any discussion of load carrying capacity, our first question must be: what is the magnitude of the signals applied to the system? More specifically, if we are considering overload, our question takes the form "Given a system equipped to carry N channels, what will be the voltage-time functions at some chosen point in the system during the busy hour"? A convenient point to choose would be an imaginary zero level point where the broadband spectrum formed by the AM terminals can be observed. From information about voltages at this point we can convert to voltages at output grids of line repeaters.*

When Holbrook and Dixon set out to answer this question, the best data available on the telephone speech load was in terms of individual channel volumes. This is still true. Volume, like noise, is measured in the Bell System using standard meters in a carefully prescribed fashion. Readings are given in "volume units", usually abbreviated to "vu". These bear a "db" relationship to each other - which is to say that if we obtain a reading of zero vu for a given talker at a zero level point, the same talker measured at a -10 db level point would

---

* Although the question of required load carrying capacity is discussed in terms of multi-channel AM systems, the results are applicable to the problem of peak frequency deviations in FM systems, as we shall see.

give a reading on the volume indicator of -10 vu. Readings are taken by watching the meter during speech and roughly averaging the highest readings, ignoring the rare exceptionally high peaks. The time constants of the volume indicator are designed to integrate the instantaneous speech power over roughly syllabic intervals. Volume is thus not a direct measure of power or peak voltage. It can, however, be related to these quantities by the results of other tests, so that if we have a large body of data on telephone talkers in terms of volume, we can make use of this data to estimate overload probabilities by finding the necessary conversion factors. Because of the "db" character of the vu scale, we can readily correct our system requirements as new volume data become available. (It should be recognized that the telephone speech load is a function of geography,* time, and other factors - for example, as better instruments and circuits become more common, fewer people assume that they must shout to be heard, especially on a long distance call; volumes increase about 1 db for each 1000 mile increase in air line distance between stations, etc.)



Talker - Volume Distribution

Figure 12-1

A typical distribution curve of the volumes of a large population of talkers, measured at a zero level point is shown in Figure 12-1 which is drawn with a mean, V_o, equal to -12.5 vu and a $\sigma=5$. Values of -10 vu correspond to data obtained quite some time ago. More recent tests indicate that V_o has decreased to -15 vu and that $\sigma=5.3$. Estimates of the volumes which will exist in the future differ widely - some indicating a continued decrease in average volume while others predict an increase. In order to be prepared for the most critical condition, most systems are being designed today on the basis of the compromise -12.5 figure. Should the average volume remain lower, the system can be administered in such a way as to

---

*Recent tests show that acoustic volumes at the telephone transmitter do not vary from city to city. Volumes measured at zero level do vary because of local differences in loops, battery supplies, and line noise.

trade the extra load carrying capacity for better noise performance. Since noise requirements are almost certain to tighten up in the future, being conservative about volumes will avoid a situation where it might be necessary to decrease noise and increase load carrying capacity - a feat that would be difficult without radically reducing the number of channels the system can carry.

In later computations we shall have occasion to use the concept of the "average power talker" - that talker whose long-term average power, multiplied by the number of talkers present, gives us the total long-term average power in a multi-channel telephone load. It should be noted that the power of the average power talker is not the power of a talker whose volume is V_o, the average of the talker-volume distribution. This distribution is normal in db, or the logarithm of power, rather than power itself. It is necessary to convert from the average of the db distribution to the average of the corresponding power distribution. W. R. Bennett has shown that for such a universe the volume "V_{op}" corresponding to the average power talker is:

$$V_{op} = V_o + .115\,\sigma^2 \qquad (12\text{-}1)$$

The power corresponding to a zero vu talker is -1.4 dbm, hence if V_o is -12.5 vu and σ is 5.0 db, the volume of an "average power talker" is -9.6 vu at the zero level point, and the corresponding power is -11 dbm at zero level.

Load Carrying Capacity Requirements

When we set up a requirement on a system, it is desirable that we express our results in terms of easily measured quantities. For load carrying capacity, it is convenient to define a sine wave test tone which the system must be able to carry without overloading - in previous chapters, the power of this test tone in dbm at zero level has been referred to as "P_s". If we define this power correctly, then a system which just carries this test signal without overloading will just carry the complex telephone load satisfactorily.

Notice the choice of the word "satisfactorily". We did not say that the complex telephone load would never overload the system. Instead we take a small calculated risk, as one usually should in seeking a good engineering solution to any problem. Good commercial service and very rare bursts of distortion or noise caused by overload are not incompatible. The question is: how rare?

The answer to this question can be sought in the following terms. Consider the rms voltage readings which might be obtained if we monitored the broad-band one-way telephone multiplex signal at some point in a four-wire system - suppose we do this monitoring with a moderately fast thermocouple. We would find, during the busy hour, that this rms voltage would vary around some average value.

Two probability distributions contribute to this variation. One of them arises from the fact that the number of channels which are active is a question of probability. A channel is "active" whenever continuous speech is being introduced into it. The probability of finding a given one-way channel active at any moment during the busy hour is about 1/4.* The other fellow is talking 50% of the time, some time is spent setting up the connections, and there are pauses in speech (the intersyllabic pauses are considered part of speech itself, but there are other much longer pauses). The average number, N_a, of active channels during the busiest hours will be about 1/4 the total number "N". In computing the largest values of rms voltage to be expected, however, it is important that we remember that the number of channels that are active at a given instant in an N-channel system may be anything from zero to N. The number of active channels may approach N an appreciable fraction of the time in, say, a 12 channel system, whereas the chance of finding all channels active in an 1800 channel system is very small.

The other distribution which causes the rms voltage to vary in average value from minute to minute is our luck in drawing our sample of talker volumes. Any infinitely large sample would give the distribution shown by Figure 12-1, but any finite sample may be skewed - it may be weighted with more than its share of loud talkers, or less than its share. Like the changes in the number of talkers, this effect is more important (produces more variation in total load) in a 12 channel system than in a 1800 channel system.

It would be unreasonable to design an 1800 channel system so that it could carry 1800 zero vu talkers all talking at once; it would also be unreasonable to design a 24 channel system so that it overloaded whenever it had more than six talkers on it - and normal ones at that. The engineering compromise that has been adopted is to design each system so that 99% of the time it will not be overloaded. In other words, we take the probability of channels being active and the probability of getting skewed volume distributions, manipulate these probability functions** to find the load that will be exceeded only 1% of the time, and

---

* More efficient use of channels (for example by use of TASI or other channel sharing schemes, or more efficient switching) would change this value.

** How one does this manipulation is outside the scope of our discussion; let us merely assume it can be done. The reader who wants to know how it was done is referred to Monograph B1183.

design to carry that load. In the remainder of this discussion, this will be called the "maximum load" value. It follows that if we precisely meet this objective (usually we add a little margin for error, for aging effects, etc.) then the system will be overloaded during 36 seconds of each busy hour. We would expect this 36 seconds of overload to be divided into a number of periods of a few seconds each. Subscribers will, we believe, prefer this amount of degradation to an increase in telephone bills.

We are not yet ready to specify the sine wave test signal which the system must carry, however. We have discussed how to define the rms voltage of the "maximum load" (caused by telephone talkers) that we must carry, but there are two more points to consider: the ratio of instantaneous to rms voltage in the telephone multiplex case, and the definition of what constitutes overload when we have such a complex signal.

In considering these two questions we again find ourselves involved in probability manipulations, but now we are concerned with instantaneous voltage variations rather than the relatively slow variations in load caused by changes in the number of talkers and their volumes. To avoid confusion while we are discussing the fast variations, let us freeze the relatively slow rms voltage at its "maximum load" value.

Consider first the instantaneous voltage distribution of the carrier frequency signal. This signal is the sum of the message sidebands of all channels that are carrying speech in one direction. (A single sideband suppressed-carrier system is assumed here.) The relationship between the instantaneous voltage and the rms voltage for such a signal, as a probability distribution, is difficult to obtain analytically. However, for systems of more than a few channels, experiments show that the voltage distribution of the carrier-frequency signal is essentially the same as the voltage distribution of a signal which is the sum of the voice-frequency signals. This distribution of the sum of voice-frequency signals was obtained experimentally by combining the recorded speeches of constant-volume talkers. For small numbers of talkers the distribution is highly peaked; i.e., voltage peaks that are much larger than the RMS voltage occur frequently. However, for large numbers of talkers (e.g. greater than 64) the instantaneous voltage distribution approaches a normal curve.

For such a curve, we can readily compute the percentage of the time that the instantaneous voltage will exceed the rms voltage by a given amount. This gets us into a difficulty, however; this easily manipulated "normal" curve never stops - it tells us, falsely, that there is a small but finite chance of having extremely large instantaneous voltages.

Partly for this reason, partly because there really are quite large but very rare voltage excursions, we are forced to think about what we mean by overload. We find that we must define overload differently for complex signals than we do for a sine wave signal.

Suppose we have determined the sine wave signal voltage that will just "overload" an amplifier. To be specific, let us say it is that instantaneous signal voltage which will just cause grid current to flow. Any increase in drive will then, we find, produce very definite and noticeable distortion. (As anyone who has tried to measure overload knows, it is not always easy to pin-point the value of the signal voltage that produces overload, but let us pass over that difficulty.) At this point we call the reader's attention to a fact which he knows but might overlook. In an amplifier which is being overloaded by a sine wave, the percentage of the time that the instantaneous signal voltage is above the "critical" or "overload" value is actually very small. We describe the amplifier as "overloaded", however - not as "overloaded 5%, or 1%, of the time". In the sine wave case, then, the definition of the overload point is that the sine-wave peaks just reach some critical value of voltage.

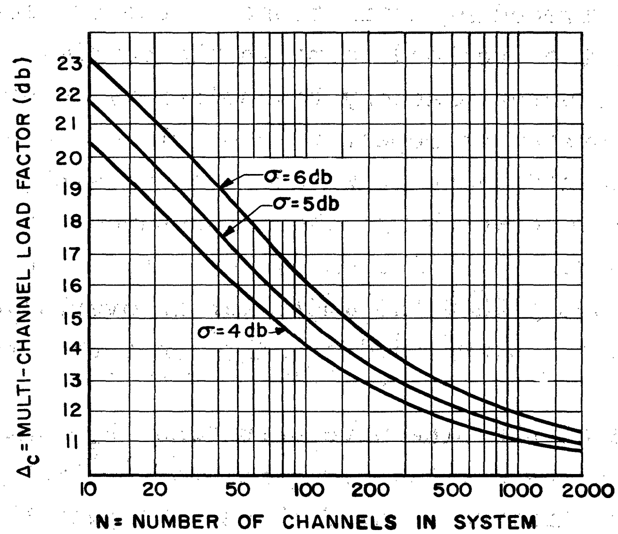
Now consider the case of a complex signal, such as a multi-talker speech load. Such a signal has the characteristic that the instantaneous voltage will, very rarely, reach very high peaks. Must we adjust levels so that these peaks never exceed the "overload voltage"- that is,the critical voltage for the sine wave case? As we might expect, the answer is "No". Very satisfactory transmission is obtained even if the peaks exceed the critical value, provided they don't do it too often or too long. In fact, there is no reasonable value of voltage that will not be exceeded once in a great while. We are forced to frame our definition of overload in different terms. We no longer speak of an overload voltage that is never reached, or to which our peaks are tangent. We speak instead of what percentage of the time we exceed some critical voltage. The voltage we found to be the critical one in the sine wave case is a convenient reference voltage to use.

To be quantitative, let us define ε as the integrated percentage of the time that the peaks exceed the critical voltage. The value of ε in a given case is a function of the long-term average rms signal, the corresponding instantaneous voltage distribution, and the critical voltage value. If the critical voltage is frequently exceeded (ε large) the amplifier will produce intolerable bursts of distortion (sometimes referred to as "cracking"). If, on the other hand, the signal is small so that the critical voltage is seldom or almost never exceeded, the amplifier will produce little distortion. What value of ε should we permit when we design a system? Dixon and Holbrook, from tests on a number of amplifiers, concluded that transmission through a repeater would be satisfactory (but just on the verge of being unsatisfactory) if the rapidly varying instantaneous voltage of the signal exceeds the critical voltage of the re-repeater 0.1% of the time. For larger values of ε, the distortion increased rapidly.

Using the preceding definitions of the critical voltage and ε, then, we can define the overload point in the complex signal case by saying that overload (i.e., unsatisfactory transmission) occurs when $\varepsilon > 0.1\%$. If ε is, say, 5%, the system or repeater is overloaded - not overloaded 5% of the time, but overloaded, just as we speak of an amplifier being "overloaded" by a sine wave even though the instantaneous voltage is excessive over only a small percentage of each cycle. If, on the other hand, $\varepsilon = 0.01\%$, we do not say that the amplifier is overloaded .01% of the time - we say it is not overloaded at all, since the transmission is seemingly perfect.

The value of the sine wave test tone "P_s" that the system should carry must be set taking all the foregoing factors into account. We might think of it as being arrived at as follows: Compute the average power of the load - this can be defined as the power corresponding to the average number of talkers (25% of the number of channels) when the talkers are all average power talkers (having volumes equal to $V_o + .115\,\sigma^2$). Add an allowance for the fact that the power varies during the busy hour because the number of talkers and the skewness of their volume distribution varies. This allowance will depend on the number of channels involved. This gets us to the 1% point on the distribution of power vs time during the busy hour, which we have defined as our "maximum load" value. Add another allowance for the peak factor of the signal, taking into account our ε definition of overload. The final value of P_s (a sine wave) has a peak voltage (3 db above rms) which is the value that the peaks of the telephone load voltage will equal or exceed 0.1% of the time that we have maximum load, if we let $\varepsilon = 0.1\%$. In system design, a more conservative limit of about 0.01% is assumed for ε. The peak voltage corresponding to P_s is then that voltage which, during maximum load periods, is equalled or exceeded 0.01% of the time by the peaks of the telephone load voltage.

R. N. Hunter has put this information into the form of a graph, shown in Figure 12-2. The abscissa is the total number of channels in the system, and the ordinate is a difference factor, Δ_c, which is the amount in db by which the required load capacity, P_s, must exceed the average <u>busy-hour</u> speech load. Δ_c is sometimes referred to as the "multichannel load factor". Curves are given for three values of the standard deviation of the talker volume distribution. In obtaining this set of curves, three assumptions were made: a), the probability that a channel is carrying speech in a given direction = 0.25; b), ε = 0.01%; and c), the system must carry the worst 1% combination of talker volumes and number of active channels. Let us now see how to make use of Figure 12-2.



Graph for Determination of Load Capacity

Figure 12-2

First, we must compute the average busy-hour speech load; this is the product of the average number of active channels during the busy hour, N_a, and the power of the average power talker, i.e., the power corresponding to the volume V_{op}. Hence, using (12-1), and recalling that the power corresponding to a zero vu talker is -1.4 dbm, the average busy hour speech load is:

$$\text{Avg. Pwr.} = (V_o + .115\sigma^2 + 10 \log N_a - 1.4) \text{ dbm} \qquad (12\text{-}2)$$

P_s is then obtained by evaluating Equation 12-2 and adding to this result the value of Δ_c read from Figure 12-2.

As an example, suppose we have a 36-channel submarine cable system with the talker volumes distributed according to Figure 12-1. What is P_s? From Figure 12-1,

$$V_o = -12.5 \text{ vu}$$

$$\sigma = 5 \text{ db}$$

Also, for a 36-channel system,

$$N_a \approx 0.25N = 9$$

Substituting into Equation 12-2 gives

$$\left.\begin{array}{l}\text{Avg. busy-hr.}\\ \text{speech load}\end{array}\right) = -12.5 + .115(5)^2 + 10 \log 9 - 1.4$$
$$= -1.5 \text{ dbm}$$

For a 36 channel system, Figure 12-2 shows the value of Δ_c to be 18 db. Therefore, the system must handle a -1.5+18=16.5 dbm sine wave at zero level without overloading.

Before we leave the topic of overload and start discussing modulation requirements, two more points might be mentioned. First, it should be noted that we have made the tacit assumption that each talker contributes equally to the probability of overload, no matter where his speech energy falls in the carrier frequency spectrum. If the telephone multiplex spectrum is not transmitted "flat", but is instead "pre-distorted" to get an improvement in signal-to-noise performance, the problem of specifying a test signal becomes more difficult. This refinement is discussed in Chapter 15.

The second point to be mentioned is the use of the "P_s" value in FM system design and analysis. If we want to transmit 36 channels over an FM system whose peak frequency deviation is to be one megacycle, the above example tells us that we should design so that +16.5 dbm of test signal at zero level will produce a one megacycle deviation of the carrier. This line of thought is made use of in Chapters 20 et seq.

Modulation Requirements

In earlier chapters, it has been stated that active channels will cross-modulate with each other to produce A+B, A-B, 2A-B, A+B-C, and other unwanted inter-modulation products, and that these will fall into various channels of the system, constituting interference in those channels. This interference may be evaluated in terms of its annoying effect, in dba, and added on a power-addition basis to the random noise to determine whether or not the channel meets noise objectives.

The number of possible intermodulation products in a given channel will be a function of the frequency allocation of the system. The magnitude of any particular intermodulation product will be a function of the non-linearity of the system, "H_x" (defined in Chapter 8). Our problem is to specify what values of H_{2A} and H_{A+B-C} (for example) will result in our meeting noise objectives for a given frequency allocation - or, in terms of previous discussions, to specify M_{2S} and M_{3S}.* To do so, we must be able to calculate the "modulation noise". The following pages describe a method developed by W. R. Bennett for making this calculation.

---

*In special cases modulation requirements may be set by considerations other than noise - e.g. intelligible crosstalk.

The method for calculating the noise (which is based on results obtained in systems transmitting relatively small numbers of channels) seeks answer to these questions:

a) How annoying is one x-type talker-modulation product arising from talkers of known constant-volume in systems of known non-linearity? Here "x" is the type of product (e.g. A+B, 2A-B, etc.). Somewhat surprisingly, this is answered by first finding how annoying the product is compared to an unwanted talker applied directly to the channel we are listening in, and then using a noise meter to find how annoying the unwanted talker is.

b) How many such products will fall in the channel of interest? This question faces the fact that not all of the channels are carrying active speech at the same time. The number sought is "N_x", the probable number of x-type products that will fall into the channel of interest during the busy hour.

c) How annoying are "N_x" products, compared to one?

d) What is the effect of the fact that talker volumes are not all the same, but distributed as in Figure 12-1?

e) What is the total modulation noise, then, for a system of given linearity? Or, what must the linearity be in order to limit the modulation noise to some given value, i.e., what are M_{2S} and M_{3S}.

The answers to a, b, c, and d will be different for each type of product (A+B, 2A-B, etc.) and must be found for each before the last question can be answered.

Single Talker-Product Noise: H_x, L_x, S_x, T_a

Suppose that a single x-type product formed by the intermodulation of talkers falls into one channel of a multi-channel system. (Such products will in fact spread over two or three channels - the A+B product, formed by two talkers each occupying a band four kc wide, would be a band eight kc wide. We will confine our attention to the channel in which most of the product energy falls.)

In order to get a measure of the system linearity, we replace the talkers by single frequency tones, and measure H_x as defined in Chapter 8. For this and following measurements, we transmit from a zero level point at system input and receive at a zero level point at the output.

We next load the disturbing channels with zero vu talkers, and listen to the x-type product in the disturbed channel. We then silence these talkers, apply a zero vu talker directly to the "disturbed" channel, and find the number of db by which we must suppress him in order to make him no more annoying than the x-type talker product was. Call this number of db "L_x".

If we made the same tests on another system of different linearity, we would find that "H_x" and "L_x" would vary together, their sum remaining a constant whose value depends only on the type product under consideration. Define this sum as "s_x". Thus

$$s_x = L_x + H_x. \tag{12-3}$$

Essentially, then, s_x is an empirical measure of the annoying effect of the x-type product formed by zero vu talkers. Bennett calls it the "Speech-Tone Modulation Factor".

By noise meter measurements, we can determine the "annoying effect" of a zero vu talker applied to the channel in which we are listening. We find that if we attenuate him by T_a db, he is just about as annoying as zero dba of any other type of noise, or causes just about as much transmission impairment. Numerically, we find T_a is approximately 82. In other words, as a "noise", a zero vu talker is equivalent to +82 dba. Therefore, we can determine W_{x_o}, which is the "noise", in dba, in the disturbed channel, caused by a single x-type modulation product formed by zero vu talkers.

$$\begin{aligned} W_{x_o} &= T_a - L_x \\ &= T_a - s_x + H_x \end{aligned} \tag{12-4}$$

<u>Multi-product Noise: 10 log N_x - ρ_x</u>

In general, there will be more than one x-type product falling in any given channel of a multi-channel system. Let the probable number of x-type products falling in the channel of interest during the busy hour be N_x. (The evaluation of N_x will be discussed later). At first glance it might seem that we should add to "W_{x_o}" a term "+10 log N_x" to find the noise in dba for the multi-product condition, since annoying effects expressed in dba are generally additive on a power basis. Actually it turns out that this is pessimistic - the annoying effect of a single product, measured by s_x, is in part due to its peak factor, and the peak energies of various fundamental talkers will not coincide in time. An experimentally determined correction term "ρ_x" is therefore needed. Having found this correction term by subjective tests, we can then write for the noise in dba in the disturbed channel caused by "N_x" products of x-type formed by many zero vu talkers applied to a system of linearity H_x:

$$W_{x_o,N} = T_a - s_x + H_x + 10 \log N_x - \rho_x \tag{12-5}$$

Values of s_x, ρ_x

Experimental values of s_x and ρ_x for the various products of interest are given in the following table.

Table 12-1

x	s_x	ρ_x
2A	3.0	-
A-B	13.3	3.0
A+B	6.1	3.0
3A	1.4	-
2A-B	- 2.2	3.0
A+B-C	+ 0.9	3.0

Very little experimental work has been done to determine ρ_x, and the values given are subject to considerable uncertainty. Further discussion of this point will be found in a later section of this chapter, where an alternative method of computing modulation noise is discussed.

Real Talker Product Distributions

Equation (12-5) gives the modulation noise caused by "reference products". (A reference product is a product formed by zero vu talkers.) As we have seen, the telephone load does not consist of zero vu talkers; real talkers form a normal distribution (see Figure 12-1) having a median "V_o" of -12.5 vu and a standard deviation "σ" of about 5 db. The following steps compute the modulation noise from such a distribution, relative to the noise from zero vu talkers.

The x-type products arising from fundamentals which are normally distributed in db will also be normal distributions in db. (We could, therefore, draw a distribution curve similar to Figure 12-1, but with a different average and standard deviation, for the modulation products of a given type that will fall into a given channel.) The average value of the product distribution will be related to the average value of the fundamentals by the usual power-series law. The standard deviation of the product magnitudes will be a function of the standard deviations of the fundamentals, the number of fundamentals which form each product, and the number of times they "enter".

The statistics of the product distribution are based on the fact that the product is formed by multiplying the fundamentals (adding db's). The 2A-B product, for example, is a third order product. If V_o is -12.5 vu, the average value of the distribution of 2A-B products will be $3V_o$, or -37.5 db, with respect to the reference product, i.e., a product of the same type formed by zero vu talkers. If we define "η_x" as "the order of the product", then we can generalize by saying that the average of the distribution of products from real talkers will be $\eta_x \cdot V_o$ db above the reference product. It will be recalled that the familiar formula for the standard deviation of a distribution formed by adding a number of normal distributions is:

$$\sigma_s = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 \cdots}$$

We can therefore write for the standard deviation of the distribution of products formed by real talkers the expression:

$$\sigma_s = \sigma\sqrt{\lambda_x}\,.$$

where σ is the standard deviation of the talker volumes and

$$\lambda_x = m_1^2 + m_2^2 + \cdots \text{ for a product of the type } m_1 A \pm m_2 B \pm \cdots \qquad (12\text{-}6)$$

For example, λ_x is 5 for 2A-B products, so that the standard deviation of the 2A-B product distribution will be 11.2 db if σ is 5 db.

The relations between average and standard deviation for the modulation products and the corresponding quantities V_o and σ for the fundamental are shown in Table 12-2.

Table 12-2

Modulation Product	Average Product in db Referred to Product from 0-vu Talkers ($\eta_x V_o$)	Standard Deviation in db ($\sigma\sqrt{\lambda_x}$)
2A	$2V_o$	2σ
A $\pm$ B	$2V_o$	$\sqrt{2}\sigma$
3A	$3V_o$	3σ
$\pm$2A $\pm$ B	$3V_o$	$\sqrt{5}\sigma$
A $\pm$ B $\pm$ C	$3V_o$	$\sqrt{3}\sigma$

Average Power Product

We cannot multiply the average value of the distributions of Table 12-2 by N_x in finding total noise, since like volume these distributions are normal in db rather than power. Recalling Equation 12-2 we find that for the distributions of Table 12-2 the average power product will be given by

$$V_{avg.\ pwr.} = \eta_x V_o + .115\ \sigma_s^2 = \eta_x V_o + .115\ \lambda_x \sigma^2$$

in db relative to a product formed by zero vu talkers. Adding this correction term to the formula of Equation 12-5 will give us the modulation noise caused by real talkers. We thus obtain*

$$W_x = T_a - s_x + H_x + 10 \log N_x - \rho_x + \eta_x V_o + .115\ \lambda_x \sigma^2 \qquad (12\text{-}8)$$

where W_x is the noise in dba in the channel of interest, caused by N_x products of x-type, arising from talkers normally distributed in db around a median value V_o with a standard deviation of σ. (It might be repeated that the talkers are applied at the zero level point at the input, and noise measured at a zero level point at the output, of the system).

For a given frequency allocation and talker population, Equation 12-8 reduces to $W_x = H_x$ + (constant). For a given transmission plan, this equation can be used to evaluate modulation noise from computed or measured values of H_x. Alternatively, it can be used to derive system non-linearity requirements by finding the limiting value of non-linearity (e.g., M_{2S}) for which the system will just satisfy a given modulation noise requirement with zero margin - under which condition $H_{2A} = M_{2S}$, for example.

Probable Number of Products: N_x

To evaluate the noise arising from modulation products, the number of talker products that will probably fall in the channel of interest during the busiest hours must be determined. If the probability that a channel is carrying speech is τ, and the total possible number of x-type products that can fall into the channel of interest is P_x, then the probable number of products is

$$N_x = P_x \cdot \tau^{\mu_x} \qquad (12\text{-}9)$$

---

*Bennett's final equation is in slightly different form. Recalling that the volume "V_{op}" corresponding to the average power talker is

$$V_{op} = V_o + .115\ \sigma^2$$

he solves this for V_o and substitutes this value in Equation 12-8. The last two terms then become

$$\eta_x V_o + .115\ \lambda_x \sigma^2 = \eta_x V_{op} + .115\ (\lambda_x - \eta_x)\ \sigma^2$$

These reduce to $\eta_x V_{op}$ for A $\pm$ B and A $\pm$ B $\pm$ C products, since for these, $\lambda_x = \eta_x$.

where μ_x is the number of fundamentals needed to form an x-type product (e.g., $\mu_{2A-B} = 2$ and $\mu_{A+B-C} = 3$.)

Thus if P_{2A-B} for the k-th channel is found to be 160, and the value of τ is 0.25, then the probability that any given possible product is present is $(0.25)^2$, and the probable number of products N_{2A-B} is 10.

The Product Count

Bennett provides formulae from which we can compute the total number of x-type products which can possibly fall in the channel of interest. These formulae assume that we have the frequency allocation indicated on Figure 12-3. Carrier slots are located at intervals of p cycles from zero frequency. The system transmits N channels with carrier frequencies from n_1p to n_2p inclusive. Channels are identified by the carrier with which they are associated; the channel of interest is identified by the k'th carrier (from zero frequency) or M carriers from the lower band edge, such that

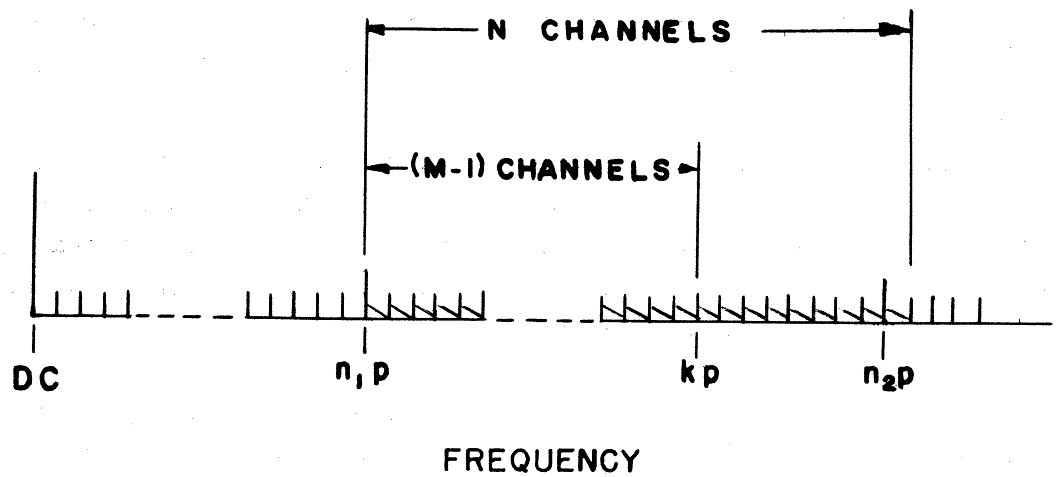
$$k = M-1+n_1$$

The "-1" occurs because the first transmitted channel is identified by the carrier, M=1, $k=n_1$. The terminology is generalized so that the k'th carrier may lie outside the transmitted band. Since we are finding the total possible number of products that can fall in the channel of interest, it is assumed that every channel in the system (except the channel of interest) is transmitting energy.

While the exact equations are straightforward to understand and use, they are rather long, and repetitive computations become laborious. Variations in bandwidth, and therefore number of channels transmitted, are commonplace during the preliminary design stage of a new system. Counting products in these circumstances can be a time consuming effort.

A useful set of approximations to Bennett's formulas is given by the graphs on Figure 12-4. The accuracy of these graphs is very poor for narrow band systems. Errors of 50 to 100% may be expected for a 10-channel system. The accuracy improves rapidly with bandwidth and for a 100-channel system is somewhat better than $\pm 5\%$. For a 1000-channel system, the error is negligible.

The noise quantities used for the abcissa of these graphs call for comment. When the abcissa is M/N, the number of products that fall



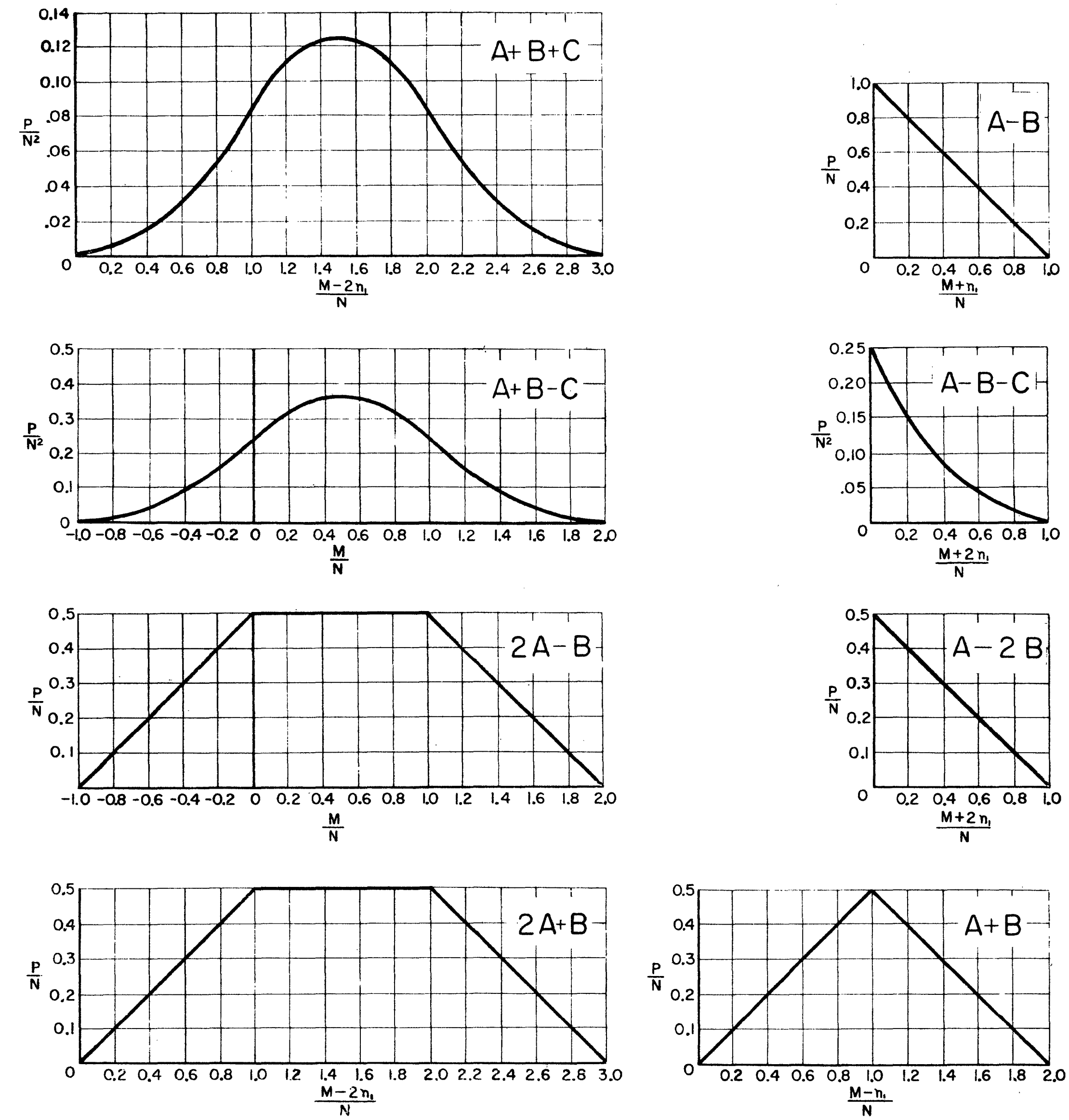
Frequency Allocation for Product Count

Figure 12-3

in any channel is independent of the position in the frequency spectrum occupied by an N-channel system. In the case of other products - for example A+B, where the abcissa is $\frac{M-n_1}{N}$, particular frequency allocations will minimize the number of products. Suppose that for any value of N we make n_1 equal to 1/2 n_2. In this case no A+B products will fall in-band. But as long as N is unchanged, there is no such escape from 2A-B or A+B-C products, where the abcissa is M/N.

Alternative Method of Computing W_x

For computing noise in very wide-band systems where the probable number of products may be in the hundreds or thousands, it does not seem reasonable to start from an uncertain determination of the single product value of s_x, and correct it by an even more uncertain value of ρ_x. It seems more straightforward to assume that by the time the probable number of superimposed products exceeds twenty or thirty, they become indistinguishable from random noise, except for a frequency weighting which can be computed. This would appear to be particularly valid if we expect to have the masking effect of an equal or greater magnitude of truly random noise power in the disturbed channel.



Legend

P = Total possible products of a given type

N = Number of channels transmitted with carriers n_1p to n_2p inclusive, where p is the base frequency in cycles per second and n_1 and n_2 are integers with $n_2 > n_1$

k = Channel of interest associated with carrier kp within the fundamental band $n_1 < k < n_2$. When values of $k < n_1$ or $k > n_2$ are of interest, the relationships are equally valid.

M = Channel of interest associated with carrier Mp where $M = k - n_1 + 1$

Note: The number of products is zero outside limits of curves.

Approximate Modulation Product Count
For Multichannel Telephone Systems

Figure 12-4

We then have two bases for computing modulation noise:

a) the method involving experimental determination of s_x and ρ_x, which has been discussed in the preceding sections,

b) the random noise basis, which involves computing power and frequency weighting of the products as if they were shaped random noise contributors. This computation is described below.

From the preceding material, the power of the modulation products in the k'th channel, at the zero level point, will be (recall that the average power of a zero vu talker is -1.4 dbm):

$$\text{Power in dbm} = \eta_x(V_o - 1.4) + .115\,\lambda_x\,\sigma^2 + 10 \log N_x + H_x \qquad (12\text{-}10)$$

To express this in dba, we recall that 0 dbm of white noise corresponds to +82 dba. But the power we are dealing with is not flat vs frequency in the disturbed channel. Assume that its frequency weighting is such as to make it less annoying than white noise by C_w db. Then the modulation noise in dba is

$$W_x = \eta_x V_o + .115\,\lambda_x\,\sigma^2 - 1.4\eta_x + 10 \log N_x + H_x + 82 - C_w \qquad (12\text{-}11)$$

Modulation noise computed on this basis can, when the number of products is large, be compared with the results of computations based on equation (12-8). If we assume that the other experimental determinations are fairly firm, and that most of the uncertainties lie in the values of s_x and ρ_x, we can consider the "noise power" method as a computational check of the experimental determination of these two quantities. Comparing (12-8) and (12-11), we find, since $T_a = 82$, that in db,

$$s_x + \rho_x = 1.4\eta_x + C_w$$

In Table 12-3 values of $s_x + \rho_x$ thus obtained are compared with those of Table 12-1. It might be expected that the noise power method would yield optimistic results, since the syllabic peaked character of the products is ignored. Instead we find the noise power method predicts that A-B products will be 11.3 db, and A+B products 4.1 db, more annoying than the experimental determination of s_x and ρ_x would predict. It is believed that this is probably because C_w is small (about one db) when modern instruments and noise weightings are used, whereas with old instruments and weightings more of the product energy fell into insentive portions of the spectrum. It would seem that the values computed by the noise power method, rather than the old experimental values should be used.

Table 12-3

x	$s_x + \rho_x$ Noise Power Assumption	$s_x + \rho_x$ from Table 12-1
A-B	5.0	16.3
A+B	5.0	9.1
2A-B	5.5	0.8
A+B-C	7.0	3.9

Bibliography

1. V. Subrizi, "A Speech Volume Survey on Telephone Message Circuits", Bell Laboratories Record, August 1953.
2. Perry, "Design for Optimization of Multi-channel Radio Systems", Communication and Electronics, (AIEE), November 1957.
3. B. D. Holbrook & J. T. Dixon, "Load Rating Theory for Multichannel Amplifiers", Monograph B-1183.
4. W. R. Bennett, "Cross Modulation in Multichannel Amplifiers", Monograph B-1252.

ILLUSTRATIVE EXAMPLE - CHAPTER 12

1. A 2000 mile telephone system is to be equipped to carry 200 channels, each occupying a 4 kc band. The lowest transmitted frequency is to be 200 kc, the highest 1 mc. The system is to meet the total noise requirement discussed in Chapter 2. The total noise requirement is to be divided between terminals and high-frequency line, allocating half the noise to terminals and half to line.

   Find P_S, M_{2S} and M_{3S}. For values of s_x and p_x, use Table 12-3 values based on the "noise power assumption".

Solution:

1. Computation of Modulation and Load Carrying Capacity Requirements

From Chapter 2 the total noise allocation for a 4000 mile system is 38 dba at a zero level point. For this problem this is to be divided equally between terminals and the high frequency line. (A more usual division would be about one quarter of the noise to be contributed by terminals, three quarters by the high frequency line).

Total:	+38 dba
Terminals:	35 dba
4000 mile line:	35 dba

If 4000 mile requirements are to be met, 2000 miles of line can contribute only one-half the line noise, or +32 dba. (In the general case, where a transmission facility is to be used as part of the Bell System plant, the noise contributed by each facility should be in proportion to its length as part of a 4000 mile circuit, if the plant is to meet over-all 4000 mile requirements. Special arguments might be made for relaxing requirements on systems designed for short end-link use. The system we are considering is obviously not an end-link but a long toll facility, since it is specified to carry 200 channels for 2000 miles).

It should be anticipated that either second or third order modulation will be controlling as discussed in Chapter 10. Hence the division of the 32 dba line requirement will be shared by random noise and modulation noise in proportions depending on which type of modulation is dominant.

If second order modulation noise turns out to be dominant, thermal noise should be 32-3 or 29 dba, whence N_S would be -53 dbm, and second order modulation noise would be set equal to thermal, whence

$$W_{A+B} = 32-3 = 29 \text{ dba}$$

and for third order modulation dominant

$$N_S = -82+(32 - 1.8) \text{ dbm} = -51.8 \text{ dbm}$$

$$W_{A+B-C} = 32 - 4.8 = 27.2 \text{ dba}$$

In order to determine M_{2S} and M_{3S} consider the A+B-C products and the A+B products in the top frequency channel where the noise will be highest because the cable loss and hence thermal noise are greatest. Modulation noise may be greater in lower frequency channels because of changes in product type and probable number of products vs carrier frequency. Consider Figure 10-4, which shows noise sources vs frequency in a system of similar frequency allocation, and Figure 12-4 We can see that at worst the variation of modulation noise vs frequency will make some lower channel just as noisy as the top channel. Certainly levels on the line cannot be properly set using values of thermal and modulation noise in lower frequency channels. Within our assumptions (flat "voltage levels" on the grid of the output stage, and flat feedback) the top channel is the one to use for laying out a transmission plan. Other channels can be checked later. In the top channel, furthermore, it can be quickly deduced that the important products will be A+B and A+B-C. Third order products that do not add in phase will be small at the end of a long repeatered system, compared to in phase addition types. 2A-B will be well below A+B-C products in both number and magnitude. (In narrower band systems, this may not be true).

The first step is to use Figure 12-4 to find number of products. In this example, assuming upper sidebands are transmitted, $n_1 = 50$ at 200 kc, n_2 at 996 kc = 249. The channel of interest is $k = 249$, whence $M = 200$. Therefore $M/N = 1$ and we read $P/N^2 = .25$ for A+B-C.
Therefore: $P = (.25)\ (200)^2 = 10{,}000$
From Equation 12-9

$$N_X = P_x \cdot \tau^{\mu x} = \frac{10000}{64} = 156$$

since

$$\mu_{A+B-C} = 3, \ \tau = .25$$

In Figure 12-4 for A+B in top channel

$$\frac{M-n_1}{N} = \frac{200-50}{200} = \frac{150}{200} = .75$$

from curve, $P/N = .37$

$$P = (.37)\ (200) = 74$$

and since

$$\mu_{A+B} = 2, \quad \tau^{\mu_x} = (.25)^2$$

we find, $N_{A+B} = \frac{74}{16} = 4.6$

Next use Equation 12-8 (footnote form) to find relation between modulation noise and H_x.

$$W_x = H_x + T_a - s_x + 10 \log N_x - \rho_x + \eta_x V_{op}$$

where

	A+B-C	A+B	
Ta	82	82	page 12-11
$-(s_x+\rho_x)$	- 5.7	- 3.8	Table 12-3
η_x	3	2	Table 12-2
V_{op}	- 9.6	- 9.6	page 12-3
10 log N_x	21.9	6.6	above

Whence

$$W_{A+B-C} = H_{A+B-C} + 69.4$$

$$= H_{3A} + 85 \text{ (using the fiction that third harmonic adds in-phase)}$$

$$W_{A+B} = H_{A+B} + 65.6$$

$$= H_{2A} + 71.6$$

Since M_{2S} is the value of H_{2A} for which second order modulation noise is the maximum permitted,

$$M_{2S} = 29 - 71.6 = -42.6$$

Similarly, $M_{3S} = 27.2 - 85.0 = -57.8$

To find P_s, first find average power from equation 12-2 (page 12-5). Avg. Pwr. = $V_o + .115\,\sigma^2 + 10 \log N_a - 1.4$ dbm at zero level.
where

$V_o = -12.5$ vu

$\sigma = 5$ db

$N_a = \tau N = .25\ (200) = 50$

$$
\begin{aligned}
\text{Avg. Pwr.} &= -12.5 + (.115)\ (5)^2 + 10 \log 50 - 1.4 \\
&= -12.5 + 2.88 + 17 - 1.4 \\
&= +6 \text{ dbm}
\end{aligned}
$$

Next use Figure 12-2 (page 12-6) to find the multi-channel load factor Δ_C. For $\sigma = 5$ db, N = 200 channels, we read $\Delta_C = 13.5$ db.
Then P_s = Avg. Power + 13.5
$= +6 + 13.5 = +19.5$ dbm

Chapter 13

FEEDBACK REPEATER DESIGN

Introduction

This chapter will consider the design of linear AM repeaters, not in the detail appropriate to a circuit design course, but as a part of our more general problem of multi-repeater system design. Perhaps the most outstanding difference between the two approaches is that in designing a repeater for use in a transmission system we do not know in advance either the objectives to be met or the components available to us. Instead the objectives themselves are to be developed as a part of an overall pattern of system, repeater and component objectives. In the framing of these objectives, preliminary repeater designs will play an important role.

Furthermore, the importance of the repeater design in the system plan leads us to put unusual emphasis on achieving the optimum performance from the repeater. This in turn will often make it profitable to develop new vacuum tubes and other components for a specific system application. One of the functions of the repeater designer, then, is to assess the benefits and penalties of various suggested component designs as they affect repeater performance. The system designer, in turn, will assess the merits of various repeater designs, in terms of their effect on band-width, repeater spacing, and signal-to-noise performance. Again it must be emphasized that system objectives and design, repeater design, and component design are inter-dependent problems.

General Design Procedure

The system designer will need values for the following quantities, usually as functions of frequency, for repeater designs which satisfy the fundamental requirement of compensating the cable loss for the frequency band and repeater spacing which have been tentatively selected.

a) Thermal or random noise contribution of each repeater.
b) Second and third order modulation performance of repeater (in the case of feedback amplifiers, this may be broken down into tube performance and amount of feedback).
c) Load carrying capacity.
d) Estimated manufacturing deviations.
e) Estimated aging and temperature effects.

In particular cases, other information may be required, such as delay distortion, power frequency modulation, regulation accuracy, etc.

In the beginning the repeater designer must rough out tentative designs, carrying them far enough so that good numerical performance estimates of the most important of the above quantities can be derived for system design use. In order to do this, he will have to make many assumptions, but must avoid becoming wedded to any of them. Such tentative assumptions must rather be subject to continuing re-examination in the light of continued system and component re-evaluations.

Eventually, however, the choices become sufficiently clear-cut to permit starting the second phase of the repeater design process. The detailed design of the chosen amplifier configuration makes use of the designs of tubes and other components which were settled on as a result of the rougher preceding amplifier designs. Estimates will still fluctuate, but with decreasing range, while an increasing tempo of laboratory work eliminates many of the initial uncertainties. Even in this stage, however, fundamental assumptions and decisions must be frequently reviewed and re-examined.

The final stage, after the paper and laboratory designs are complete, is the realization of the expected performance from production models. Here it is dangerous to assume that objectives will be met without continuing effort. Much of the performance that took so much design effort to obtain can be easily lost in the course of production by ill-advised economies or inaccurate testing.

During the first preliminary stages of roughing out initial designs, advantage can be taken of a number of principles relating achievable performance to circuit constants. Although these principles are indispensable aids, it is still true that the validity of the many assumptions required at the beginning, and hence the speed and efficiency with which usable designs are crystallized, will depend to a great extent on the skill, judgment and experience of the circuit designer. Perhaps the best way of making clear the processes involved is to examine in some detail a fairly typical problem.

Illustrative Example

Let us take as our illustrative problem the design of a repeater for a submarine cable system, to employ electron tubes as the amplifying devices. The system development in this case is a long-range project, to involve development of new types of cable and components if our studies show such developments to be economically justified. This is to be a telephone system and it should meet so-called 4000 mile requirements. As usual we probably will not be transmitting over this single facility for 4000 miles. Our objective, however, is to design the system so that a 2000 mile submarine system combined with a similarly good 2000 mile land system will just meet 4000 mile requirements.

Aside from this broad general objective, very few of the design parameters of the system are really fixed during the initial stages of such a problem. However, it is necessary before we can start at all to assume what we hope are reasonable values for bandwidth to be transmitted, repeater spacing, the impedance and loss versus frequency characteristics of the cable to be used, the voltage limitations on the cable, and the available designs of vacuum tubes.

Here it should be kept in mind that going from a particular system layout to a similar system, say of greater bandwidth, involves more than just some arbitrary change in repeater spacing or in signal levels on the high frequency line. The use of Eqs. 10-6, 10-7, 10-8 with proper ultimate values for the various parameters, system requirements and margins, leads one into a very tight situation for any particular bandwidth choice, where repeater spacing or signal levels cannot be adjusted up or down without failing to meet requirements by the desired margins. In other words, "n" and "C" are uniquely determined. In the early stages of the layout there may be some uncertainty (at most about 3 db) as to what value of "N_S" to use in the three equations referred to. Yet the final solution is a definite one.

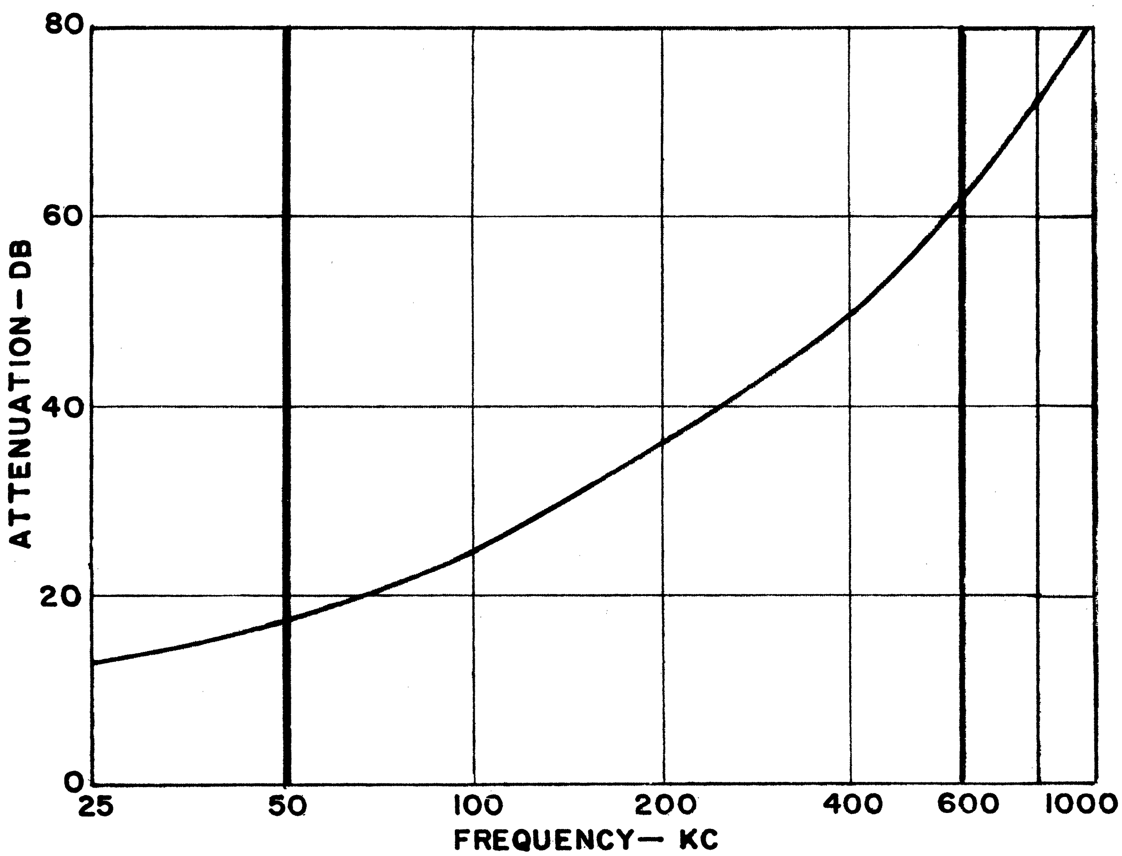
<u>Preliminary Design Steps</u>

A good starting point is to select a practical cable. In this case we shall consider an armored coaxial having an outer conductor diameter of 0.620 inches. Figure 13-1 gives the loss characteristic versus frequency for a 20 nautical mile length of this cable. The loss varies from 18 db per 20 nautical miles at 50 KC, to 72 db at 800 KC, following the square root of frequency law.* In giving this data we have tacitly made our first assumption, since we have roughly defined

---

*This is always approximately true for coaxial cable; in the final polished design the fact that any physical coaxial will show a characteristic deviation from a precise square root of frequency shape must be taken into account.

the bandwidth in which we will be interested. Our roughed out design will tell us whether or not our assumption was reasonable. Later we shall probably consider other possible cables in an effort to optimize system bandwidth, reliability, and cost. The same re-examinist philosophy applies to the vacuum tubes whose assumed characteristics are shown in Table I. We can always ask for an improved vacuum tube design - improved in the sense that the tube is better suited to our purposes. We might for example find it very advantageous to obtain a higher figure of merit. On the other hand we may conclude that it would be better to sacrifice some figure of merit in favor of lower voltage operation or increased life.



ATTENUATION VS. FREQUENCY
FOR 0.620-INCH SUBMARINE CABLE
Length equals 20 nautical miles

FIGURE 13-1

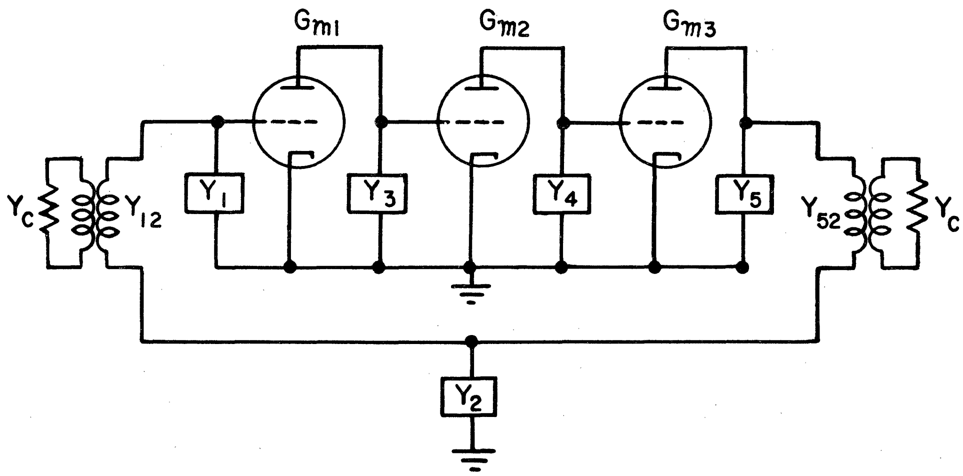
TABLE 13-1

Assumed Characteristics of Vacuum Tubes

Transconductance	3400 μmhos
Input Capacity (hot)	12 μμf
Output Capacity	5 μμf
Equivalent Noise Resistance	2000 ohms
Cathode Current	3.5 ma.
Plate-Cathode Voltage	40 volts
Heater Voltage	11.4 - 13.0 volts
Heater Current	.23 - .25 amp.
Grid-Cathode Spacing	8.5 mils.

We must now decide, very tentatively, on an amplifier configuration and a repeater spacing, and the bandwidth to be transmitted. At this stage of the process we are interested in what is possible rather than in the exact solution of the circuit problems involved. It will therefore be some time before we define exactly the circuit elements which we would use in the repeater design that we shall rough out.

Experience with similar systems leads us to assume that in order to meet signal-to-noise requirements it will be necessary to use feedback to suppress intermodulation products in the repeaters. This feedback will, of course, also help to stabilize the gain of the repeaters against manufacturing deviations and tube and element aging, thus reducing the amount of misalignment in the system and further improving the signal-to-noise ratio. Since this is to be a submarine cable system where repair is difficult, simplicity and a minimum number of elements are desirable goals. We shall therefore select a fairly simple amplifier configuration for initial study. A three stage series feedback amplifier is a good first choice. It has the merit of being the easiest feedback circuit to design, which means that we shall start getting rough estimates of performance more quickly than we would from a more complex multi-loop configuration. Furthermore, experience shows that although more complex circuits may give better performance, the difference will not be great enough to make substantial changes in system layout. Figure 13-2 shows the configuration and gives the formulae for feedback and insertion gain of a series feedback amplifier as derived from nodal analysis.



Feedback:

$$-\mu\beta = T = F - 1 = \frac{\Delta}{\Delta^{o}} - 1 = P_i \frac{G_{m1}}{Y_3} \frac{G_{m2}}{Y_4} \frac{G_{m3}}{Y_2'} P_o$$

$$P_i = \frac{Y_{12}}{Y_1+Y_{12}} \qquad P_o = \frac{Y_{52}}{Y_5+Y_{52}}$$

$$Y_2' = Y_2 + \frac{Y_1 \, Y_{12}}{Y_1+Y_{12}} + \frac{Y_5 \, Y_{52}}{Y_5+Y_{52}}$$

Insertion Gain:

$$G_A \doteq 2\,\theta_i\,\theta_o \frac{Y_2}{Y_c} \left(\frac{T}{1+T}\right)$$

$$\theta_i = \frac{E_{12}}{E_o} \qquad \theta_o = \frac{I_c}{I_{52}}$$

E_{12} = open circuit voltage in series with Y_{12} appearing in response to open circuit input cable voltage E_o.

I_c = current delivered to the output cable admittance in response to the current I_{52} through Y_{52}.

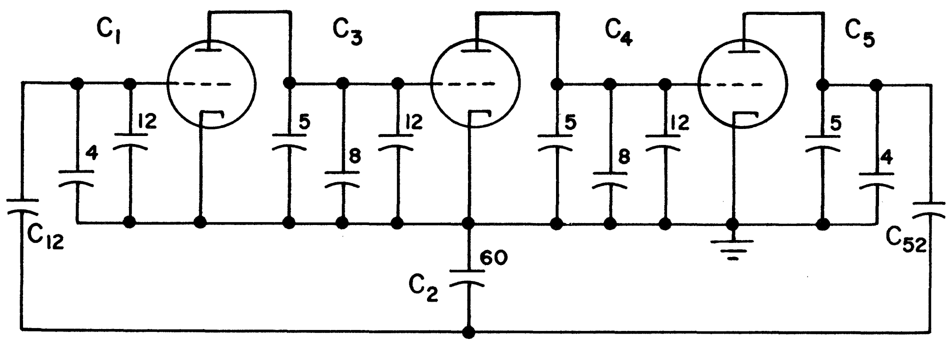
FIGURE 13-2

For the bandwidth, let us assume that we will start at 50 KC and transmit up to 600 KC. Past experience tells us that somewhere between 40 and 60 db of repeater gain is a reasonable amount to call for, so we might select a 20 nautical mile repeater spacing for the first round. This bandwidth and spacing calls for 62.3 db of gain at 600 KC (Fig. 13-1), which is on the high side of what we may guess to be the practicable range, but a long span is a very desirable goal in a submarine system.

The next step is to compute some of the important repeater constants; to do this we need to know something about the parasitic circuit capacities of our repeater. These are rather difficult to specify except from a knowledge of previous circuits of the same general physical configurations. There are two suggested methods of building repeaters for submarine cable systems. One is to build the amplifier as a series of protected cylinders which form an integrated flexible part of the cable itself - a sort of bulge in the cable. This might be concisely referred to as the articulated design. The other method is to make the repeater a much bigger rigid bulge in the cable - cylindrical in shape - which gives considerably more freedom in the layout of vacuum tubes and circuit elements than the six foot long, one and a half inch diameter articulated design gives. We might choose either of these physical configurations for our first round.

From an electrical standpoint, the rigid design offers substantial advantages. The component parts of the repeater may be grouped closely together thereby minimizing the parasitic capacities and lead inductances which limit performance at high frequencies. We shall therefore concentrate on the rigid container approach as offering possibility of an improved system. If, however, later studies indicate that this design is impractical from a cable laying standpoint, we must then redesign the amplifier to fit the alternative articulated design. In a land system, similar choices would usually have to be made - for example, between printed circuits, or a mechanical design adapted to automation, and conventional wiring.

Laboratory work involving the construction of models and the measurement of the parasitic capacities which limit obtainable performance is essential at an early stage. Since we are at an even earlier stage, however, we shall go ahead on the basis of guesses for the moment. The parasitic capacities associated with the tubes (see Table I) and the estimated circuit element and wiring parasitic capacities are shown on Figure 13-3.



C_1 = 4+12=16 μμf
C_2 = 60 μμf
C_3 = 5+8+12=25 μμf
C_4 = 5+8+12=25 μμf
C_5 = 5+4=9 μμf
C_{12} and C_{52} to be determined

ESTIMATED PARASITIC CAPACITIES

FIGURE 13-3

The values for C_{12} and C_{52}, as yet undetermined, will be discussed later.

Theorems Relating Circuit Capacity to Gain and Bandwidth

At this point it is convenient to review some of the relationships between circuit parameters and physically realizable performance.

No general rules can be set down as to how practical or impractical it may be in a given case to achieve the maxima listed below. In general, one may say that characteristics which either peak or slope off at high frequencies are more easily obtained than flat characteristics. It might also be noted that all the theorems assume that circuits degenerate to capacities at high frequencies, whereas in fact the inductances of lead wires also become very important at frequencies around 100 megacycles. In certain cases this may result in an appreciable reduction in realizable performance

a) Transformer Gain

The voltage step-up of the input coupling transformer, and the current step-up of the output coupling transformer, are important contributors to the total repeater gain.

For ideal transformers, these gains, in db, are equal to $10 \log_{10} R_H/R_C$, where R_H is the resistive component of the impedance seen looking into the high side of the coupling transformer when the low side is terminated in the cable impedance R_C. Unfortunately it is not possible to increase these gains indefinitely by merely increasing R_H, since in the presence of a given shunt capacity C across the high side of the coupling network, R_H is limited by the resistance integral theorem*

$$\int_0^\infty R_H d\omega = \frac{\pi}{2C} \tag{13-1}$$

b) Interstage Gain

For a two-terminal interstage, we have the limitation on maximum gain given by the theorem that in the presence of a given shunt capacity C the interstage gain area within a band of frequencies from zero to f_o cycles can equal but not exceed the area which would be obtained from the shunt capacity alone, the area in each case being obtained by plotting db gain versus Φ where**

$$\Phi = \sin^{-1} \frac{f}{f_o} \tag{13-2}$$

Figure 13-4 illustrates the type of plot referred to, for the well known case of the flat interstage. The great usefulness of the Φ scale lies, of course, in the fact that it is applicable for any shape of gain vs. frequency curve and for any arbitrary choice of f_o.

c) Maximum Obtainable Flat Feedback***

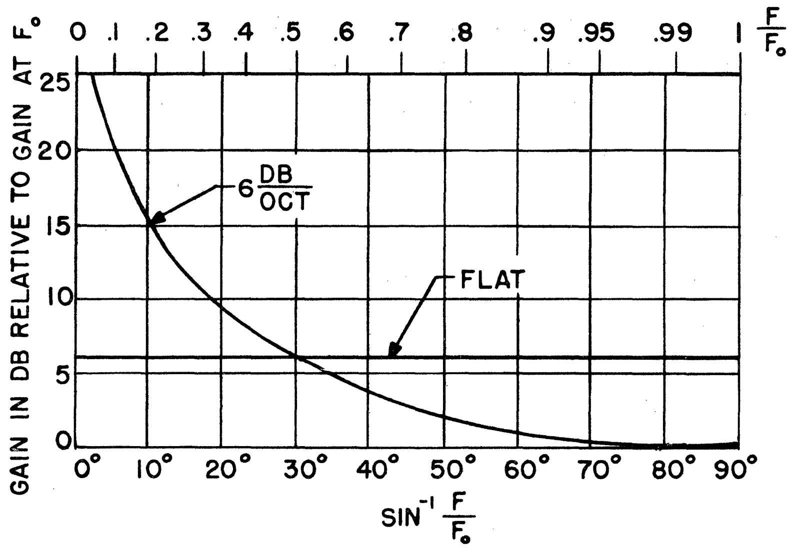
In a single-loop circuit, the flat feedback obtainable over a given band from zero cycles to f_o can be found from an equation of Bode's. However, it is more informative to use the well known graphical method of Figure 13-9, since it better illustrates the frequency band of interest in the cutoff region - a matter of considerable practical concern.

---

*H. W. Bode, "Network Analysis and Feedback Amplifier Design", pp. 280-283.

**op. cit. p. 415.

***op. cit. pp. 464-468.



THE Φ SCALE

FIGURE 13-4

d) Maximum Obtainable Shaped Feedback*

The Φ scale used for plotting interstage gain is also useful in computing the maximum obtainable shaped feedback. If the same phase margin is to be obtained, the area under the shaped feedback characteristic up to the top transmitted frequency can equal but not exceed the area under the flat feedback curve when both are plotted against Φ.

Insertion Gain

With the foregoing theorems in mind, we are now ready to return to our consideration of the design procedure. It is an obvious requirement that the gain of the repeater must match the loss of the cable section over the frequency band to be transmitted. The accuracy of this match we shall ignore for the moment, and concentrate on initial decisions on impedance level and such other gross questions. The formula for the insertion gain given on Figure 13-2 (expressed here in db) is:

$$G_A = 20 \log \left[2\ \theta_i\ \theta_o \frac{Z_c}{Z_\beta} \frac{T}{(1 + T)} \right] \qquad (13\text{-}3)$$

---

*op. cit. p. 456.

We are faced with three problems: 1) to achieve the required maximum insertion gain at the top transmitted frequency; 2) to distribute the required insertion gain shaping among the various networks so as to match the cable slope; 3) to obtain the maximum possible feedback. The insertion gain formula includes three important gain parameters -- beta circuit impedance and two coupling network gains. (Z_c and T/1+T can obviously be neglected in this discussion.) Since the over-all repeater insertion gain characteristic must match the loss of the cable, it follows that the required slope is divided among these three networks.*

Coupling Networks

Let us first consider the design of the coupling networks which couple the vacuum tubes to the cable impedance at the input and output of the amplifier. In addition to their importance in achieving the desired insertion gain, noise figure, and modulation performance, the coupling networks play an important role in the design of the feedback loop since they determine the potentiometer terms (P_i and P_o of Figure (13-2).** They are also, in general, the most sensitive of the amplifier networks, in the sense that element deviations in the coupling networks are not suppressed by feedback, tend to be large, and are therefore dominant contributors to misalignment. Optimum design of the coupling networks is therefore an important objective.

If we know the high side capacity of these networks (C_{12} or C_{52} in Figure 13-3), we can, from the resistance integral theorem, find the maximum gain which could be gotten from the coupling networks over the transmitted band. It is not advisable, however, to attempt to achieve this maximum. If we attempt to do so, we find that the high-frequency gain of the network becomes extremely sensitive to small

---

*Some of the slope may be allocated to a constant-R equalizer preceding the amplifier, but this is to be avoided if possible in order to optimize signal-to-noise performance.

**Near upper band edge, these potentiometer terms may become large losses in transmission around the feedback path, and tend to make the circuit conditionally stable - or, at best, lead to difficulty in obtaining the theoretically possible feedback. See also BSTJ Vol. 32, July 1953, p. 890.

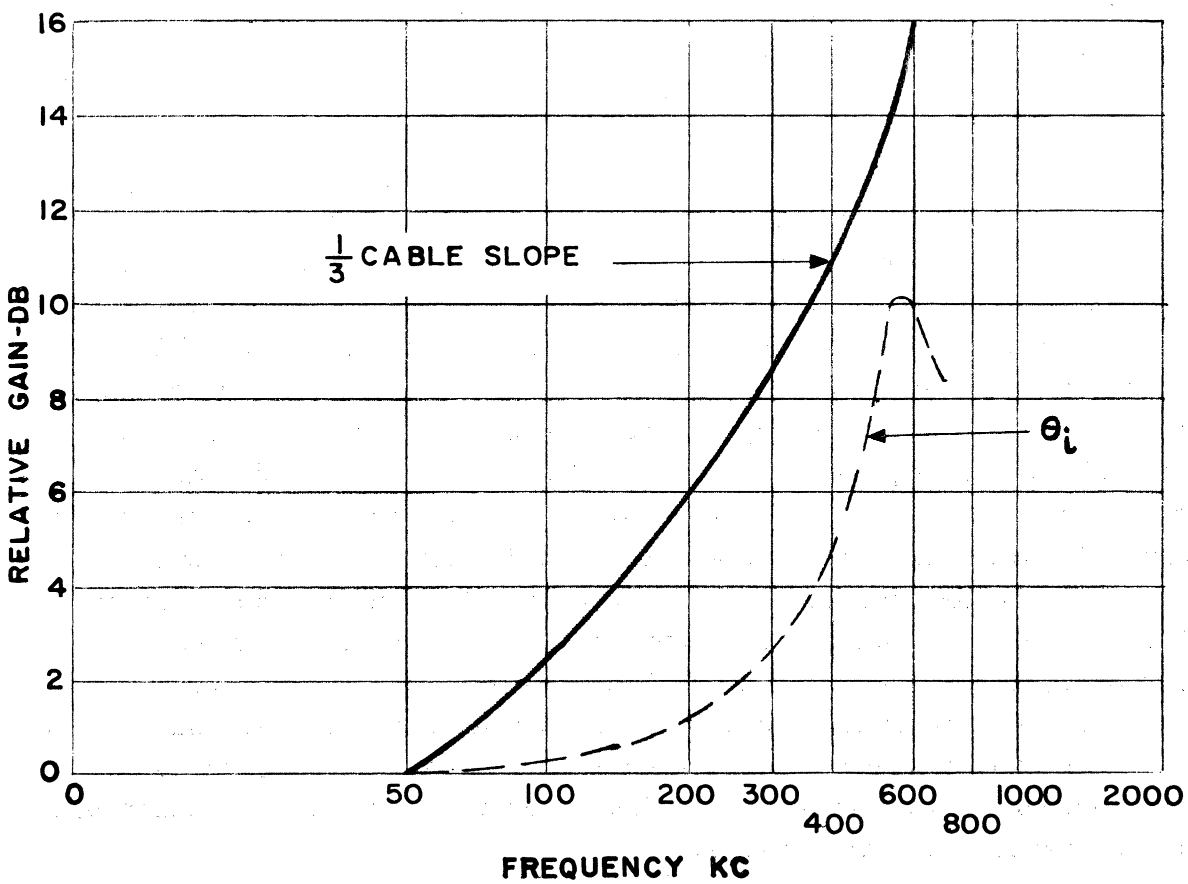
variations in any of the element values. Since in transformer manufacture such variations cannot be avoided, and since gain deviations are in general one of the most important problems in a multi-repeater transmission system, it is advisable to settle for somewhat less than the maximum achievable gain in order to minimize equalization difficulties.

Experience indicates that a reasonable compromise is to allocate about one-half the total available resistance gain area to the transmitted band, allowing the other half for out-band cut-off -- a so-called "50% efficiency" design. Since the coupling network gain is equal to 10 log R_H/R_C, it follows that the "50% efficiency" design has three db less than the theoretical maximum gain over the transmitted band.

Referring to our amplifier schematic (Figure 13-3) however, we recall that the gain determining high side capacity (C_{12} or C_{52}) had not been specified. Its value is one of the first questions that we have to settle. The smaller the capacity, the more coupling network gain we can get (Eq. 13-1), but the more difficult we make the feedback loop design since the available feedback is related to P_i and P_o (Fig. 13-2), which in turn are functions of C_{12} and C_{52}. As a compromise, we find it a good rule of thumb to specify that C_{12} shall be at least equal to or, even better, two to three times as great as C_1, the capacity from the grid of the first tube to its cathode. For the capacity values illustrated on Figure 13-3, this gives a C_{12} value of somewhere between 16 and 48 mmf. Assuming for the moment that we will use 32 mmf, we can compute the gain area at our disposal for the input coupling network. In view of the 50% efficiency assumption, however, only half of this area is within the band. We may assume for the moment that the half within the band can be distributed at pleasure. A similar gain term may be assumed for the output coupling network, or we might as a refinement in our design decide that the input and output coupling networks will not be identical. This would permit somewhat higher gain in the output coupling network case, since C_5 is in general smaller than C_1, and hence C_{52} could be made smaller than C_{12}.

At this point we should consult the transformer expert to get an estimate of the minimum achievable high side capacity which, incidentally, will be in part determined by the lower frequency cutoff of the transmitted band. This is a consequence of the mutual inductance requirements (hence the construction) of the transformer. The mutual is difficult to control accurately in manufacture, and hence is made large so that the effect of its variations on amplifier gain will be small. If the transformer high side capacity is too great, a more complex coupling network may be used with advantage.

In actual practice we are very limited in our freedom to prescribe the function of gain versus frequency for the coupling networks. We might want to specify, for example, that about one third of the required slope be contributed by each of these networks. In that case one third of the characteristic shown in Figure 13-1 would be allocated to each of the coupling networks, and the other third to the β circuit. Each coupling network would then have the gain-frequency shape shown by the solid curve of Figure 13-5. This is



COUPLING NETWORK CHARACTERISTIC

FIGURE 13-5

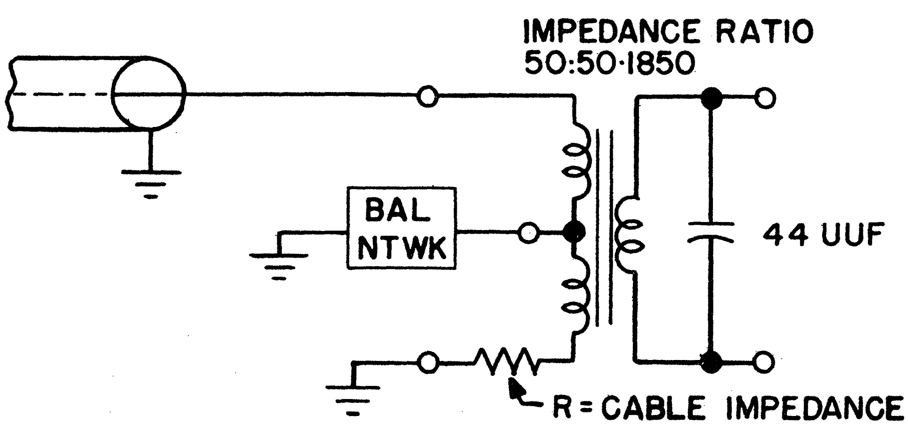
theoretically achievable provided we are willing to invest in a fairly large number of elements in order to shape the insertion gain to this prescribed curve. However, using a large number of elements to meet a prescribed shape in coupling network gain is hardly in the direction of increasing reliability, which makes this approach particularly undesirable in a submarine cable system. It is even less desirable if virtually the same over-all repeater performance can be obtained by using a simpler coupling network. We also know, from experience, that coupling networks giving 15 db slope over the band will be (like maximum gain area networks) quite sensitive to element variations.

If we restrict ourselves to using the leakage inductance of the transformer, the high side capacity, and perhaps two additional elements in shaping the coupling network gain, and ask for 10 db of peaking, we find that we tend to arrive at a gain versus frequency shape more like that shown by the dotted curve of Figure 13-5. Alternatively, it is possible to obtain a fairly flat gain characteristic with the same number of elements. In either case the beta network must supply whatever additional shaping is needed to match the cable loss.

Another point to be considered in connection with the coupling network design is the problem of making the input and output impedance of the amplifier, or one of them, match the cable impedance. This may not be a necessary requirement on the repeater. One of our problems is to decide whether or not it is worth the price. Matching the cable impedance, at one end or both, has the effect of reducing gain deviations caused by reflections or by departures of the cable impedance from the assumed value. These departures result from manufacturing deviations, changes in cable impedance with pressure, temperature, or age, or errors in correctly simulating the cable impedance during the measurement of repeater gain before the insertion of the repeater in the cable and its final laying on the ocean floor. Any one of these factors can cause misalignment, penalizing system equalization and hence signal-to-noise performance. On the other hand, one must pay a price for terminating the cable in its own impedance. The price is a decrease in the amount of coupling network gain that can be built up over a given transmission band, and a degradation of noise figure.

As mentioned in an earlier chapter, there are several methods of achieving an impedance match. A terminating network on the low or high side of the coupling network may be used*, or a hybrid balancing scheme at the repeater input or output. The effect is to reduce the achievable gain by somewhere between 3 and 6 db per terminated coupling network as against the unterminated case. The system performance degradation which this represents is a relatively complex question, depending on the relative magnitudes of modulation noise, tube noise, and cable resistance noise. A complete consideration of the merits of various termination methods is beyond the scope of this discussion.

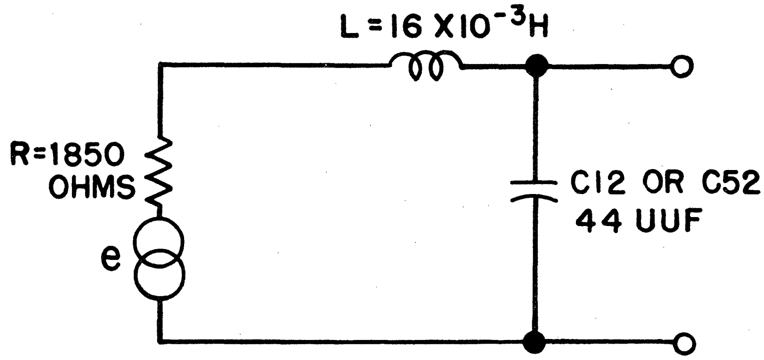
Without attempting to settle the question finally at this time, however, let us assume that we will use some method of termination at each end of the repeater. Instead of the theoretical maximum in-band gain area in the presence of the assumed high side capacity, we would then obtain 6 db less than this gain area for each coupling network. The 6 db gain reduction for each coupling network is made up of 3 db resulting from the 50% efficiency assumption and 3 db for the termination. If we now specify further that a relatively simple network be used and that we want a fair degree of peaking in order to obtain part of the slope equalization required, we might arrive at the coupling network design shown in Figure 13-6 and 13-7.



ELEMENTS OF COUPLING NETWORK

FIGURE 13-6

*The practical use of this method is restricted to coupling networks which are nearly flat across the transmitted band, if we want a good termination at all transmitted frequencies.



R = High side impedance of transformer
L = Leakage inductance + buildout
C_{12} or C_{52} = End capacity

HIGH SIDE EQUIVALENT CIRCUIT

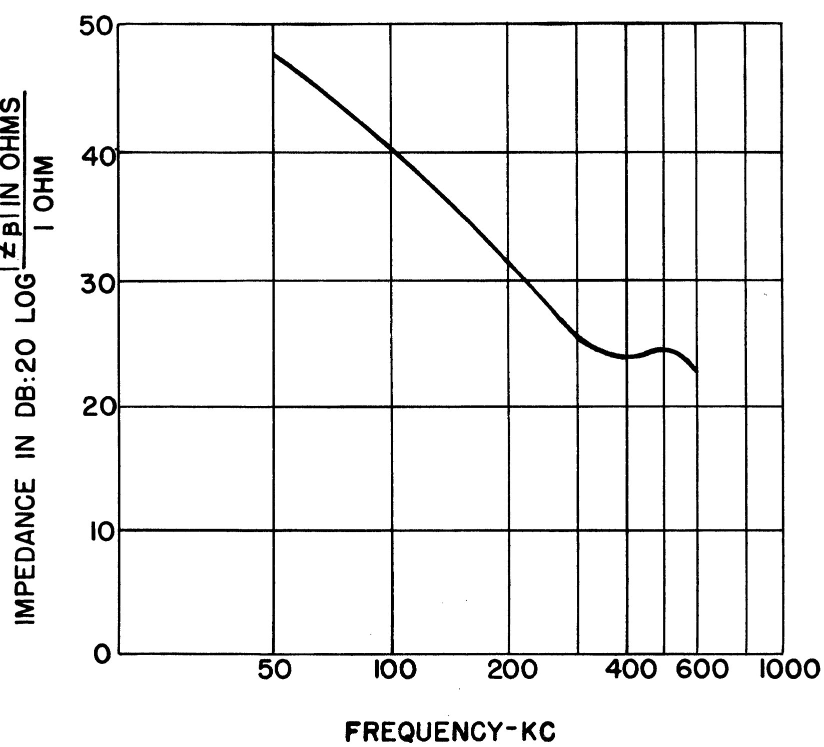
FIGURE 13-7

Here, in an attempt to approximate the desired shape without making potentiometer terms too severe, we have gone from 32 to 44 mmf for C_{12}.

Beta Circuit

Using the coupling networks of Figure 13-6 and knowing the cable loss to be matched, we can deduce the combined value of the equalizer and beta circuit components of the amplifier. Since an equalizer represents dead loss in the circuit it is preferable to incorporate as much of the remaining gain shaping as possible in the beta circuit. Setting the equalizer loss to zero for the moment, we arrive at the in-band beta circuit characteristic shown in Figure 13-8 by using Equation 13-3, expressing the coupling network gains, the cable loss and the cable impedance in db. We are asking for a slope of about 25 db over the band, shaped in a particular fashion. This characteristic requires the impedance of the beta circuit to vary from about 240 ohms at low frequencies to about 13 ohms at the top transmitted frequency. This appears to be an achievable beta circuit impedance characteristic for the bandwidth in question. There are several tests that we can apply. One of them is to compare the impedance of the beta circuit with a theoretically achievable two-terminal impedance as limited by the parasitic capacity across the beta circuit. In order to do this we use the Φ-scale plot.

Another criterion is the minimum or maximum value of impedance called for from the standpoint of element value considerations and associated parasitics. Thus, for example, if we were asking for a two-terminal network having an impedance of 0.5 ohm at the top transmitted frequency we might question this design because such a low beta circuit impedance would be sensitive to parasitics (e.g. lead inductance or ground plane resistance.) Similarly even if the Φ-scale criterion were not violated, we might be skeptical of a beta circuit design which called for extremely high impedance at some portion of the band - say 20,000 ohms at very low frequencies. The element values which would be involved, and the d-c considerations, would probably be intolerably difficult to satisfy. Later detailed circuit design may lead to the conclusion that the exact shape specified by Figure 13-8 would be too difficult to attain. In that case we can always decide to place a passive equalizer ahead of the repeater to help in shaping the gain characteristic. It will frequently turn out to be easier to meet the gain requirements with simple circuits if we do in fact employ such an equalizer, but it remains desirable to keep the loss of the equalizer to a minimum.

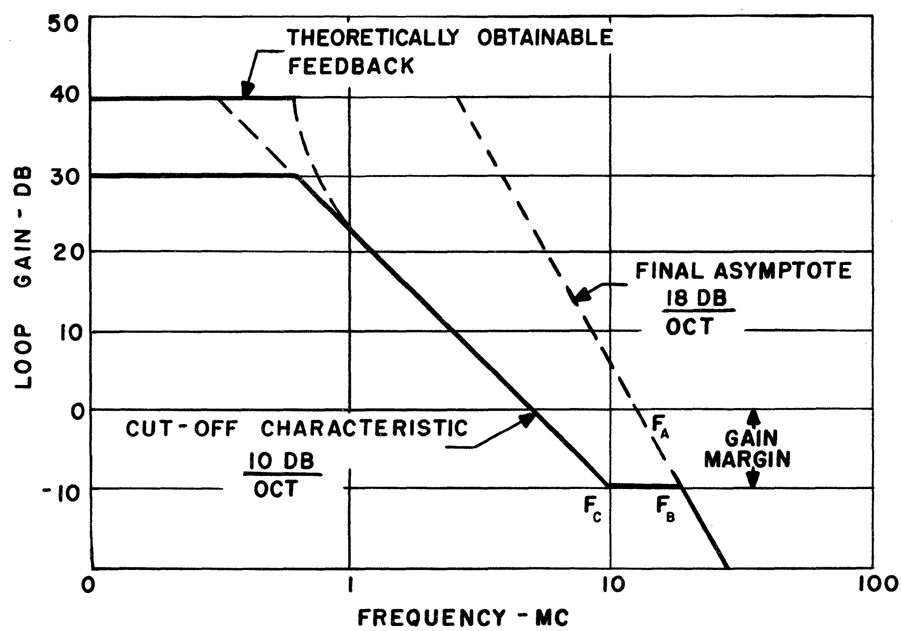


BETA CIRCUIT CHARACTERISTIC

FIGURE 13-8

Feedback Loop Design

Since we began with the assumption that we were going to use a rather simple feedback amplifier configuration, the limitations on obtainable feedback are fairly well known and rather simply stated. In the first place the Nyquist stability criterion must be satisfied. A graphical method for computing how much feedback can be obtained over a given band without violating this criterion is illustrated in Figure 13-9. In arriving at this sort of plot, however, we are again forced to make a number of assumptions, as usual to be re-examined later in the light of system considerations. Among the more obvious are the stability margins which we wish to incorporate into the design.



LOOP GAIN CHARACTERISTIC

FIGURE 13-9

Usually it is considered reasonable to choose about 7 to 10 db of gain margin and about 20° to 30° of phase margin. Since we are in the very rough initial stages of our design problems, it is wise to be somewhat conservative in choosing margins. Another assumption which affects the computation of obtainable feedback is our choice of coupling network high side capacity. We have made the assumption that the coupling network capacity will be somewhere between 2 and 3 times the adjacent vacuum tube capacities. This approximately sets the so-called potentiometer term loss in the asymptotic cut-off region. Another problem is that of excess phase due to transit time. At first sight it would appear that for the frequency range which we are considering this might be ignored. For low voltage tubes with large plate-cathode spacing, however, excess phase may not be negligible in the asymptotic cut-off of even a relatively low-frequency amplifier. In very wide band designs or very high-frequency band-pass designs excess phase will always be a limiting factor in feedback loop design.

Using the preliminary capacity estimates of Figure 13-3 and the formula for T of Figure 13-2, we can determine the frequency at which unity gain around the loop occurs. In our case this is approximately 12.5 megacycles. We know from the form of the equation and can see from the circuit configuration that this gain around the loop will decrease 18 db per octave in the asymptotic region. We can therefore draw the asymptote line passing through 0 db loop gain at 12.5 megacycles (f_a in Figure 13-9). Following the design technique outlined by Bode, the loop gain characteristic joins the 18 db/octave asymptote at the frequency where the asymptotic loop gain equals the assumed gain margin. Here a gain margin of 10 db has been chosen, and hence f_b in Figure 13-9 falls at 18.4 megacycles. We also selected a phase margin of 30°. A 30° phase margin is associated with a slope in the cut-off region of 10 db per octave. If we ignore excess phase for the present, this fixes f_c in Figure 13-9 since the ratio of f_b to f_c is equal to the ratio of the final asymptotic slope to the assumed slope in the cut-off region. Hence

$$f_c = \frac{10}{18} \cdot f_b = \frac{10}{18} \cdot 18.4 = 10.2 \text{ mc.}$$

We can draw a line having a slope of 10 db per octave through f_c. At 600 KC this line intersects with the 30.3 db gain line. If we are planning to use flat feedback across the transmitted band the Nyquist criterion tells us that we can have 10 db more flat feedback than the above value (see Fig. 13-9)*.

Whether or not we can obtain this 40.3 db of flat feedback (assuming that we want flat feedback) is still a question. It will be recalled that the in-band beta circuit impedance has already been

---

*If we had decided on a 15° phase margin, the cut-off slope would have been 11 db per octave, the 10.2 megacycle frequency would have shifted up to be 11.2 megacycles (11/18 of 18.4 megacycles instead of 10/18) and we could theoretically obtain flat feedback 11 db greater than the intersection of the 11 db/oct slope and the 600 kc ordinate.

specified. If in order to obtain the desired insertion gain it was necessary to set the beta circuit impedance at a very low value, we may find that we do not have sufficient μ-circuit gain area to achieve the theoretically possible 40.3 db of feedback - theoretically possible, that is, on the basis of asymptotic considerations alone. Deciding whether or not we are gain-limited would be easy if the potentiometer terms and the beta circuit impedance and interstages were flat with frequency. Since none of these is true, however, it is generally necessary to carry through a rough design of the interstages at this point to determine whether or not we are exhausting our fund of gain area. To do this we must compute the potentiometer terms and select interstage designs which roughly compensate for the potentiometer term and beta circuit shapes in band. At this stage a sufficiently good approximation can be found by selecting from some of Bode's interstage curves* a couple whose combination matches moderately well the combined effect of potentiometer terms and beta circuit to give a flat over-all feedback characteristic. If we find that we are close to being gain limited, we can expect difficulty if we attempt to achieve the theoretically obtainable maximum feedback.

This sort of design approach gives us a pretty good idea of the flat feedback that can be obtained over a given transmitted band and assures us that we are not violating the Nyquist criterion or the gain area limitations. It frequently turns out, however, from a study of the system modulation properties and requirements, or from a study of the gain deviations of the amplifier as tubes age, that flat feedback across the transmitted band is not the optimum solution. Instead it may be desirable to have more feedback at low frequencies, allowing the feedback to fall off at the upper edge of the transmitted band. We can use the flat feedback value found to be practicable as a guide in determining how much shaped feedback we can expect to achieve. For this purpose the Φ-scale referred to above is useful.

It should be emphasized again at this point that we are carrying through a very rough design to get an idea of the size of the quantities of interest. Before we freeze on any final repeater design for the system, we should consider other circuit configurations, and even assuming that we decide to use the series feedback amplifier circuit, we have left many questions unanswered. For example, the outband design

---

*op. cit. Chapter XVIII.

of the beta circuit and the manner in which this joins the in-band values tentatively determined from gain objectives is a question to exercise the art of the network and amplifier designers. This is because most of the burden of shaping the out-band cutoff characteristic of the feedback loop falls on the beta circuit, since the potentiometer terms and interstages tend to be so restricted by in-band and band-edge considerations that they are practically unavailable as design parameters in the cutoff region. If we are not gain-limited, some cutoff shaping can be incorporated in the interstages, but usually not much can be obtained without sacrifice of in-band feedback.

A number of new techniques of network design have been developed in recent years; these permit us to arrive at economical exact network designs for limited problems, but are usually not applicable to beta circuit design in the earliest stages of the problem. We have many freedoms in our choices of exact coupling network values, and in our specification of passive equalizer characteristic. A network design may turn out to be extremely difficult under one set of assumptions and easy under other assumptions where the difference between the two sets of assumptions is almost immaterial from the system standpoint. A few db more or less shaping in the beta circuit or a smoother transition from in-band to out-band values achieved by using a slightly different equalizer may make a vast difference to the beta circuit network designer. A certain amount of art is therefore necessarily called for.

All this discussion has been predicated on the assumption that we are using the simple series feedback amplifier circuit. If we go to more complex amplifier configurations, other limitations on obtainable feedback begin to enter or, on the other hand, ways may be found to achieve slightly more feedback than the Nyquist criterion for single loop circuits would permit. A full consideration of these topics would be beyond the scope of this section.

Performance Estimates

Once the preliminary repeater design has been roughed out, it is relatively simple to make rough estimates of the quantities, listed at the beginning of this section, which the system designer needs.*

---

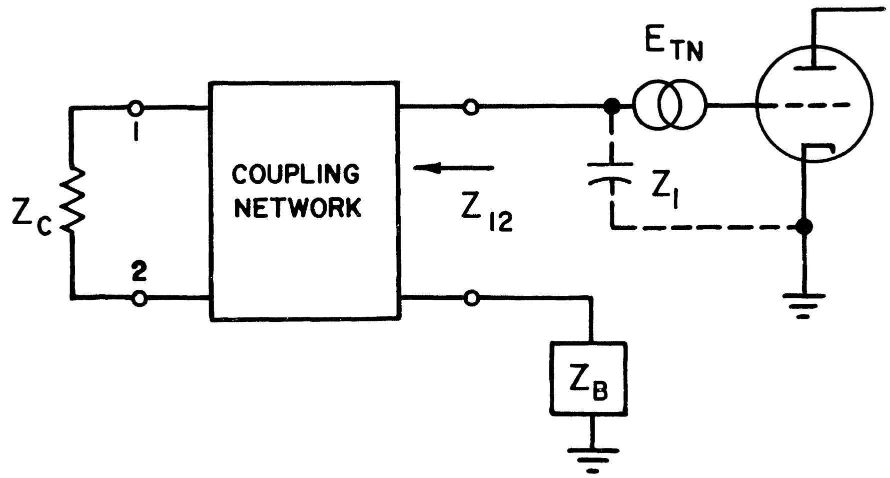
*It should be emphasized, however, that these estimates are of necessity only approximate, and must be checked against laboratory measurements as early in the system design as possible.

We have found already, for example, the factors determining what gain-bandwidth we can obtain, and the corresponding feedback. For reasonable variations in frequency or in repeater spacing we will be able to estimate the corresponding variations in repeater parameters. Thus we can anticipate that for a given repeater structure, the available feedback will decrease by 10 db for each additional octave of top transmitted frequency (Fig. 13-9). At the same time the coupling network high-side impedances will be halved, decreasing the gain of each coupling network by 3 db (see page 13-9), with corresponding changes in thermal noise, load carrying capacity, and modulation for a given output power from the repeater.

It might be pointed out that although the feedback is a function of bandwidth, it is not necessarily a function of repeater spacing. When we are not gain-limited, the tie between available gain area and the sum of insertion gain and feedback is a loose one. In designing feedback amplifiers for other applications (e.g., a home audio system) one might, as a matter of course, design for all the feedback possible (as limited by Nyquist's criterion) and use the rest of the available gain area for forward gain. In such a case, if we were later asked to provide more forward gain, we would have to decrease the feedback. Very often we do not have this situation in system design work. Instead we may be considering an amplifier whose insertion gain, determined from the repeater spacing as in Chapter 10 is less than the achievable maximum although the feedback is as great as the Nyquist criterion permits. In other words, we could increase the beta circuit loss, thus increasing the insertion gain, and still get the same feedback as before by raising the interstage impedances. We can often, therefore, examine the effect on system performance of a longer repeater spacing without at the same time assuming a smaller value of feedback in the equations of Chapter 10 Eventually in such a process we would reach a spacing for which we were gain limited. From then on, any further increase in spacing (i.e., in forward gain required) would, of course, necessitate a reduction in feedback.

<u>Noise</u>

To find the noise in dbm delivered to the load impedance at amplifier output, we are usually justified in assuming that the first stage circuits are the only significant noise contributors. The contribution of the first tube is conveniently expressed in terms of an rms voltage generator in series with the grid, E_{TN} in Figure 13-10. The value of this generator can be computed for a 3 kc band by formulae given in Chapter 7, repeated here in a form convenient for our purposes.



NOISE CONTRIBUTORS

FIGURE 13-10

$$E_{TN} = 7.03\ 10^{-9} \sqrt{R_{TN}} \text{ volts,} \qquad (13\text{-}4)$$

where

$$R_{TN} = \frac{I_p}{I_p + I_s}\left(\frac{2.5}{G_m} + \frac{20\ I_s}{G_m{}^2}\right) \text{ ohms} \qquad (13\text{-}5)$$

and

I_p = plate current in amperes

I_s = screen current in amperes

G_m = transconductance in mhos.

The tube noise component of the total amplifier noise could equally well be produced by a generator in series with Z_{12}. To a very close approximation, this equivalent generator would be

$$E'_{TN} = E_{TN}\ \frac{Z_1 + Z_{12} + Z_B}{Z_1} \doteq E_{TN}\ \frac{Z_1 + Z_{12}}{Z_1} = \frac{E_{TN}}{P_i} \qquad (13\text{-}6)$$

In other words, the "tube noise voltage" may be added on a power basis to the noise of Z_{12} if it is first multiplied by the reciprocal of the input potentiometer term.

In addition to tube noise, we will also have important contributions to total output noise from the resistive component (R_{12}) of the high side impedance of the input coupling network, Z_{12}. Since we may have transformer and termination losses, this Z_{12} noise is not necessarily equal to the noise from the resistive component of the input cable impedance Z_c stepped up by θ_i.

Defining the total effective rms noise voltage in series with Z_{12} as E_{12N}, we can find the resulting noise voltage across the output load impedance as

$$E_{OUT} = E_{12N}\, \theta_o \frac{Z_c}{Z_\beta} \left(\frac{T}{1+T}\right) \doteq E_{12N}\, \theta_o \frac{Z_c}{Z_\beta} \qquad (13\text{-}7)$$

where T is the return ratio around the feedback loop, so that $\frac{T}{1+T}$ is nearly unity. Or we may, if we wish, transfer E_{12N} to be a voltage in series with Z_c by dividing it by θ_i. Usually the beta circuit resistance is a negligible noise contributor. When it must be taken into account, it can be added directly to the resistive component of Z_{12} with negligible error.

Modulation Coefficients and K_F

Experience with a number of tube types justifies the use of the following assumptions in making first-order estimates of M_{2R} and M_{3R}. While only approximations, these assumptions seem to hold fairly well for pentodes, tetrodes, and for triodes working into low load impedances.

1. The tube modulation will be such that for a plate current swing whose peak excursion is 10% of the d-c plate current, the second harmonic will be 40, the third harmonic 70 db below the fundamental.

2. The tube modulation will be reduced by feedback. Second order modulation (e.g. second harmonic, A-B products, etc.) will be reduced by the full amount of the feedback F.* Third order modulation, however, will be produced in a feedback amplifier by virtue of the combining of fundamentals and fed-back second order modulation even if the third order modulation coefficient of the tubes themselves is zero. The effective reduction of third order modulation is less, therefore, than the full amount of the feedback. This effect is taken into account approximately by the factor K_F which appears in the equation

$$M_{3R} = M_3 - F + K_F \,.$$

An estimate of K_F can be obtained from the equation

$$K_F = 20 \log \left(1+2\,\frac{m_2^{\,2}}{m_3}\right)$$

where m_2 and m_3 are the numerical ratios corresponding to M_2 and M_3.

---

*F = 1 + T, thus, if T is a factor of 10 (20 db), with a phase angle of zero degrees, the feedback reduction in db will be 20 log (1 + 10) or 20.8 db.

Load Carrying Capacity

As discussed in Chapter 9, the load carrying capacity of an amplifier is difficult to define with any exactness. A very rough estimate of load carrying capacity, which is more valid for comparison of various circuit choices than for any absolute use, is to assume that the amplifier can deliver to the load the peak power equal to the power which the d-c plate current would deliver to the resistive component of the a-c impedance of the coupling network, minus any losses due to terminations, etc., in the output coupling network. Usually the system designer will limit the drive on the tubes so that modulation exceeding some given amount will not be obtained; this implies a knowledge of how the modulation will depart from the power series law as the drive is increased. Estimates of this effect are difficult to make.

Manufacturing and Aging Deviations

Estimates of the shape and magnitude of the gain changes which will be caused by manufacturing and aging deviations are not difficult to arrive at when we know definitely the values of circuit elements, although the computations are tedious. Such estimates are, of course, subject to constant refinement as the design and the manufacturing process become better known. For preliminary figures on manufacturing deviations, it is generally sufficient to compute the changes in θ_i (and θ_o) caused by variations in leakage inductance, mutual inductance, and high side capacity of the transformer. For estimates of probable transformer variations, which are a function of the physical design, we must rely on the transformer design group. The dominant cause of aging deviations in the repeater will be tube aging; the computation of the change in feedback amplifier gain associated with a change in vacuum tube transconductance is straightforward. In the final stages of design and system analysis, the aging effects in passive elements and cables must also be considered.

Realization and Quality Control

The final crystallization of any particular repeater design, and the realization of this design in the form of a manufactured product, is a process of increasing concentration on detail. Where the importance of the application justifies such a process, it is desirable to carry out a thorough program of cross-checking computations and laboratory measurements. (For multi-loop circuits or off-ground coupling networks, this can be quite laborious). From such a program we can gain the complete understanding of the circuit and its causes of variation which is almost essential to the engineering of the final manufacturing process. A statistical approach is essential here - the use of "designed experiments", the analysis of variance technique, and the avoidance of drawing conclusions from single models can greatly expedite the program.

In the final stage of manufacture and use, the employment of quality control procedures can result in a better and cheaper product - may, in fact, make possible the realization of system objectives which would have been unrealizable without these tools.

Bibliography

1 - W. T. Duerdoth, Some Considerations in the Design of Negative-Feedback Amplifiers, Proceedings of the I.E.E. (Brit.) Part 3, vol. 97, pp. 138-158, May 1950.

ILLUSTRATIVE EXAMPLES - CHAPTER 13

In the following, assume the transmitted band is from almost d.c. to 1 mc.

1. a) Assuming that the high side capacity of a coupling network is 45 UUF, compute the maximum value of high side resistance in the transmitted band, (100% efficiency) assuming gain is to be flat vs frequency.

   b) If the coupling network is to be only 50% efficient and be terminated as in Figure 13-6, what would θ_i be?

   c) If the coupling network of b) above is to have a gain (θ_i or θ_o) shaped vs frequency (relative to the 1 mc gain) as shown in the following table, what (approximately) will be the gain at 1 mc?

f, kc	θ, db	f, kc	θ, db
up to 100	-10.0	600	-5.25
200	- 8.5	700	-3.2
300	- 7.5	800	-1.9
400	- 6.6	900	-0.85
500	- 5.85	1000	0

Solution:

a) If the high side capacity seen looking into a coupling network is 45 UUF, then the resistive component, flat over a band extending from dc to 1 mc is given by the resistance integral theorem as:

$$\int_0^\infty R \, d\omega = \frac{\pi}{2C}$$

which reduces to

$$R \cdot 2\pi \, 10^6 = \frac{\pi}{2 \; 45 \; 10^{-12}}$$

$$R = 5560 \text{ ohms}$$

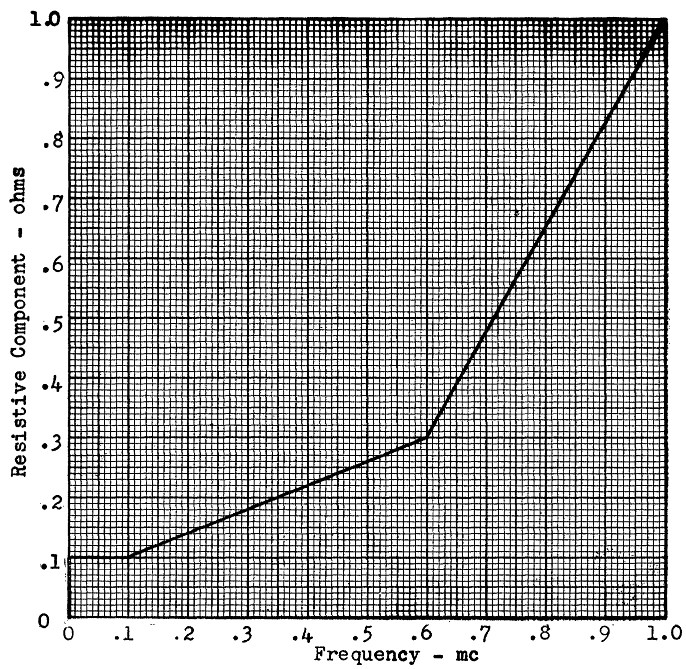
b) If only 50% of the resistance-frequency area is to be in the 1 mc band, and if hybrid termination is to be used, then the gain of the coupling network will be

$$\theta_i = 10 \log \frac{5560}{R_c} - 6 \text{ db}$$

$$= (10 \log \frac{1390}{R_c}) \text{ db}$$

where R_c is the resistive component of the generator or cable impedance.

c) Replotting the assumed curve of relative gain vs frequency, we obtain the plot shown below, assuming the resistive component of the high side impedance at 1 mc to be unity. The area under the



curve is approximately 0.41 megacycle-ohms. The top frequency gain, corresponding to one ohm, is therefore 3.87 db higher than the gain which could be obtained from a flat coupling network having the same high side resistance area, since

$$10 \log \frac{1}{.41} = 3.87 \text{ db}$$

The gain which could be obtained at 1 mc from the coupling network of part b, shaped as prescribed, would be

$$\theta_i = 10 \log \frac{1390}{R_c} + 3.87 \text{ db.}$$

$$= 10 \log \frac{3390}{R_c} \text{ db}$$

Illustrative Example #2

For a submarine cable repeater which carries 250 channels (0.2 to 1.2 mc) on a 2000 mile circuit, estimate the following top channel performance figures:

$$N_R \; M_2 \; M_3 \; e_p \; F \; K_F \; Q$$

Assume a 3 stage series feedback amplifier.

The following information and assumptions are given:

Feedback: Flat over the transmission band. Sing margins for new tubes shall be

$$A_m = 10 \text{ db} \quad \varphi_m = 30°$$

In finding F from T, assume angle of T is zero degrees.

Tube Parameters:

g_m = 12,200 micromhos	E_{SG} = 50 v.
I_b = 12 ma	E_K = 1.25 v.
I_s = 4 ma	C_{IN} = 31 mmf (hot)
E_{bb} = 50 v.	C_{OUT} = 10 mmf

Circuit Capacities (excluding tubes)

Interstage = 28 mmf
At input grid = 8 mmf
At output plate = 8 mmf
Transformer end capacity = 45 mmf
Beta Circuit = 60 mmf.

Coupling Networks

Input and output networks are identical, with "hybrid" terminations. Real part of high-side impedance (top channel) = 3350 ohms. Absolute value of high-side impedance (top channel) = 3450 ohms. Cable impedance = 44 ohms.

Top Channel Gain

$$20 \log \theta_i = 20 \log \theta_o = 15.3 \text{ db}$$

Potentiometer Terms (top channel)

$$20 \log P_i = -2.2 \text{ db}$$

$$20 \log P_o = -0.5 \text{ db}$$

Additional Assumptions

1) Beta circuit noise is negligible in the top channel.
2) The only important source of modulation is the output tube.

Solution:

N_R

To find noise figure:
Tube noise - equivalent resistor

$$R_{TN} = \frac{I_p}{I_p+I_s}\left[\frac{2.5}{G_m} + \frac{20\ I_s}{G_m^2}\right]$$

$$= \frac{12}{16}\left[\frac{2.5}{12.2\ 10^{-3}} + \frac{20\text{x}4\ 10^{-3}}{149\ 10^{-6}}\right]$$

$$= 557 \text{ ohms}$$

The noise voltage in series with Z_{12} which would produce the same noise at repeater output would have to be 2.2 db greater than the noise voltage associated with this 557 ohm resistance. Since $E_N = k\sqrt{R}$, we can regard the equivalent generator in series with Z_{12} (to simulate tube noise only) as a fictitious 924 ohm resistor. (Since $20 \log \frac{k\sqrt{924}}{k\sqrt{557}} = 2.2$ db).

High side resistive component of input coupling network is given as 3350 ohms. Total effective noise voltage in series with Z_{12} is therefore that corresponding to a resistance of 924+3350=4274 ohms.

If input generator (44 ohms) were only source of noise, noise voltage in series with high side of coupling network would be 44 ohm noise voltage increased 15.3 db, since θ_i is given as 15.3 db. This would correspond to the noise voltage associated with a resistor of 33.9x44 ohms or 1492 ohms.

Comparing real and ideal high-side noise voltages we find the real voltage to be greater than the ideal by

$$20 \log \frac{k\sqrt{4274}}{k\sqrt{1492}} = 10 \log \frac{4274}{1492} = 4.6 \text{ db}$$

This is the noise figure for the top channel. N_R is therefore $-139 + 4.6 = -134.4$ dbm.

$\underline{M_2, M_3}$

Assume tube modulation, in terms of equivalent rms voltage at grid to produce observed plate current components for a ± 10% swing of quiescent plate current, to be

$$\frac{e_{2f}}{e_f} = -40 \text{ db} \quad \frac{e_{3f}}{e_f} = -70 \text{ db.}$$

The peak grid voltage E_g required to swing plate current 10% can be found

$$G_m E_g = 0.1 \, I_p$$

$$E_g = \frac{12 \; 10^{-3}}{10 \; 12.2 \; 10^{-3}} = .098 \text{ V peak}$$

This is .069 volts rms, or 23.2 db below one volt rms. The 40 and 70 db ratios given above thus correspond to a fundamental grid drive of -23.2 dbv rms. If the drive were increased to one volt rms these ratios would become

$$M_2 = -40 + 23.2 = -16.8 \text{ db}$$

$$M_3 = -70 + 46.4 = -23.6 \text{ db}$$

$\underline{E_p}$ (maximum permissible grid swing in rms volts as defined in Chapter 3)

Max. plate swing:

$$|I_b \, Z_{52}| = (12 \times 10^{-3}) \; (3450)$$

$$= 41.4 \text{ volts peak.}$$

This is less than 50 V, therefore the repeater is current rather than voltage-limited. Max. grid swing:

$$E_p = \frac{I_b}{g_m} = \frac{12 \times 10^{-3}}{12.2 \times 10^{-3}} = 0.98 \text{ volts peak}$$

$$= 0.692 \text{ volts rms}$$

$$= \underline{\underline{-3.2 \text{ dbv}}} \text{ rms.}$$

$\underline{F}$ (maximum obtainable feedback).

To find F:-

$$T = \frac{P_i P_o G_{m_1} G_{m_2} G_{m_3}}{Y_3 Y_4 Y_2'}$$

let $f = f_a$(frequency at which T = 1).

Solving this equation for f_a:

$$f_a = \frac{1}{2\pi}\sqrt[3]{\frac{P_i G_{m_1} G_{m_2} G_{m_3} P_o}{C_3 C_4 C_2'}}$$

Asymptotic P-terms:

$$P_i = \frac{C_{12}}{C_1+C_{12}} = \frac{45}{(31+8)+45} = 0.535$$

$$P_o = \frac{C_{52}}{C_5+C_{52}} = \frac{45}{(10+8)+45} = 0.714$$

See Figure 13-3 of text:

$$C_3 = C_4 = 10 + 28 + 31 = 69 \text{ mmf}$$

$$C_2' = C_2 + P_i C_1 + P_o C_5$$

$$= 60 + (0.535)(8+31) + (0.714)(8+10)$$

$$= 93.5 \text{ mmf.}$$

$$f_a = \frac{12.2\text{x}10^{-3}}{2\pi}\sqrt[3]{\frac{(.535)(.714)}{(69)(69)(93.5)(10^{-36})}}$$

$$= 18.6 \text{ mc.}$$

Whence, from a plot like that of Figure 13-9,

$$f_b = 27.5$$

$$f_c = \frac{10}{18} \times 27.5 = 15.4$$

T = 37 db if the angle of the feedback were 180°, i.e., T = 70.7 $\underline{/0°}$, then F would be 71.7 $\underline{/0°}$ or about 37.1 db.

$\underline{K_F}$

$$M_2 = -16.8 \text{ gives } m_2 = \frac{1}{6.9} \qquad (13\text{-}8)$$

$$M_3 = -23.6 \text{ gives } m_3 = \frac{1}{15.2}$$

$$K_F = 20 \log\left(1+2\frac{15.2}{(6.9)^2}\right) = 4.3 \text{ db}$$

Q:

For e_g = 1 volt rms on the final grid, how many dbm are delivered to the line?

$$i_p = g_m e_g = 12.2 \text{ ma rms}$$

(Since Q is simply a translation relating grid volts to line dbm, we need not be concerned with "overload" in computing it).

High side load current $i_H = i_p P_o = \frac{12.2 \text{ ma}}{1.06} = 11.5$ ma rms.

(Here top channel value of P_o, rather than asymptotic ratio of capacities, must be used.)

Low side load current $i_L = i_H \theta_o = (5.83)\ (11.5) = 67.0$ ma rms.

$$P_L = {i_L}^2 R_c = (67.0 \times 10^{-3})^2 \quad (44)$$

$$= 198 \text{ mw.}$$

$$Q = 10 \log P_L \quad (e_g = 1 \text{ v. rms})$$

$$= 23.0 \text{ db.}$$

Chapter 14

EQUALIZATION AND REGULATION

Introduction

If a system is to transmit telephone, television or special service signals satisfactorily, its loss-frequency characteristic must be flat within narrow limits. Some signals also require a linear phase-frequency characteristic. The problem of achieving and maintaining the required characteristics is the equalization problem. "Regulation" refers to that part of equalization which corrects for relatively rapid changes in transmission by automatic means. As in any system problem, we have here an interaction of requirements, devices, and performance. Just as we cannot consider transmission systems apart from the environment in which they operate - the whole Bell System plant, from telephone subsets to switching plans - so we cannot discuss "equalization systems" without reference to their environments: the transmission systems of which they form parts, and for whose imperfections they must compensate.

Unfortunately, the equalization problem does not lend itself to a definitive treatment like that given the signal-to-interference problem in earlier chapters. In the discussion which follows, therefore, we can be specific only by referring to problems encountered in particular system designs, and thus illustrating the general principles involved. The L-3 coaxial system will be used as a source of such examples.

We shall consider first the objectives and requirements which affect the design of equalization systems. Next we shall discuss the more important sources of transmission deviations, and their nature. Finally we shall describe some of the methods and devices which form the parts of an equalization plan - such as reducing deviations at the source, complex manually-adjusted equalizers, and pilot-controlled automatic equalizers with their associated regulators.

Requirements

General

Equalization can be regarded as a network problem or as a system problem. As a network problem, the transmission objectives for the system and the system deviations from these objectives are translated into requirements on equalizer networks. The problem then is to design networks that economically meet those requirements.

As a system problem equalization must be related to a much more diverse set of conditions. The distribution of equalizers along the transmission line, the relation between that distribution and other factors such as power, maintenance switching, population density, and signal-to-interference performance are typical of these diverse conditions. Equally diverse requirements are imposed on the equalization system by these environmental effects.

Let us consider first the transmission requirements imposed on the system. These will differ, depending on the types of signals to be transmitted. Telephone, monochrome and color television, telegraph, telephotograph, program, and data signals - each has peculiarities which require equalization treatment in someway different from the others. A full discussion of these requirements would be outside the scope of this text; we shall confine ourselves to a few examples to illustrate the dimensions involved.

Telephone Requirements

In setting up a broad-band system of given length, a certain value of nominal net loss (from transmitting to receiving toll switchboard) is assigned, the value being determined by such factors as the necessity of controlling echoes. The magnitude of the equalization problem for the telephone transmission can best be appreciated by noting that the operating companies specify that corrective action shall be taken whenever:

a) The average net loss of a moderately large group of circuits (say 100) departs from the nominal value by more than 0.25 db.
b) The statistical distribution of net losses of such a group of circuits exhibits a "standard deviation" (sigma) greater than 1.0 db.
c) The net loss of any one circuit departs from the assigned nominal value by more than 1.5 db.*

Of these requirements, the first is the most stringent in its effect on the equalization of broad-band systems. In a system like the L-3 coaxial, for example, carrying 1800 channels, any transmission deviation is likely to affect several hundred channels, so that if the gain over part of the band changes by one db per week, corrective action would be called for every other night. And this sort of gain stability is demanded of a system which, for transmission from New York to Chicago, requires about 250 amplifiers in tandem compensating for a total mid-band cable loss of the order of 7500 db.

---

*Actually these requirements (a, b, and c) vary somewhat depending on circuit length.

Television Requirements

Television transmission requirements are discussed in more detail in the next chapter. For 4000 mile systems, they are of the order of $\pm$ 0.25 db in gain, and $\pm$ 0.1 microsecond in delay distortion. In the case of telephone, the stability versus time is the most severe factor; the transmission across the entire broad carrier frequency spectrum is not necessarily required to be flat, provided it stays put. This follows from the fact that the adjustable group and super-group equipment in the terminals fragmentizes the spectrum anyway. For television, stability is still an extremely stringent requirement, but now in addition we must consider initial equalization to obtain a characteristic which is gain and delay equalized over a 4 mc spectrum.

Other Requirements

In addition to the fundamental transmission requirements mentioned above, a number of other requirements are imposed on the equalization system - some obvious, others perhaps unexpected. Some of these are listed below with brief comments.

Flexibility: An earlier chapter discussed the way in which circuits (telephone or TV) are set up and taken down on a minute-to-minute or hour-to-hour basis in accordance with flexible nation-wide switching plans. It follows that equalization must be done on a link basis in such a fashion that satisfactory long-circuit transmission can be obtained by adding together various links in any possible pattern. To plan to equalize to perfection using one set of equalizers at the end of a 4000 mile circuit is to miss the point entirely.

Adjustment: The most advanced design of equalizer is useless in practice if it cannot be accurately and quickly adjusted by field personnel of limited training. As a general principle, no more information should be gathered from the system than is necessary to provide sufficiently accurate equalizer control setting information for the equalizer operator. Since the shapes (gain and delay characteristics) of the equalizers must often be complex functions of frequency, "ease of adjustment" may be difficult to achieve.

Signal-to-noise: The effects of misalignment on signal-to-noise and overload performance were discussed in an earlier chapter. These, along with considerations of flexibility and maintenance, determine the location of equalizers along the line. Deciding how much misalignment penalty is tolerable versus equalizer costs and equalizer signal-to-noise penalties is a question calling for nice judgments.

Maintenance and Protection: In order to permit replacement of defective equipment (e.g., tubes which, although they have not failed, have aged beyond limits) or to protect service in the event of a sudden failure, it is necessary to switch service to a spare line or channel. If circuits are to be switched between "working" and "standby" lines without causing service reactions, the transmission characteristics of the two lines must be very nearly alike. Since the only practical way to make them alike is to make them both flat, a rather formidable array of equalizers is needed for each line in a switching section. There may be a number of switching sections in a "link", so this requirement, although similar to, is not identical with the flexibility requirement mentioned earlier.

Special Requirements: The foregoing are rather general requirements. Others, of more specialized nature, may arise in particular systems. For example: in the L-3 coaxial system, which transmits telephone and television signals simultaneously, a particular third order modulation product tends to limit system signal-to-interference performance. The requirements imposed on this type of product were met by requiring that the telephone and television signals be split apart at the end of every maintenance switching section and that delay equalizers for the television signal be located in the television branch at every such point. By imposing this requirement on the equalization system, sufficient phase shift between fundamentals is introduced to result in random addition of the modulation products arising in successive switching sections.

Summary

To recapitulate, then, the requirements on the equalization system include:

1. Initial line-up to desired transmission accuracy.
2. Holding the transmission within limits between reasonably infrequent manual readjustments.
3. Flexibility, to meet nationwide switching plans (telephone or TV).
4. Ease and speed of adjustment.
5. Reducing signal-to-noise penalties due to misalignment - without introducing too much noise from equalizer loss itself.

6. Permitting switching to spare lines for maintenance and service protection.
7. Reasonable equipment costs - both for equalizers and for adjustment test sets.
8. Geographical accessibility for personnel.
9. Minimizing special patching procedures and the need for liason over long sections of line.
10. Special features, such as breaking up systematic addition of particular third order modulation products.

Design Philosophies

A number of general ideas about equalization system design can be stated as worthy of consideration in any new design effort. The first of these is the idea of "cause associated shapes". In most systems there are two phenomena which cause variable transmission deviations - temperature and aging. These cause-associated characteristics tend to be relatively simple functions of frequency. They can usually be predicted, to a fair degree of accuracy, from computations and laboratory measurements.

When this is so, it is possible tc design variable equalizers of complementary shape in advance of the first field installation of the new system. Good accuracy can be obtained, especially when the characteristics are simple. Such variable equalizers can be adjusted by varying a single element (e.g., a thermistor) controlled either by the original cause of variation (e.g., temperature) or by a pilot frequency placed at a sensitive point in the spectrum. By this approach, initial system operation is improved, misalignment reduced, and range of adjustment in the complex manually-adjusted equalizers is conserved.

A second idea that is useful is that of providing manual mop-up that is extremely flexible. Unexpected effects tend to plague the operation of systems - flexible mop-ups with adequate range prove to be good insurance, especially during early prove-in years of operation while "bugs" are being worked out of the system and personnel is being trained to operate it efficiently.

Recognition of the difficulties of providing adequate equalization facilities under present methods of operating transmission systems leads to the thought of one day designing a fully automatic system of equalizer adjustment. To attain such a goal will require the development of new measuring techniques and the application of information theory, computer design, and servo control of equalizers. Some general thinking has been directed at these possibilities, but no development work for specific system use has been initiated.

Transmission Deviations

If we list the sources of transmission deviations in such a system as the L-3 coaxial, we obtain Table 14-1, which gives some of the important characteristics of each deviation. The descriptive terms used are relative, of course. A deviation is "fast" if it calls for corrective action as often as once a week in order to keep fairly long circuits within limits - but it is difficult to say just where the line should be drawn. The distinction must be made on the basis of economics and practicality. Similarly we have some shapes which are simple monotonic functions of frequency, and some which are complex by any standard, but some others are difficult to classify.

A brief discussion of each deviation follows.

Design Error: The failure to exactly match the average repeater gain to the average cable loss characteristic is a deviation that, by definition, is exactly the same in every repeater section. By its very nature, it accumulates systematically from repeater to repeater- that is, the

Table 14-1

Transmission Deviations And Types Of Equalizers

	Deviations			Equalizer
Source	Shape	Rate	Addition	Type
Design Error	Complex	Zero	Systematic	Fixed**
Manufacturing	Complex	Slow	Random*	Manual
Maintenance	Complex	Slow	Random	Manual
Temperature:				
Cable	Simple	Fast	Systematic	Dynamic**
Repeaters	Complex	Fast	Systematic	Dynamic**
Aging:				
Active devices	Simple	Fast	Systematic	Dynamic**
Passive elements	Complex	Slow	Systematic	Manual
Residues	Complex	Moderate	Random	Dyn + Man

*Assuming quality control techniques to minimize systematic manufacturing deviations.

**Cause-associated shapes can be used to advantage.

deviation after transmission through "n" repeaters is "n" times the deviation in one repeater. Being the residue which remains when a large number of elements are used to approximately match a truly smooth characteristic, it will be complex in shape. It will, however, be constant vs time unless the design is changed. Therefore fixed equalizers at the end of each switching section can be used to correct for it. Since the number of repeaters between equalizers will be a function of geography, several designs of fixed equalizer are needed - e.g., one for 15 repeaters, one for 25 repeaters, etc. A section consisting of 17 repeaters will then have a residue which can be taken care of by a "mop-up" manual equalizer.

Manufacturing Deviations: There are no average repeaters, of course. Any batch of manufactured product will show both systematic and random deviations from the design values. If we are successful in using quality control techniques, the systematic component can be kept small, and the total manufacturing deviation of, say, a 25 repeater line section may be largely a random effect. The deviation of n repeaters is then in general about $\sqrt{n}$ times the deviation of one. The shape, however, will be complex, since it arises from the random deviations of many elements.

Maintenance: The distinction between this and the effects of manufacturing deviations is somewhat artificial. New product tends to be installed rather systematically in some new section of line. Established systems have associated with them a population of amplifiers (some in service, some removed and taken to central points for tube replacement). At routine intervals a number of these - say 20% of those in service - will be interchanged. Maintenance operations thus affect equalization somewhat differently than new manufacturing operations.

Cable Temperature: Changes in cable loss with temperature occur rapidly even though almost all the cable in the plant is buried. The cable loss of an L-3 New York to Chicago circuit, for example, can change 25 db per week at 7 mc. This corresponds to a ground temperature change of 2°F per week, and is not great on a percentage basis, since the total cable loss at this frequency is about 10,000 db. But compare 25 db per week with telephone transmission requirements! Obviously such a variation in transmission calls for automatic, practically continuous equalization - and these "dynamic" equalizers must be placed at frequent intervals along the line to minimize misalignment penalties. The design of the regulators which automatically adjust these equalizers will be discussed in a subsequent section. Fortunately the shape of this deviation is monotonic and precisely known.

Repeater Temperature: Since every effort is made to use elements which are not temperature sensitive, and any one element which is outstandingly sensitive is a focus of design attention until it has been brought into line, repeater temperature effects will tend to be complex functions of frequency. But they are systematic though small, and a New York-Chicago L-3 circuit would vary about 1.5 db per week if no corrective action were taken. (This is all of .006 db per repeater!) Again automatic regulation is called for, but at less frequent points along the line. In this case the equalizer shape is difficult to specify with certainty and the resulting network is difficult to design.

Active Device Aging: If we assume that tubes age in transconductance by about 1 db per year, we compute a gain change of about 1 db per week for a New York-Chicago L-3 system. The shapes are fairly simple, but several may be required if the amplifier is a multi-loop feedback configuration. The variation is fast enough to require automatic correction.

Passive Element Aging: Obviously a slow, complex shape, which can be corrected by manual equalizers.

Residues: Here we lump together all our ignorances and imperfections. We find that after our best efforts to regulate "fast" known shapes there are "fast" residues left over, and additional dynamic equalizers are needed. These can hardly be specified except as a result of observing the transmission deviations of the system vs time - if there were any other way to determine the required shapes, they wouldn't be classed under residues. In addition to fast residues, there will be slow or fixed residues arising from imperfections of design error equalizers and intermediate manual equalizers if these are used.

Control of Deviations at Source

In theory, deviations can be corrected or minimized at their source or at equalization points. Good repeater design practice dictates the minimizing of deviations at the source, within economic bounds. Even though the economic limitations are poorly defined, certain techniques can be profitably employed. Some of the more important of these are given below with brief comments on their application.

1) The repeater gain characteristic can be designed by the use of advanced mathematical tools such as Tchebycheff polynomials and the

potential analogue. Such a sophisticated design approach will help minimize the design error due to the finite number of elements in the repeater.

2) The use of negative feedback reduces gain changes due to vacuum tube aging, power supply fluctuations and other mu-circuit variations. While feedback is very effective in reducing the magnitudes of gain changes, it usually converts flat changes into moderately complex shapes which are ultimately more difficult to equalize than larger flat deviations would be. If feedback were not needed to suppress intermodulation products, one might prefer, from the equalization standpoint, to use an AGC type of regulation to compensate for tube aging, rather than internal feedback.

3) The provision of input and output impedances which closely match the characteristic impedance of the cable minimizes transmission ripples due to reflections. This usually is accompanied by signal-to-noise penalties as described in Chapter 7.

4) Quality control techniques may be imposed as a manufacturing requirement to insure that close tolerances on important transmission elements are maintained, with respect to both "average value" and "range".

5) The choice of a circuit configuration which minimizes the sensitivity of repeater gain variations to element variations is important, particularly with respect to the control of unwanted parasitic effects.

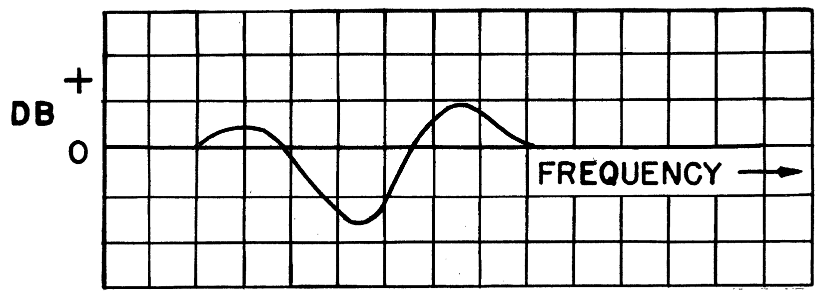
Full use of all these design techniques is still not enough; deviations will exist, will accumulate, and must be corrected by equalizers. Let us now inquire as to the devices and techniques available for equalizer design.

Manually Adjustable Equalizers

There are many "families" of characteristics that may be used for manually adjustable equalizers. Among these are "bumps", "power series", "cosines", and "time domain" types. Each has advantages and each has been used or considered for use in existing systems. The general characteristics and some of the main advantages and disadvantages of each of these families will be described briefly. It must be appreciated, however, that a full treatment of this subject is beyond this text, involving, in great measure, the theories of passive and active network design.

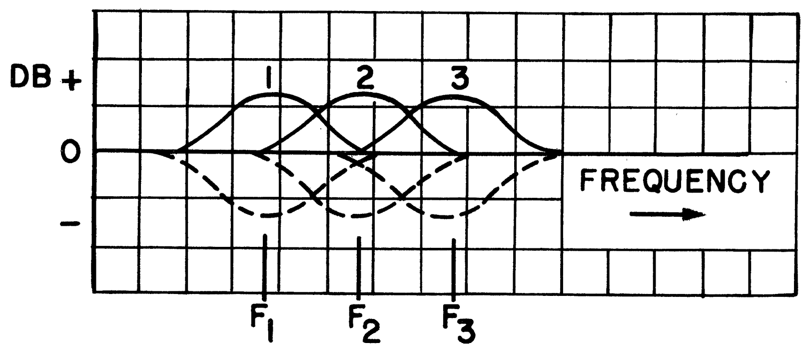
"Bumps"

One way of correcting a transmission deviation such as that shown by Figure 14-1 is by means of a series of equalizers of the form shown by Figure 14-2.



Transmission Deviation To Be Corrected

Figure 14-1



Bump Equalizer Shapes

Figure 14-2

These equalizers consist of a number of separate units each of which can be varied to give adjustment over a small range of frequencies. Thus unit #1 can be varied between the limits shown by the first solid and dotted pair of lines, unit #2 supplies the next, overlapping, bump, etc. The adjustments could be made by transmitting test tones at frequencies f_1, f_2 and f_3, using their amplitudes at the receiving end as criteria for proper adjustment - or the characteristic might be displayed by sweep techniques on an oscilloscope, and adjustment made on the basis of optimum visual flatness. Such equalizers are characterized by ease of design and of manufacture. They are difficult to adjust accurately, however, because the slopes of the bumps must overlap to give continuous coverage of the frequency range. Because of this, there tends to be large interactions between bumps - that is, the adjustment of one equalizer spoils the characteristic in the region overlapped by an adjoining bump. For wide band systems, the number of bumps required becomes very large - 88 were used in the 3 mc L-1 system.

"Power Series"

This family of equalizer characteristics is made up of related terms which form a power series in db versus frequency. Mathematically, the equalizers may be represented by

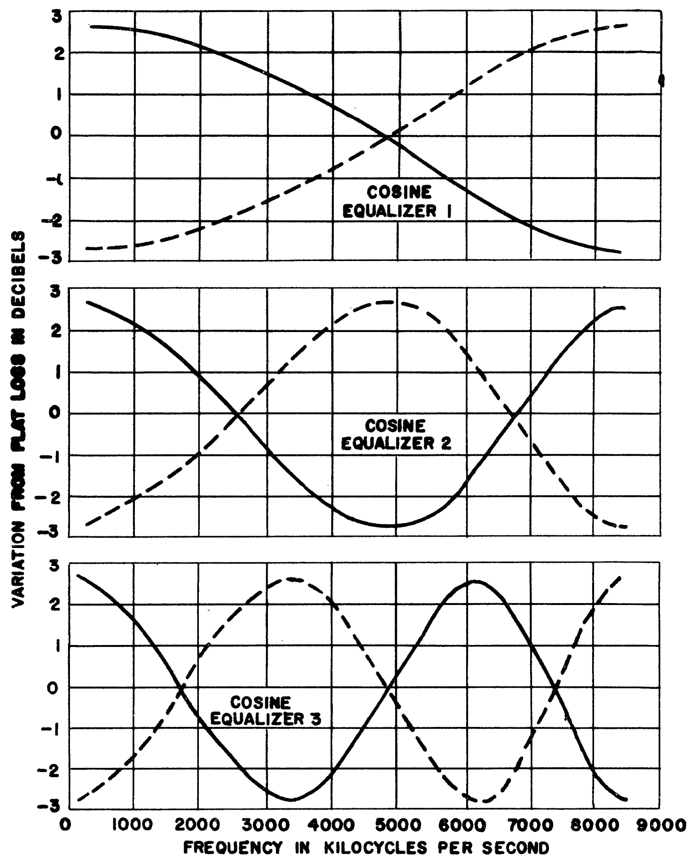
$$F(f) = A_1f + A_2f^2 + \text{---} A_nf^n \qquad (14\text{-}1)$$

Each term on the right side of equation (14-1) represents an equalizer characteristic. The first term, for example, is linear in db versus frequency and has a coefficient A_1 which is adjustable between predetermined plus and minus limits. This family of equalizers may be expected to be a powerful tool in the adjustment of sharp band edge characteristics. Although they have often been considered, such equalizers have never been designed for system use. Methods of adjustment have not been worked out.

"Cosines"

The use of cosine equalizers was suggested by the observation that manual equalizers are called upon to correct a wide range of complex deviation characteristics. Each of these characteristics is, in theory, expressible as a mathematical function in db versus frequency, which can be analyzed in terms of its Fourier components. If equalizers can be designed to correspond to each of these components and if each can be adjusted so that its magnitude is the same as that of the corresponding Fourier component in the deviation characteristic, then, in theory, any deviation characteristic can be corrected. Figure 14-3 shows, for example, the first three terms of a Fourier series applicable to equalizing a transmitted band extending from 0 to 8.5 mc. (It can be seen that the relationship between transmission frequency and the phase angle of the cosine is not linear. It was slightly warped for the reason mentioned below and because it simplified the equalizer network.) The accuracy of the correction will depend on the accuracy with which each term can be set and on the number of terms for which equalizers are provided.

Such equalizers were invented many years ago but were not used in a system until recent years because a suitable method of adjustment could not be found. A solution to this problem is found in the application of sweep techniques. By these means, a db-frequency characteristic is converted to a voltage-time characteristic. To adjust the equalizer, one need only minimize the electrical power represented by this characteristic. Instead of being constant, the rate of change of frequency of the sweep is slower at high frequency than at low. The warping mentioned above was such that, combined with this non-linear sweep, a non-distorted cosine time function is obtained. The net result of all



Shapes Introduced By The First Three Cosine Harmonics

Figure 14-3

this is to give more emphasis to transmission deviations at the high end of the band. In the L-3 system this is where the TV signal is transmitted, and as we shall see, TV is very sensitive to such deviations. The simple criterion - minimum power - theoretically permits adjustment of the knobs in any order, since correct compensation of any cosine term in the original wave results in a reduction of total power at diode output. In practice, of course, masking effects dictate that large deviations be corrected first.

This family of equalizers is thus characterized by ease and accuracy of adjustment, flexibility in the number of characteristics that can be corrected and flexibility in application to a system; if it is found that a given complement of equalizers does not achieve acceptable accuracy, equalizers corresponding to more terms in the Fourier series may be added. The number of terms required turns out to not very great - a 15 term equalizer is very effective for the 8 mc band of the L-3 system.

"Time-Domain" Equalizers

In Chapter 16 the equivalance of echoes and transmission deviations is discussed - deviations can be considered as causing echoes, or vice versa. If we add echoes which cancel those existing in a system, we will also correct transmission deviations. By providing taps on delay lines, and means for adjusting magnitudes of the signals picked off, and adding them to the original signal, any desired pattern of leading and lagging echoes around a signal can be provided. Used in conjunction with sweep adjustment techniques, such time-domain equalizers provide adjustment of gain and delay distortion. They are an essential tool in the equalization of long broad-band carrier systems for television transmission.

REGULATORS

In the course of transmission systems development, a number of types of devices have been called regulators. Of these, the most interesting are dynamic or back-acting regulators, analogous to the AVC circuit in an ordinary home radio receiver for AM reception. Automatic feedback regulation of this type is a fairly simple problem for an individual circuit. The rather extended treatment given here is because of the chain-action difficulties that are encountered when the principle is applied to long systems with several hundred regulated amplifiers in tandem. Before considering these problems in more detail, it will be of interest to consider briefly some of the various types of regulators.

Non-Feedback Regulators

In this type something is measured which is an indication of the system gain and a gain control or regulating network (capable of providing a cause associated shape) adjusted accordingly. For example:

1) In the type C system (open wire) a reference signal or pilot is transmitted continuously. At the receiving end it is monitored by a meter. When the transmission loss is too low or too high, a contact is closed, an alarm sounds, and the attendant manually readjusts the loss of the receiving terminal equipment.

2) In the K1 system (twisted pair in lead sheathed cable) a pair of wires in the cable is made one arm of a mechanically self balancing bridge located at the adjacent repeater. The bridge is continuously balanced in accordance with the resistance, and hence the temperature, of the pair and is linked to a gain control to offset the associated transmission variations.

3) In the L-3 (coaxial), at some line repeaters, a thermistor is buried in the ground adjacent to the repeater hut. Its resistance, and hence temperature, is a quite accurate measure of the cable temperatures for the preceding four miles of cable. This thermistor, as part of a simple resistance network, adjusts the gain of the amplifier as the ground temperature varies.

The attraction of the non-feedback regulator lies in its simplicity, reliability and economic advantage. Its disadvantage is that it is not self-checking. Whatever residual error it leaves is passed along to the next repeater. If the error is systematic it will accumulate in direct proportion to the number of repeaters traversed. It is the ability to overcome this difficulty which is the chief virtue of the feedback regulator.

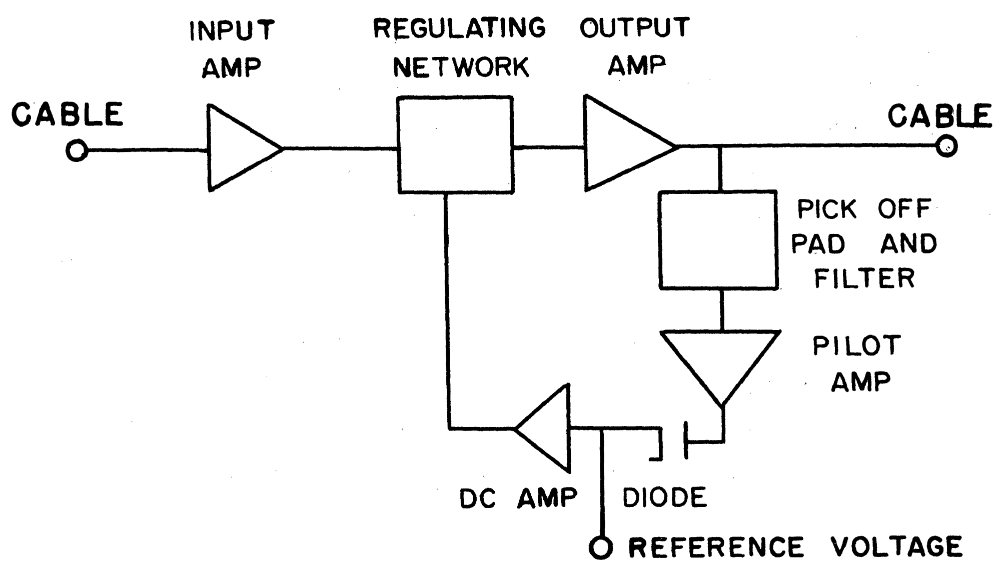
Feedback Regulators

In a feedback regulator the transmission is measured continuously by comparing a pilot signal with a reference. The difference, or error, is used to actuate a transmission control in the repeater until the system is in an equilibrium determined by the feedback around the control loop. For example:

1) In the type J system (open-wire) a pilot is measured and the gain mechanically adjusted until the pilot level is within prescribed limits. The gain control is not continuously controlled by the pilot but operates in steps as the signal varies. This avoids the dynamic stability difficulties encountered in a long chain of feedback regulators.

2) In the K2 system (pair in cable), and L-1 and L-3 systems (coaxial), a pilot is measured and controls the current through a thermistor in a regulating network, which in turn adjusts the gain of the repeater. The L-3 feedback regulator is shown schematically in Figure 14-4.

In the latter cases the feedback loop is always closed, and the individual regulators as well as long chains of them must be designed to be dynamically stable. In the case of the individual regulation loops this means meeting Nyquist's criterion for stability and is not a difficult design problem. The loops have moderate feedback and can be (in fact



L3 Repeater, Showing Regulation Loop

Figure 14-4

must be) cut off at very low frequencies and at low cut off rates so that the loop loss is very great before the phase shift exceeds 180°. The stability of a long chain of regulators is the controlling factor in the design.

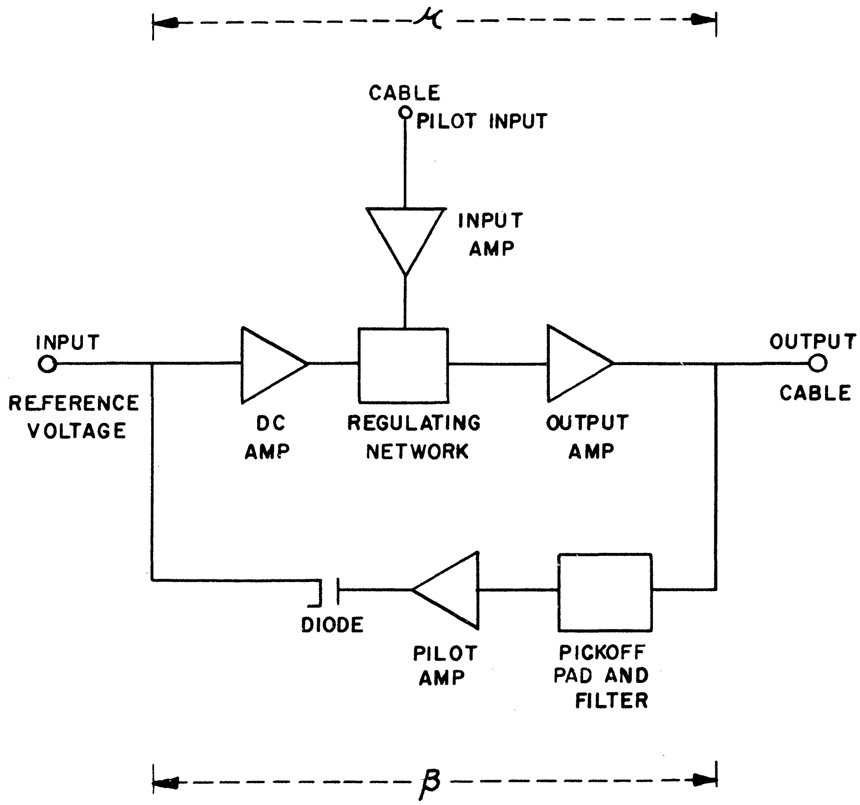
Chain Action

It was stated above that the chief advantage of the feedback regulator was that it does not permit the accumulation of errors from repeater to repeater. To illustrate this, let us assume that the line loss between repeaters changes by 1 db and that each repeater is equipped with a regulator with 20 db of loop feedback. The first repeater will reduce the deviation by .9 db leaving a .1 db residual error. The following repeater will have an input deviation of 1.1 db and will be forced to make a correction of .99 db leaving a residual error of .11 db to be passed along. Table 14-2 illustrates the action for additional repeaters. In effect, the error of the first regulator forces the other regulators to make a more accurate correction. Or, considering the total system deviation, the regulating system may be considered to have the feedback of one regulator, multiplied by the number of regulators. Thus, 100 regulators in tandem each having 20 db feedback are equivalent to a single regulator with 60 db feedback.

To generalize chain action effects, it is useful to draw an analogy between the feedback loop of the regulator and a conventional feedback amplifier. By comparing the various parts of the circuit in such an analogy, it is possible to formulate most of the important design requirements. The circuit of Figure 14-4 is redrawn from this point of view as Figure 14-5.

Table 14-2

Regulator Number	Input Pilot Change	Inserted Correction	Output Pilot Change
	db	db	db
1	1	0.9	0.1
2	1.1	0.99	0.11
3	1.11	0.999	0.111
4	1.111	0.9999	0.1111



L3 Repeater, Showing Feedback Structure

Figure 14-5

The d-c reference voltage against which the amplified and rectified pilot is measured is considered to be the "input". The RF pilot at the output of the amplifier is the "output", linearly related to the amplitude of the "input". The pick-off pad, crystal filter, and RF pilot amplifier make up the "β circuit" by means of which a portion of the output is fed back. Any change in these β circuit elements will affect the "gain" from input to output proportionately. The d-c amplifier, regulating network and output amplifier constitute the "μ circuit". The stability of these components will not be as important for accurate regulation, since the effect of changes will be reduced by the loop feedback. Note that as in other feedback devices the net input to the μ circuit is an error voltage, in this case the difference between the reference d-c and the rectified pilot.

In this view of the regulator, variations in the input pilot level are seen to be μ circuit disturbances. They will affect the output pilot level in the same way as changes in the gain of the output amplifier, a μ circuit component. The resultant disturbance to the output pilot will depend on the amplitude of the input disturbance and the loop feedback - which can be denoted as in any feedback circuit by the amount of μ gain and β circuit transmission.

Having developed the feedback nature of the structure and the roles of the components, the conventional feedback art can be used for the analysis of the individual regulator. One can show that

$$\frac{\text{Change in output pilot level}}{\text{Change in input pilot level}} = \frac{1}{1 - \mu\beta} \tag{14-2}$$

$$\frac{\text{System gain change to pilot frequency}}{\text{Change in input pilot level}} = \frac{\mu\beta}{1 - \mu\beta} \tag{14-3}$$

This result is not very surprising but it can be used to analyze the chain action of regulators. Many different cases must be considered. Sometimes the pilot levels change because of an effect distributed all along the system. In other cases the change occurs only at the input to the line. Sometimes the pilot level changes are the important effect. In other cases the importance resides in the gain change to the signals. In all cases the results may be complicated by the fact that $\mu\beta$ is, in general, a complex number and thus phase as well as amplitude is important.

Gain Enhancement

Gain enhancement is the ratio of output pilot variation to input pilot variation, for either a single repeater or a system. Numerically in db it is equal to 20 $\log_{10}$ of this ratio. Thus if the input pilot varies 0.1 db, and the output pilot 0.2 db, we express this by saying we have 6 db of gain enhancement. Since the magnitude of the gain enhancement is determined by the magnitude and phase of the feedback around the regulation loop, it will in general be a function of the frequency of the pilot variation.

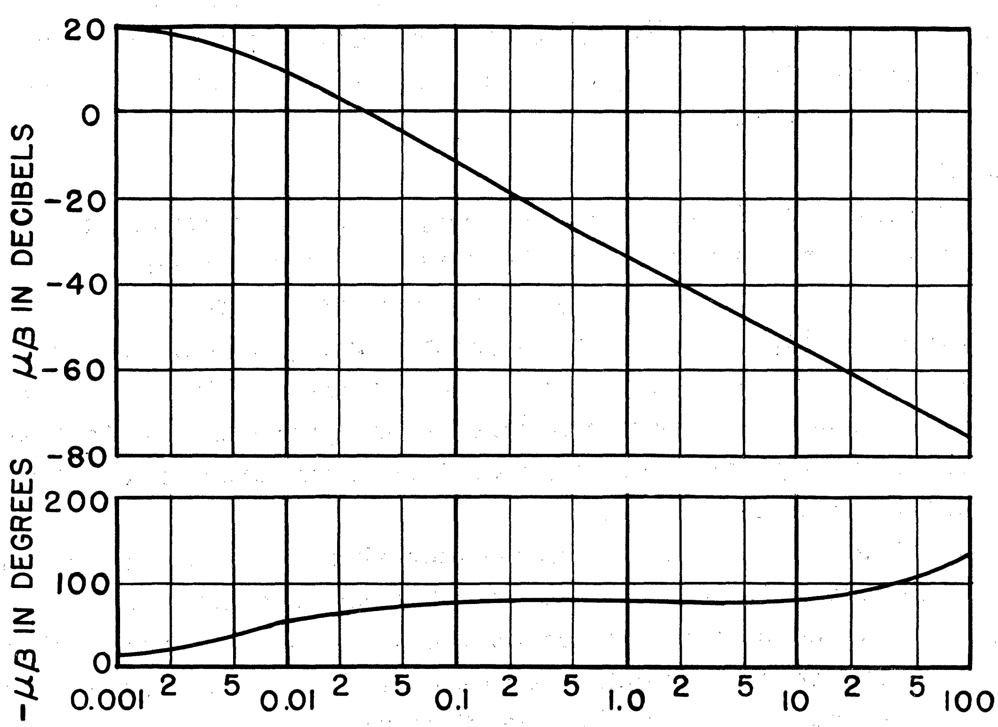
As an illustration of gain enhancement phenomena, consider the results when the feedback loop of a regulator has the characteristics shown by Figure 14-6. Suppose the pilot at the input to a system having a large number of regulators in tandem varies at a rate of 60 cycles per second. Using the values of Figure 14-6 we find that the input pilot change at each dynamically regulated repeater will be amplified, rather than suppressed - by about .0012 db. In passing through a thousand regulators, the original input pilot variation will be amplified by 1.2 db. As a function of frequency, the gain enhancement corresponding to the $\mu\beta$ characteristic of Figure 14-6 is shown on Figure 14-7 for a single regulator.

Some gain enhancement is inescapable in any conventional regulator design. Bode has demonstrated that in any feedback loop in which the $\mu\beta$ declines more rapidly than 6 db per octave, that is, those loops which will have an ultimate phase shift of greater than 90°, the average regeneration or degeneration over the complete frequency spectrum is zero. This means that if a feedback regulator decreases the effect of a pilot change for some rates of change (disturbance frequency) it must increase the disturbance over some other band such that:

$$\int_0^\infty \frac{1}{1-\mu\beta}\, df = 0 \qquad (14\text{-}4)$$

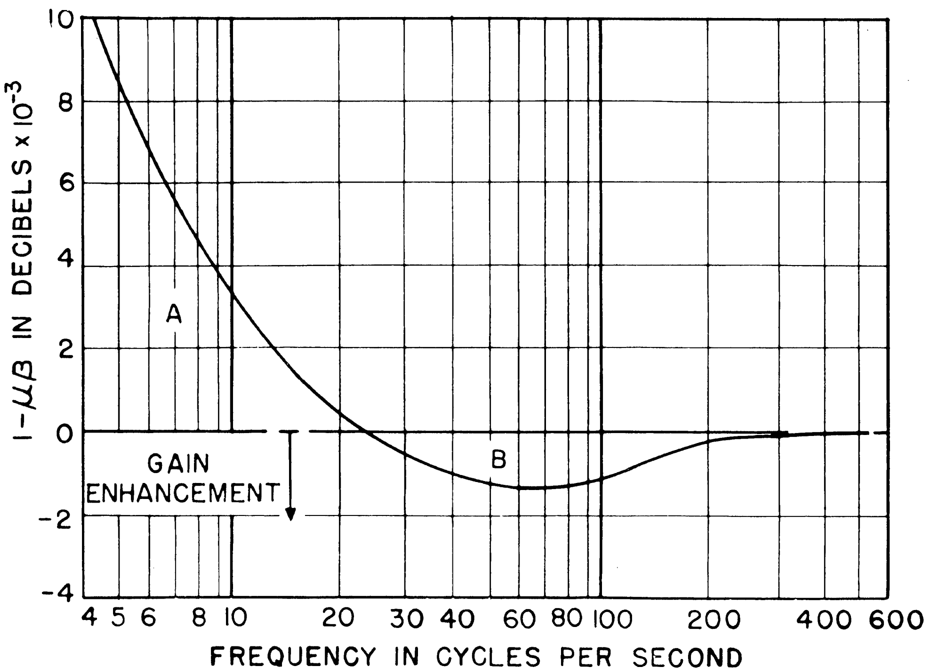
$\frac{1}{1-\mu\beta}$ = feedback term in db

f = frequency in cps



$\mu\beta$ of Regulation Loop in db and **Phase**

Figure 14-6



Magnitude of Gain Enhancement

Figure 14-7

Graphically, if the abcissa of Figure 14-7 were arithmetic rather than log frequency, area A would equal area B. The ordinate of the area of gain enhancement (area B) is ordinarily only a few thousandths of a db and the frequency abcissa very great, the ordinate of the area of suppression is ordinarily tens of db and the frequency abcissa relatively small. The design objective must be to control $\mu\beta$ so that its phase cannot exceed 90° (a necessary condition for $\frac{1}{(1-\mu\beta)}$ to exceed unity) until the amplitude of $\mu\beta$ is so small that the increase in disturbance and the gain fluctuations due to it will be tolerable. This, of course, means a $\mu\beta$ cut-off rate of 6 db per octave or less until the required loss is achieved.

Other Effects of Chain Action

Even when the design of the regulation system is such that there is no gain enhancement at a particular disturbance frequency, because the angle of the feedback around the regulation loop has not exceeded the 90° point, long chains of regulators may penalize certain system effects. As an example, consider the effect of television-induced compression. The presence of the television signal reduces the gain of the line amplifier. The effect is small but cumulative. In the absence of regulator action it merely compresses the television signal slightly and makes a negligible change in the contrast rendering of the picture. However, the regulators observe a gain change to the pilots and attempt a correction. The very rapid changes are ignored but 60 cps, for example, is partially corrected. This introduces a 60 cps gain change which will lag the picture and therefore must meet 60 cps bar pattern reouirements. This problem is minimized by keeping the regulator response low at 60 cps.

The degree to which pilot compression is converted into unwanted gain changes at signal frequencies can be evaluated from Figure 14-8. If

n = number of regulated repeaters in tandem

ΔP = gain change of each section in db, at pilot frequency, caused by compression (before regulation)

ΔG = total system gain change in db, at pilot frequency, at end of n sections, after regulation

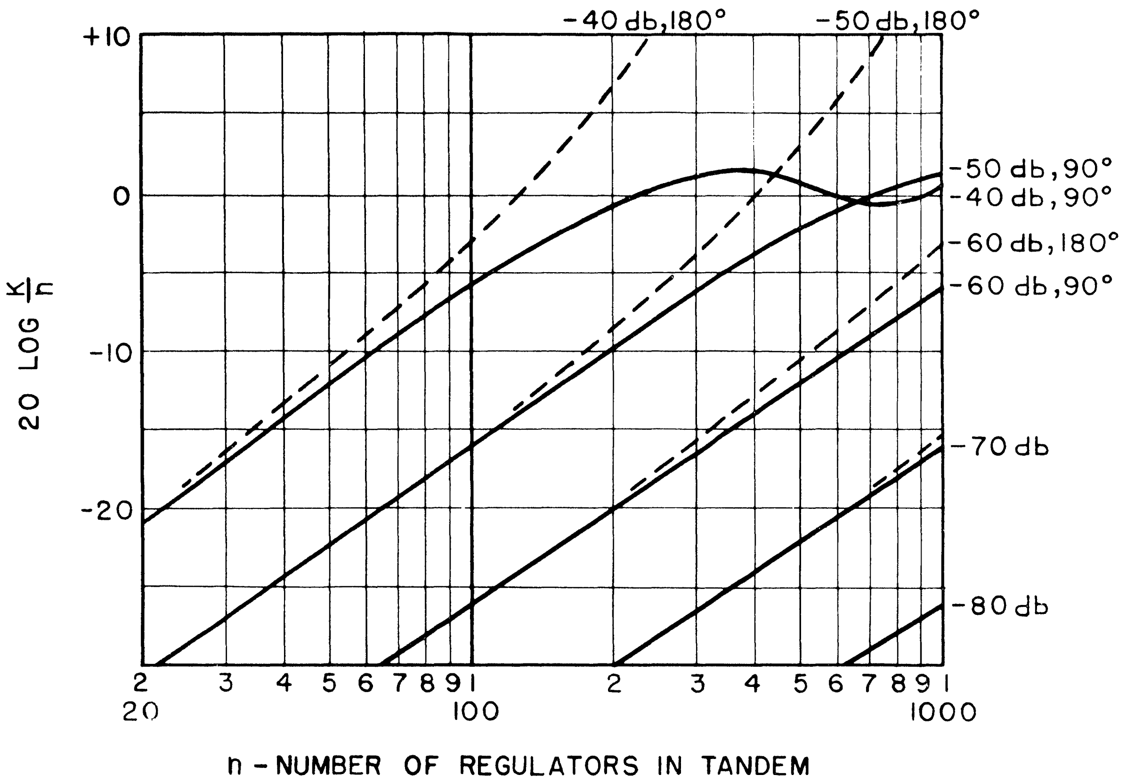
then it can be shown that for small values of compression

$$\frac{\Delta G}{\Delta P} = \frac{1}{\mu\beta}\left[\left[\frac{1}{1-\mu\beta}\right]^{n} -1\right] \quad -n = K \tag{14-5}$$

or

$$\Delta G = \frac{K}{n} \cdot n\,\Delta P$$

The gain change of the system at signal frequencies, especially near the pilot, will be very nearly equal to ΔG. Expressing K/n in db, we can plot it as a function of the number of regulators, with feedback as a parameter, obtaining Figure 14-8. From these curves we can see, for example, that if $\mu\beta = -70$ then a chain of 700 regulators will insert a total gain change approximately one tenth of the total compression. If $\mu\beta$ were -50 db, the total gain change would equal the compression, which could seriously degrade certain pictures. Since we are concerned with the part of the envelope spectrum where $\mu\beta$ is small and decreasing at 6 db per octave, the 90° phase value for $\mu\beta$ is of most interest. Values are also given for $\mu\beta$ shifted 180° from its in-band



Chain Action - Unwanted Gain Change
vs Compression

Figure 14-8

value, which is the worst condition for such chain action effects as this, as it is for gain enhancement. However, it will be observed that for many values of $\mu\beta$ this effect is, unlike gain enhancement, nearly independent of $\mu\beta$ phase.

Design Limitations of Regulation Loop

The need to control adverse chain-action effects, and other system considerations, impose a number of limitations on the design of the regulation loop and its components. A few of the more general principles will be discussed with reference to Figure 14-4.

Output Amplifier

The transmission of the output amplifier is dictated by the insertion gain requirements of the repeater and obviously cannot be varied to suit the requirements of the regulation loop.

Pilot Level

The pilot is transmitted at as high a power as can be tolerated in view of intermodulation effects in the transmission band. The highest tolerable level is obviously advantageous for good pilot-to-noise ratio, and because high pilot power will require the least gain in the pilot amplifier or permit the greatest loss in the pick-off pad and crystal filter, both desirable objectives.

Crystal Filter, Pick-Off Pad and Nick Effect

The transmission of the system is to be controlled by the pilot signal. The economical use of the system transmission band requires that speech and other varying level signals be very close in frequency to the pilot and, if they are not to affect the regulator in a chaotic way, we must provide discrimination against them. This is the function of the crystal filter which passes the pilot but provides from 50 to 60 db loss to signals more than 1 or 2 kc removed from the pilot frequency. The pass-band of the filter is made somewhat wider than would be required by filter design limitations because the phase shift close to its cut-off would otherwise increase gain enhancement effects to an excessive degree. A filter designed for a characteristic impedance of 25,000 ohms and shunted directly across the 75 ohm coaxial would have negligible effect on the through transmission if it were not for sharp resonances and anti-resonances in the crystal outside the pass-band. These abrupt variations in shunt impedance would cause small but exceedingly sharp deviations in the main transmission band that would be impossible to equalize. In order to hold this "nick effect" of the crystal filter to a tolerable value a simple L pad is used between the filter and the line. With a voltage loss of 17 db in this pad the "nick effect" is held to the order of .15 db for a 4000 mile system, a tolerable value.

Pilot Amplifier

The design of the pilot amplifier is determined by the following considerations:

1) Input is set by the pilot level, the loss of the pick-off pad, and the loss of the terminated crystal filter. The gain of the pilot amplifier must raise the input level to a voltage equal to the reference voltage.

2) The pilot amplifier is part of the "β" circuit of the regulation loop and its gain must therefore be very stable. This is accomplished by about 25 db of feedback around the pilot amplifier alone. The forward gain without feedback must therefore be equal to the sum of the gain above and the feedback. This gain can be quite high; in L3 it exceeded 100 db.

DC Amplifier

One of the design objectives for the regulator is reasonably constant $\mu\beta$ for the entire range of operation. The requirement is not very exacting but less feedback would obviously give poorer regulation and more would increase the gain enhancement. Since the sensitivity of the regulating network (change in transmission gain per unit change in thermistor current) is not constant over the regulation range and all other components of the regulating loops are linear, the dc amplifier must provide a nonlinear transconductance to complement the regulating network sensitivity.

Bias Drift

The input signal to the dc amplifier, which is the "error voltage" difference between the reference voltage and the rectified pilot, will also be affected by a third voltage - the contact potential of the tube. Drifts in contact potential will therefore have the same effect as changes in the reference voltage or in the pilot. One way of minimizing the effect of this drift is to make the reference voltage and rectified pilot voltages as large as possible, so that small percentage pilot changes result in large dc voltage differences. Contact potential changes are then equivalent to very small pilot changes in their effects on through transmission.

Regulating Network

Since the regulating network is primarily an element in the main transmission path, its design as a component in the regulation loop is hedged about with restrictions. Whether the regulating network employs thermistors, grid bias, or other means of regulation, the voltage, current, and impedance limits will be dictated by transmission path requirements as well as by regulation loop objectives.

Gain Enhancement vs Regulator Signal to Noise Ratio

There is a necessary compromise between a good design for gain enhancement and a signal to noise problem in the regulation loop. It is apparent from the previous discussion that there is a large gain from the pilot pick-off to the dc amplifier output. This gain is such that in a regulator at the end of a long L-3 system the noise at pilot frequency in the crystal filter pass-band (2 kc) would overload the dc amplifier. The crystal filter, although it restricts the noise power that may reach the diode rectifier, is not sufficiently narrow to eliminate this difficulty, and cannot be made much narrower because of the large phase shift and hence gain enhancement it would produce near its cutoff. To control the noise, therefore, it is necessary to introduce an RC cutoff in the dc amplifier after the diode. This reduces the output noise to a tolerable value without excessive gain enhancement.

Multi-shape Regulation

The application of feedback regulators can be extended to the automatic regulation of several cause associated shapes simultaneously. This requires the transmission of as many pilots as shapes anticipated and the use of a simple computer between the diode rectifiers and dc amplifiers to determine the unique combination of shapes represented by the pilot levels. It is by this means that the several shapes, mentioned earlier in this chapter, are controlled without interaction in the L3 system. The details of circuitry are beyond the scope of this discussion.

Bibliography

1 - C. H. Elmendorf, et al, The L3 Coaxial System, BTL Monograph 2090, Bell Telephone System, 1953.

2 - H. W. Bode, Variable Equalizer, Bell System Tech. Journal, April, 1938.

3 - R. L. Lundry, Attenuation and Delay Equalizer for Coaxial Lines, A.I.E.E., Trans., 68, 1949.

4 - P. H. Richardson, Variable Attenuation Network, Patent 2,096, 028.

5 - C. R. Wylie, "Advanced Engineering Mathematics", McGraw-Hill Book Company, Inc., 1951.

6 - M. D. Stoughton, "Use of Tchebycheff Polynomials in the Analysis and Synthesis of Networks", MM 51-2120-34, October 17, 1951.

Chapter 15

SHAPED LEVELS, FEEDBACK; COMPANDORS; TASI

Up to this point we have made a number of restrictive assumptions in our discussion of the design of single-sideband AM telephone systems employing cables and electron tube repeaters. For example, we have assumed that, for a given power of signal in any channel at the zero db transmission level point, we will obtain the same signal voltage at the grids of the output stages of the repeaters in the high frequency line, regardless of the carrier frequency slot assigned to the particular message channel. In terms of previous definitions, C, the relationship between dbv at the grid of the output stage and dbm at zero level, has been assumed to be constant vs carrier frequency. Since cable loss, and therefore thermal noise at repeater output, increase with frequency, flat transmission from output grid to zero level usually results in system performance which is better at low carrier frequencies than in the top transmitted channel. If the top channel just meets requirements, the lower channels may have unnecessary margins in signal-to-noise performance. A similar assumption has been that feedback in the repeaters is flat vs carrier frequency.

In this chapter we shall discuss the effect of removing such restrictions on system design, and also the effects of compandors, and of more efficient utilization of channels on a time-sharing basis.

Shaped Feedback

The effects, on signal to noise performance, of shaping feedback are relatively simple insofar as modulation noise is concerned. An exact analysis would show that to compute rigorously the degree to which a $2\alpha-\beta$ product, for example, will be suppressed by feedback, we must take into account the feedback at the fundamentals, at the 2α frequency, and at dc, as well as at the product frequency $2\alpha-\beta$. The magnitudes involved, however, are such that in most practical cases, only the feedback at the product frequency need be considered. The values of feedback at other frequencies affect only small correction terms, which can be ignored with negligible error if the feedback at these other frequencies is greater than about 6 db. Shaped feedback is therefore a tool which we can use to make modulation noise any desired function of carrier frequency, within the limits of our ability to prescribe the shape of the curve of $1-\mu\beta$ vs frequency.

A second, less easily estimated, result of shaping feedback is the change which occurs in the so-called $\mu\beta$ effect - the variation of insertion gain as active devices age. As a contributor to misalignment, the $\mu\beta$ effect affects signal to noise ratio. Shaping the feedback may make the $\mu\beta$ effect either more or less severe, and the resulting change in system signal to noise ratio must be evaluated for each individual case. We shall find, however, that in general we are led to ask for less feedback at high frequencies, and we thus eliminate the sharp corner of the feedback curve at top transmitted frequency which is a feature of maximum flat in-band feedback. Associated with this sharp corner is a sharp $\mu\beta$ effect. The result of shaping feedback, then, is likely to be a $\mu\beta$ effect which may be larger, but which can be more easily equalized because it is a simpler shape.

"Shaped Levels"

By the insertion of suitable networks at the transmitting and receiving terminals, we can make the signal voltage at the output stage grids of the repeaters follow any desired function of carrier frequency. By combining this mechanism with shaped feedback, an improvement in overall performance can often be obtained. Consider, for example, the curves of Figure 15-1. This is the same system shown in Figure 10-4. There levels were set to optimize signal to noise ratio, with third order noise dominant; now more margin against overload has been asked for, and levels have been lowered to achieve this additional margin, thus incidentally decreasing modulation noise. Third order, of course, decreases more rapidly than second order noise.* In Figure 15-1, the total noise in the lowest channel is about seven db less than the noise in the top transmitted channel.

These computed results are based on a repeater having flat feedback, and assume networks at the terminals which are designed to make the output grid a "flat voltage level point". The thermal or random noise therefore has, at zero level, almost the shape of the cable loss characteristic - differing only by the shaping of the noise figure and the shaping of the output coupling network. The thermal noise voltage from grid to cathode of the output stage of the line repeaters will, since C

---

*The change in slope of second order modulation noise in Figure 15-1 vs that shown in Fig. 10-4 is a consequence of using "noise power" values of $\rho_x + s_x$ instead of the subjective values; this also changes the relative magnitude of 2nd and 3rd order.

is constant, have the same shape vs frequency as the noise in dba at zero level. The shaping of modulation noise vs carrier frequency is caused by the variation in the numbers of intermodulation products vs the carrier frequency position of the disturbed channel, and by the different values of s_x for the various products entering into the total second and third order noise.

If random noise were the only contributor, equal signal-to-noise ratio in all channels could be obtained by "shaping" the signals so that "level" at output grid would follow the random noise curve of Figure 15-1, this would mean transmitting the lowest channel at a level (at output grid) 26 db below top channel level, as in Figure 15-2.

Obviously, however, modulation noise cannot be ignored, and any shaping of levels immediately raises the question: what does modulation noise then become? (A similar question as to the effect of shaping on the load carrying capacity requirement also arises, to be considered later).

We have seen that to compute the modulation noise in the flat-level, flat-feedback case we must compute the probable number of products of each type falling in the disturbed channel, and also must have data on the magnitudes of the fundamentals at the point where inter-modulation occurs. Making fundamental signal magnitudes a function of carrier frequency will obviously complicate this computation.

There is, as yet, no rigorous general solution to this problem, but a seemingly adequate approximate method of computation has been worked out. This method involves approximating the shaped levels at output grid with a series of steps (1, 2, 3) as indicated in Figure 15-2. Two or three steps are usually adequate, but the process can be extended to a greater number of steps if necessary, and the steps can be made unequal in width. The height of the step is intended to approximate the "average power" channel in the band covered by the step, rather than the average of the db spread.

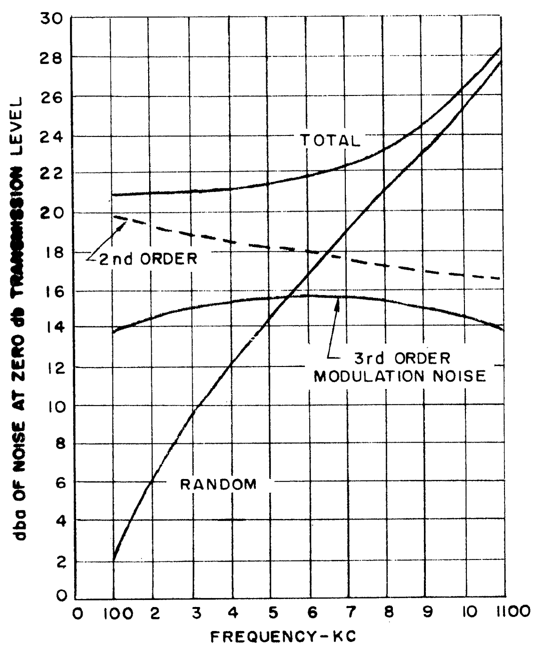
Consider now the $\alpha+\beta-\gamma$ products falling in the k channel, formed by fundamentals A cos $\omega_\alpha t$, B cos $\omega_\beta t$, C cos $\omega_\gamma t$. In earlier chapters we have assumed that except for differences between the volumes of individual talkers, A=B=C at the grid of the output stage of a line repeater. Now we have the situation that some $\alpha+\beta-\gamma$ products will be formed by three fundamentals of "Step 1" magnitude, some by two fundamentals in "Step 1" and one in "Step 3", and so on. In all, there will be ten different types of $\alpha+\beta-\gamma$ products, formed by fundamentals from various steps of Figure 15-2 as shown in Table 15-1.

Table 15-1
Fundamental Magnitudes

Product Magnitude	Step 1	Step 2	Step 3
1	A, B, C		
2	A, B	C	
3	A, B		C
4	A	B, C	
5	A	B	C
6	A		B, C
7		A, B, C	
8		A, B	C
9		A	B, C
10			A, B, C

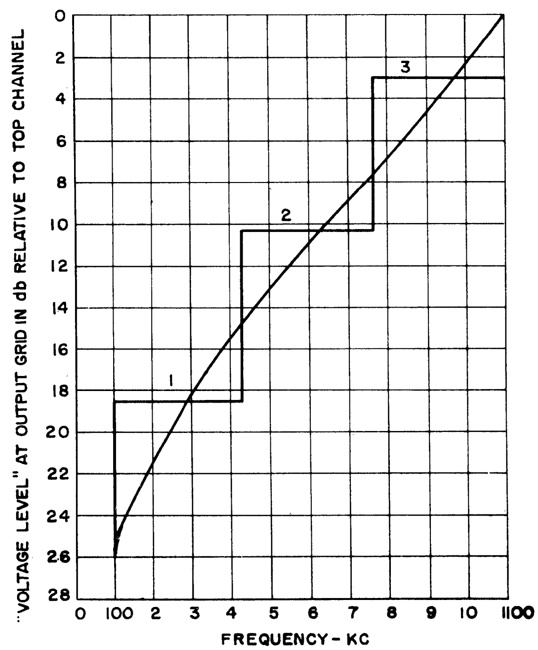
Similar tables might be constructed for other types of product, such as $\alpha+\beta$.

Basically, the same counting methods that were used to determine the total number of $\alpha+\beta-\gamma$ falling into channel k can be used to break this total number of products down into component types of the sort shown in the table. Like the original product count, this is a tedious process, and a rough guide to the probable results is useful. Table 15-2 shows, for various products and values of k, the product distribution found for a system of channels extending from 100 kc to 1100 kc. It would be a reasonable first approximation to assume, for a slightly different system, that the relative number of products of each magnitude would be about the same. In analyzing a system of 50% greater bandwidth, for example, one would construct a new 3-step staircase of levels, find the total number of x-type products, by the methods of Chapter 12, and scale the values given in Table 15-2 by the ratio of the total numbers of products for the two systems. If, however, the frequency allocation were substantially different, so that the approximation would be too inaccurate, more fundamental methods of counting must be used.



Noise vs Frequency From Various Sources, Computed for a Particular System: Flat Levels, Flat Feedback

Figure 15-1



"Voltage Level" at Grid of Output Stage Shaped to Follow Random Noise Curve, and Approximating Steps

Figure 15-2

Thus far we have spoken of the shaped signal in db relative to the top channel level. There is, however, no reason to assume that this top channel level would, for the shaped signal case, be the same as in the flat level case. The new level chosen will depend, as the old level did, on what limitation (modulation noise of either type, or overload) is of dominating importance. Assume for the moment that a particular level choice has been made. The dba of noise to be expected from each group of the sort shown in Table 15-2 may then be computed. (At this point one may begin to wonder how the empirically determined values of s_x and ρ_x given in Table 12-1 would be affected by such widely differing values of fundamentals magnitudes, however). These values of dba may then be added on a power basis to find the total modulation noise of each type in various disturbed channels.

Table 15-2

Product Distribution for a System 100 kc - 1100 kc

I "α-β" Products

From Step Combinations	No. of Products Bottom Chan 100 kc (k = 25)	Mid-Chan 600 kc (k = 150)	Top Chan 1100 kc (k = 275)
1, 1	58	0	
1, 2	25	16	
1, 3	0	67	
2, 2	58	0	
2, 3	25	17	
3, 3	58	0	
	224	100	0

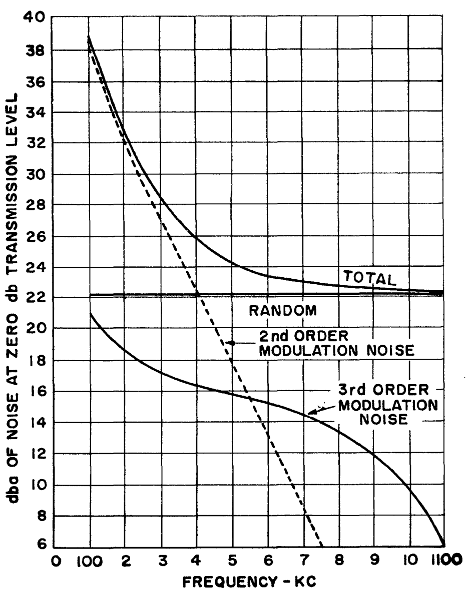
II "α+β" Products

From Step Combinations	No. of Products Bottom Chan (100 kc)	Mid-Chan 600 kc	Top Chan 1100 kc
1, 1		32	
1, 2		18	24
1, 3	No	0	59
2, 2	Products	0	29
2, 3		0	
3, 3		0	
	0	50	112

III "α+β-γ" Products

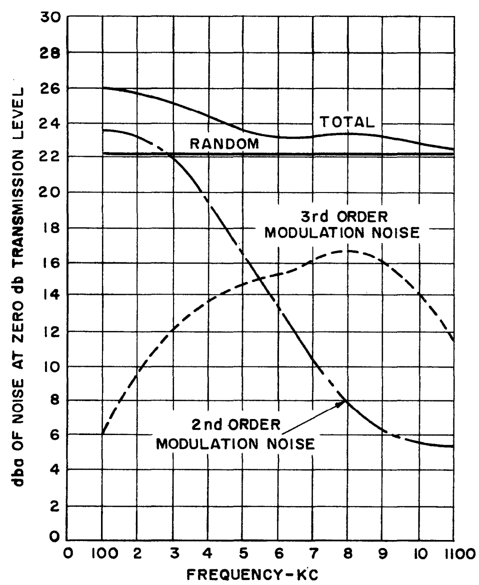
Step Combinations	No. of Products Bottom Chan 100 kc (k = 25)	Mid-Chan 600 kc (k = 150)	Top Chan 1100 kc (k = 275)
1, 1, 1	1640	400	0
1, 1, 2	1681	5083	0
1, 1, 3	0	903	3403
1, 2, 2	3403	1261	1681
1, 2, 3	3403	5208	3486
1, 3, 3	3485	861	0
2, 2, 2	0	2480	0
2, 2, 3	1763	1303	3403
2, 3, 3	0	5165	1722
3, 3, 3	0	462	1681
	15, 375	23, 126	15, 376

In this way we arrive at the results shown in Figure 15-3 for the same system as that of Figure 15-1, but now with levels shaped as in Figure 15-2. In this particular case, top channel level has been set 5.5 db higher than in the original flat level case, for reasons to be discussed later. Feedback is still flat. We see that modulation noise is now excessive at low frequencies, but negligible at high frequencies. Shaping the feedback will bring the modulation noise into better balance. We recall, from the discussion of the ϕ scale in Chapter 13, that if we ask for more feedback at low frequencies we must pay for it with high frequency feedback, but it appears that we can afford to. Theoretically one might specify an optimum shape of feedback vs frequency, but in practice this exactitude is not worthwhile. There will be a number of other considerations that will govern the final design of the feedback loop of the repeater, ($\mu\beta$ effect, manufacturing ease, number of components, for example) and our computations of modulation noise are only approximate. A rough estimate of what slope of $\mu\beta$ vs frequency is reasonably good would appear to be sufficiently accurate. If we let this slope be 6 db per octave we obtain the final results shown in Figure 15-4. The top channel signal-to-noise ratio has been improved by 5.5 db, as compared to Figure 15-1.



Random and Modulation Noise at Zero Level With Flat Feedback - "Voltage Levels" Shaped to Follow Random Noise at Output Grid

Figure 15-3



Random and Modulation Noise at Zero Level With Shaped Feedback and Shaped Signals

Figure 15-4

Noise in the lowest channel, however, is +26 dba, which is only 2 db less than top channel noise was originally. It can be seen from Table 15-2, and from consideration of the "voltage levels" of the three approximating steps of Figure 15-2, that modulation noise in the lower channels must be dominated by products formed by intermodulation of fundamentals being transmitted at high frequencies and high voltage levels. Modulation noise at output grid in the lower frequency channels is, therefore, almost like thermal noise in its independence of low frequency channel "voltage level" at output grid. In the high frequency channels, modulation noise is not dominant, and would be very little changed by a moderate increase in "voltage levels" of low frequency channels at output grid. We can therefore improve the signal to noise ratio in the low frequency channels almost db for db by raising their "voltage levels" at output grid, without appreciable penalty at higher frequencies. This would be a second, and better, approximation to optimum shaping. The selection of a roughly optimum level adjustment in a case like this is so much a function of the parameters of the individual system that it does not seem worthwhile to work out general equations for the purpose.

Shaped Signal Levels - Overload

Another problem that arises when transmission "levels" as measured at the output grid are not flat vs frequency is the computation of the required load carrying capacity. Here again no rigorous mathematical approach has yet been worked out and approximate solutions are used instead.

The average rms voltage at the output grid during the busy hour can be found by integrating the area under the curve of signal voltage "level" vs frequency. A good approximation is to divide the total band into a number of steps as in Figure 15-2, and find the rms voltage corresponding to each group of channels. Thus for the 250 channel system of Figure 15-2, the average power at the zero level point, assuming an activity factor of 0.25, is (see Eq.12-2) -11 dbm + 10 log 62.5 = +7 dbm. If the output grid were a flat voltage level point the rms voltage there would be

$$(+7 - C) \text{ dbv.}$$

Since each step contains one-third of the channels, the rms voltage contribution of each third of the spectrum in the "flat voltage level" case would be (since 10 log 1/3 is -4.77)

$$(+7 - C - 4.77) \text{ dbv}$$

The steps are not flat at the level of the top channel, however. They are, relative to top channel, at levels of -18.5, -10.3 and -3.1 db. Letting C_t now be the value of C for the top channel, then, the three contributors to total rms grid voltage are:

Step 1	$+7 - C_t - 4.77 - 18.5 =$	$-C_t - 16.27$ dbv
Step 2	$+7 - C_t - 4.77 - 10.3 =$	$-C_t - 8.07$ dbv
Step 3	$+7 - C_t - 4.77 - 3.1 =$	$-C_t - 0.87$ dbv
Total		$-C_t - 0.01$ dbv

The evaluation of the multi-channel load factor "Δ_C" is not so readily made. At present the approximation used is to take the value of Δ_C (given by Figure 12-2) corresponding to the one-third or one-quarter of the total number of channels, on the reasonable assumption that only the channels transmitted at high levels will contribute appreciably to the difference between average rms voltage and the instantaneous voltage during 1% at the busy hour. (Obviously if the shaping of voltage levels is less steep, the value of Δ_C can be chosen to correspond to, say, one half the total number of channels). Thus for the 250 channel system of Figure 15-2, a reasonable value for Δ_C would be that corresponding to an 83 channel system, or about 15.7 db. The repeater must therefore carry, without overloading, a single frequency sine wave having an rms voltage at output grid of $(-C_t + 15.69)$ dbv.* One could test the system by applying a +15.7 dbm sine wave signal in the top channel at zero level, or about 40.7 dbm in the bottom channel at zero level.

It is of interest to compare this sine wave output grid voltage to be carried by the repeater with the sine wave to be carried in the flat level case, which would be -C+7 plus the multi-channel load factor (Δ_C of Figure 12-2 for N=250) or -C+7+13.3 or -C+20.3 dbv, rms. For the same signal to thermal noise ratio in the top channel (i.e., if C_t for the shaped level case were equal to C of the flat level case), the overload requirement for the repeater eased by 4.6 db (15.7 vs 20.3 dbm). The rms voltage change is 7 db, but Δ_C has changed, unfavorably, by 2.4 db.

- -

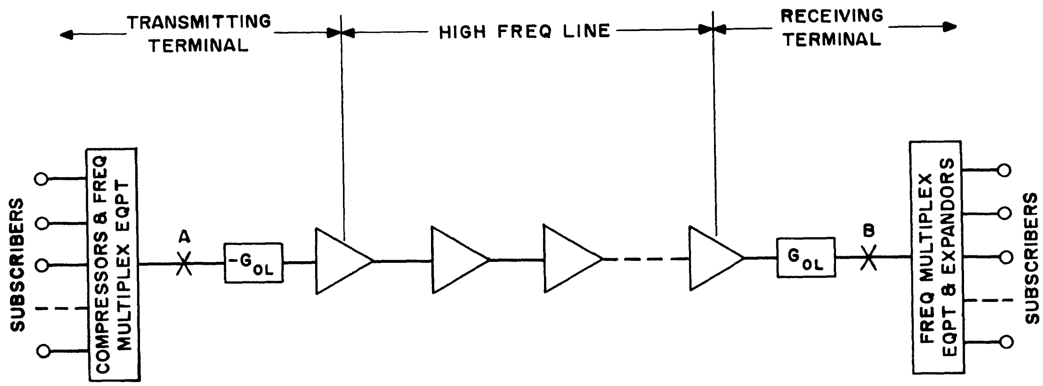
*The .01 db is carried along to show the method, of course.

Which brings us, finally, to the question of how one sets C_t (shaped level) relative to the value of C of the flat level case. In the example we have been discussing, the attempt was to hold overload margin constant, and take the 4.6 db as a top channel noise advantage. (there is a discrepancy here, since the noise advantage computed turned out to be 5.5 db, indicating that signal voltages were set a little high, penalizing overload margin by 0.9 db. This is due merely to the fact that inconsistent assumptions as to step approximations of total rms power and choice of multi-channel load factor were made by the various engineers involved.) The final choice of C_t would depend on how one chose to invest the profit of shaping levels - in overload margin, in noise margin, or in lengthening repeater spacing, and this decision would depend on the particular situation and, for example, expected future aging effects.

TASI

The word "Tasi" is formed from the initial letters of the phrase "Time Assignment Speech Interpolation". It has been pointed out that on the average the one-way channels of a broadband carrier system will be carrying speech only 25% of the time. This does not represent very efficient use of an expensive transmission facility. With appropriate switching and supervisory signalling, it is possible to increase this percentage appreciably - up to about 60%. A considerable amount of terminal equipment is required - as soon as a subscriber A (whose interest in another has already been established) starts to talk, a one-way connection to the distant subscriber B must be set up thru the high frequency line, without appreciable loss of the initial syllable. When A is silent, his high frequency channel must be released for the use of another talker. His next speech spurt will, in all probability, be transmitted in a different carrier frequency slot than his first. In the opposite direction, high frequency channels are similarly assigned to B when he "demands" them by speaking, and released when he is silent.

The problems involved in recognition of speech spurts, making the connections, and protecting against the condition of "no channel available", or lock-out, are outside our scope here. So far as the layout and design of the transmission system exclusive of terminals is concerned, the application of TASI to a system is merely an increase in the activity factor. This will increase the modulation noise by making the probable number of products more nearly equal to the total possible number; it will have two effects with respect to load carrying capacity required - it increases the rms load, but decreases the multi-channel load factor Δ_C.



Block Schematic of System - Compandors

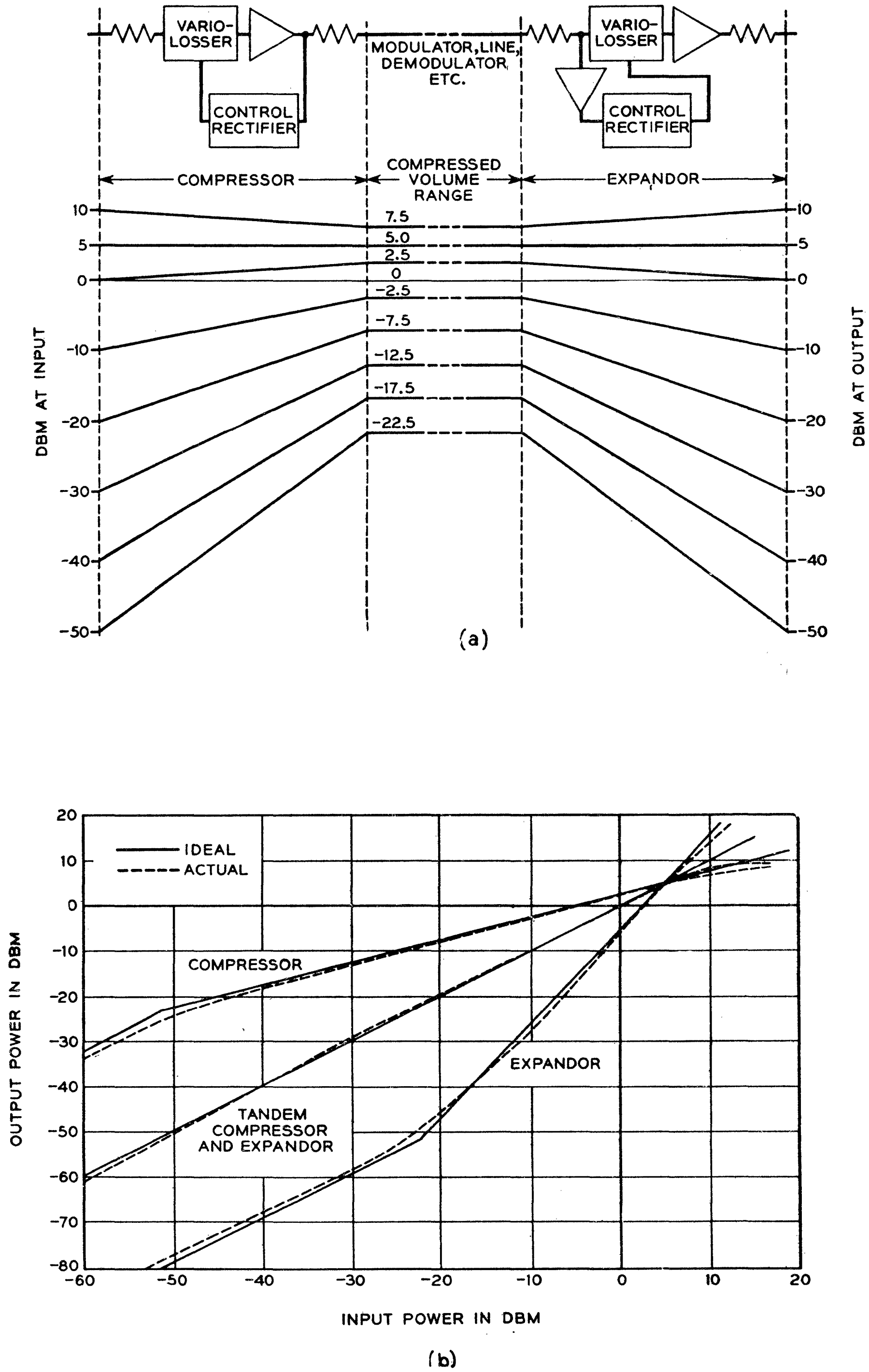
Figure 15-5

To find the appropriate value of Δ_C, one should have curves like those of Figure 12-2 computed for the expected value of τ. Lacking such curves, a good approximation would be to use, for a 100 channel system with Tasi, the Δ_C given by Figure 12-2 for a 300 or 400 channel system.

Compandors

The compandors we shall consider are voice-operated devices used to improve the signal-to-interference ratio. The word "compandor" is a contraction of "compressor" and "expandor". As indicated in Figure 15-5, a voice-frequency operated compressor is associated with each channel in the transmitting terminal and serves to compress the range of speech volumes before the signals are modulated and transmitted over the high frequency line. A compensating expandor, also voice-operated, is used with each channel at the receiving terminal to restore the compressed range of speech volumes to their original range.

The level diagram of Figure 15-6a shows the action of the N-system compandors for sine-wave signals of different strengths. The input-output characteristics of the compressor and expandor which produce this level diagram are shown in Figure 15-6b. The time constants are such that we can consider that each device has, for the duration of a syllable, the gain (loss) corresponding to the average power of the talker. Between syllables, the gain (loss) corresponds to the noise power in the channel. The compressor acts as an amplifier of the input speech signals, but the amount of amplification decreases as the speech volume increases. This results in an output from the compressor which, over the compression range, varies one db for every two db change in input signal. At the receiving terminal the expandor performs the opposite function, inserting loss which increases with decreasing speech volumes.



Input-Output Characteristics of N-System Compressors and Expandors and the Resulting Compandor Action

Figure 15-6

Compandors have two important attributes which can be taken advantage of in a system design. As we have seen, the action of the compressor causes the range of speech volumes which must be transmitted by the system to be reduced. We would expect that reducing the range of signal magnitudes to be transmitted would be beneficial in any transmission problem. For example, if our system had satisfactory performance in every respect except for the fact that it was nearly overloaded, adding compandors would permit us to lower levels while still maintaining an adequate signal-to-noise ratio on the line for the weakest volume talker. We would thus increase the margin against overload. Or we might have had enough margin against overload, but too much noise or crosstalk to let low volume talkers come through satisfactorily. Adding compandors would let us lift low volume talkers out of the noise without letting the high volume ones overload the system.

Equally important is the expandor action at the receiving terminal. By inserting a loss equal to the gain of the compressor, the expandor acts to attenuate, by the same amount, any noise or crosstalk which comes into the channel from the high frequency line. In the absence of speech, Figure 15-6b shows that any noise at the input to the expandor having a power of about -22 dbm or less will be attenuated by about 28 db. When speech is present, noise from the high frequency line will be attenuated, but by an amount less than 28 db depending on the power of the speech signal. The overall result, as we shall see, is to permit a system to operate with considerably more noise at the receiving terminal, in the absence of speech, than could be tolerated if compandors were not used.

The effects of using compandors on the requirements and performance of a single sideband telephone system using AM modulation can best be analyzed by considering that we have a new signal with new values of tolerance to noise, as in the following example.

Assume that we are to compute the effects of applying compandors to a particular broadband repeatered cable transmission system of the sort discussed in previous chapters, equipped, let us say, for 100 channels. Feedback, if any, and "levels" at the grids of the output stages of the line repeaters are assumed flat vs frequency. We shall confine our discussion to the noise, intermodulation, and overload effects arising in the high frequency line, ignoring terminal problems. Particular questions we might answer are: How do a), the margin against overload, b), the thermal noise, and c), the modulation noise, change when compandors are applied to each channel?

For convenience, we shall postulate that we have, in the transmitting and receiving terminals, points at which we can observe and define the nature of compressed signal and the interferences just as we previously could at similarly postulated zero level points. Let us define these observation points (A and B in Figure 15-5) by saying that before the application of compandors they were zero level points, and that so far we have made no other change in system levels except adding compandors.

We immediately notice, in passing, that our old definition of transmission level - for example, at line repeater output - has become useless in a compandored system. Formerly we could ask "What is the ratio, in db, of the power applied in a given channel at transmitting switchboard to the resulting signal power at repeater output?" The answer was a definite number of db, before compandors - now the db ratio depends on the applied power, since transmission has deliberately been made non-linear.

How will the universe of talker signals at point A compare to the universe we would have observed before application of compressors? In terms of a distribution such as that shown in Figure 12-1 for volumes of talkers at zero level we see that the standard deviation will be reduced from 5 db to 2.5 db by virtue of the slope of the compressor characteristic shown in Figure 15-6. (The 2:1 portion of the input-output characteristic of the compressor covers the entire range of talker powers for all probable values of level at compressor input.) The distribution will still be a normal one. Its median value will shift, however. If the compressor input is a zero level point, and the power of a V_o (median) talker is -13.9 dbm at zero level (corresponding to V_o= -12.5 vu), then the new median talker will have at point A a power of about -4.5 dbm, obtained from Figure 15-6b. The difference between the average power talker and the median for this new broadband signal would still be $.115\ \sigma^2$ db. Thus, if the power of the median talker is -4.5 dbm (true only if input to compressor is a zero level point), the average power talker will be $.115\ (2.5)^2$ db higher, or .72 db above -4.5 dbm, which is -3.8 dbm. This is 7.2 db greater than the average power talker before compressors were applied. (It would not be valid to arrive at the power of the average power talker of the compressed distribution by asking what the compressor output would be for an input power corresponding to the average power talker of the uncompressed distribution.)

This, then, is the new broadband signal which is to be transmitted over the high frequency line. There it will generate intermodulation products, be contaminated with thermal noise, and bear some relation to the overload point of the repeaters. In computing these effects, let us first describe them in terms of the high frequency line up to point "B", ignoring the effects of the expandors for the present.

Load Carrying Capacity Requirements

We shall couch our discussion of overload in terms of the question: what sine wave power at point B must we carry without overloading in order to meet the usual requirements for overload expectation?

We postulate that compressors will operate at syllabic rates, but will not compress instantaneous voltage excursions. The new value of "P_S" can still be specified, therefore, in terms of an rms power plus a peak factor. As we have seen, however, the value of the multi-channel load factor, Δ_C, is a function of σ. Values of Δ_C as a function of number of channels equipped are given in Table 15-3 for $\sigma = 2.5$ vu. The table also gives the difference between Δ_C for $\sigma = 5$ and Δ_C for $\sigma = 2.5$. Since we can find the power, at A or B, of the average power talker of our new distribution, the computation of total average power presents no difficulties. As we have seen, the average power talker after compressors is 7.2 db higher power than before if compressor input is a zero level. The change in Δ_C partly counteracts this 7.2 db increase, giving a net increase of about 5 db for a 100 channel system. It is therefore usually necessary to lower line levels by appropriate pads in the transmitting terminal.

Table 15-3

Number of Channels	Δ_c for $\sigma = 2.5$	Decrease in Δ_c for $\sigma = 2.5$ vs 5.0
10	18.4 db	3.4 db
20	16.9	2.9
50	14.6	2.5
100	13.0	2.2
200	11.9	1.8
500	11.0	1.2
1000	10.4	1.0

For the present, let us ignore the fact that we shall probably have to change levels on the high frequency line.* When we do so, we shall, as always, affect the overload, modulation and thermal noise performance of the system, but this is now a familar procedure. These level adjustments can be deferred until we have completed our analysis of the new broadband signal, the system performance, and sensitivity to interferences.

Noise Contributors

The thermal noise at point B can be computed by methods developed in Chapter 7; the fact that signals have been altered has no effect, of course. Modulation noise is another matter. It was shown in Chapter 12 that modulation noise due to the important x-type products could be computed from the formula:

$$W_X = T_a - s_x + H_x + 10 \log N_x - \rho_x + \eta_x V_{op}.^{**} \qquad (15\text{-}1)$$

where, W_x is the intermodulation noise, from talkers, in dba at zero level

T_a is the annoyance, in dba, caused by a zero vu talker

s_x measures the relative annoying effect of different types of products

H_x is system linearity expressed as ratio of 2f or 3f power to fundamental (f) power when fundamental power is one milliwatt at zero level (3f being evaluated as if it added in phase from repeater to repeater).

N_x is the probable number of products of type x in the disturbed channel (busy hour)

ρ_x is a correction allowing for the fact that the annoying peak values of products do not usually coincide

η_x is order of x-type product (for example,

$$\eta_x = 3 \text{ if } x = \alpha+\beta-\gamma)$$

---

*It will be recalled that we are explicitly ignoring terminal problems. In actual practice, we would put in pads at the transmitting terminal, and compensating gain in the receiving terminal at locations chosen so as to optimize the performance of the frequency multiplexing modulators and demodulators - for example, pads might be inserted at compressor output. Additional gain and loss blocks would then be used to optimize high frequency line performance.

**Here, and in the discussion which follows, it is assumed that $.115\ (\lambda_x-\eta_x)\ \sigma^2 = 0$, which is true for A+B, A-B and A+B-C products, but not for 2A-B.

Considering the system between points A and B, and recalling that these are equal level points, we see that this formula for W_x at point B is still valid with our new talker distribution at A and our re-evaluation consists of substituting a new value V_{op}. (V_{op} =-2.4 vu vs the old value of -9.6 vu). H_x as measured between A and B (which were zero level points before we compandored the system) has not changed, altho it will when we eventually change levels on the high frequency line. From computations of W_x, and the values found for thermal noise, we can compute the total interference at point B as a function of carrier frequency.

To compare this interference with the subscribers' tolerance, we can either translate our zero level requirements to point "B", or carry the total noise from B to zero level. In either case we must ask what relationship will exist between noise at B and at zero level with expandors in circuit. From Figure 15-6 we see that the noise between syllables will not be expanded unless it exceeds -22 dbm in power at expandor input. Intersyllabic noise will therefore suffer a loss of 28 db in passing thru the expandor. Not all of this loss can be taken as an advantage, however, since the expandor is not instantaneous, and because noise during syllables will not be attenuated relative to signal. Listening tests indicate that a 5 db allowance should be made for these effects, giving a net compandor advantage of 23 db. If we express our noise requirements in terms of point B in a 4000 mile system, then, we would allow +61 dba where before compandors the tolerable total was +38 dba.

To conclude this discussion, let us consider an illustrative example.

Illustrative Example

At the zero level point in the receiving terminal of a 100 channel system we find the following values of top-channel noise from the high frequency line:

Thermal	+34 dba
Second order noise	+30 dba
Third order noise	+15 dba

The system levels and repeater spacing have been dictated by overload considerations. Repeater spacing is to remain unchanged.

Suppose now that we apply compandors to the system, and then readjust levels to maximize the margin against overload, subject to the restriction that total noise in dba heard by the subscriber shall not be increased. What pads or amplifiers should be inserted between the high frequency line and the A and B points in the terminals (see Figure 15-5)? What will be the new margin against overload?

Solution: The total noise at zero level is 35.5 dba. After compressors are applied at the transmitting end, the noise at "B" at receiving terminal will change as follows, (before any steps are taken to readjust levels):

	Change in Noise at B
Thermal	No Change
Second Order	2 x 7.2 = 14.4 db
Third Order	3 x 7.2 = 21.6 db

The modulation noise changes are, of course, deduced from the 7.2 db change in V_{op}.

It has been stated that the effect of the expandors will be to increase tolerance to noise at B by 23 db. Still leaving levels unchanged, then, we find after compandors:

Original Noise	Compressor Effect	Expandor Effect	Noise at Zero Level
+ 34 dba	0	-23	+11 dba
+ 30 dba	+14.4	-23	+21.4
+ 15 dba	+21.6	-23	+13.6

We have not been asked to optimize the noise, in the sense of setting levels for optimum S/N ratio, but rather to make levels on the high frequency line as low as possible without engendering more than +35.5 dba at zero level. In terms of point B, this is +58.5 dba. (This approaches the noise power which would start to operate the expandor!) Inspecting the last column of the table above, we see that we can drop levels on the line by 24.5 db before thermal noise exceeds +58.5 dba at B. Since we do this by changing G_{OL} (see Figure 15-5), signal magnitudes at A and B will remain unchanged. Therefore thermal noise at B will change db for db as we change levels on the high frequency line, just as it formerly did at a zero level point.

After a 24.5 db decrease in magnitudes of signals on the line, modulation noise will be negligible.

How much have we improved our margin against overload? Adding compandors will cost us about 5 db (7.2 db increase in average power, but a better 100 channel peak factor by 2.2 db). Our net profit, then, is 24.5 - 5 or 19.5 db.

References

1 - "Cross-Modulation in Multi-Channel Amplifiers," Bell System Monograph B-1252. W. R. Bennett.

2 - A. C. Dickieson, D. Mitchell, C. W. Carter, Jr. - Application of Compandors to Telephone Circuits, Bell System Monograph B-1512.

INDEX

Where discussion of an item extends over several pages, reference is to the initial page.

A Partial List of Symbols Used in This Text

The following list is restricted to those symbols which have been given a special meaning and used in a number of places in the text. No attempt is made to list such common places as ω for frequency, ω_c for carrier frequency, etc.

Where several references are given, the first is to the definition; subsequent references list particularly important uses of the symbol.