NASA CR-165,774

NASA Contractor Report 165774

NASA-CR-165774
19820021635

SYSTEM DATA COMMUNICATION STRUCTURES FOR ACTIVE-CONTROL TRANSPORT AIRCRAFT Volume I

A. L. Hopkins, J. H. Martin, L. D. Brock, D. G. Jansson,
S. Serben, T. B. Smith, and L. D. Hanley

THE CHARLES STARK DRAPER LABORATORY, INC.
555 Technology Square
Cambridge, Massachusetts 02139

CONTRACT NAS1-15359
JUNE 1981

LIBRARY COPY

AUG 3 1982

LANGLEY RESEARCH CENTER
LIBRARY, NASA
HAMPTON, VIRGINIA

NASA
National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23665

NASA Contractor Report 165774

SYSTEM DATA COMMUNICATION STRUCTURES FOR ACTIVE-CONTROL TRANSPORT AIRCRAFT Volume I

A. L. Hopkins, J. H. Martin, L. D. Brock, D. G. Jansson,
S. Serben, T. B. Smith, and L. D. Hanley

THE CHARLES STARK DRAPER LABORATORY, INC.
555 Technology Square
Cambridge, Massachusetts 02139

CONTRACT NAS1-15359
JUNE 1981

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23665

N82-29510

PREFACE

This is the first volume of a two-volume report covering work performed in the period between June, 1978, and April, 1981, on a project entitled "Definition and Analysis of Systems Data Communication Structures." This project was sponsored by the National Aeronautics and Space Administration, Langley Research Center, Hampton, Virginia. The Technical Contract Monitor was Mr. J. Larry Spencer.

This volume is primarily concerned with communication methodology, while the second volume treats communication issues at the aircraft system level.

The authors would like to express their gratitude to the personnel of NASA Langley who, along with Mr. Spencer, have made significant technical contributions to this work, especially Messrs. Brian Lupton and Nicholas Murray. Thanks are due also to Mr. Billy Dove, whose foresight and confidence made this project possible.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-165773		

4. Title and Subtitle	5. Report Date
SYSTEM DATA COMMUNICATION STRUCTURES FOR ACTION-CONTROL TRANSPORT AIRCRAFT - VOLUME I	June 1981
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
A.L. Hopkins, J.H. Martin, L.D. Brock, D.G. Janson, S. Serben, T.B. Smith, and L.D. Hanley	R-1469
	10. Work Unit No.

9. Performing Organization Name and Address	11. Contract or Grant No.
The Charles Stark Draper Laboratory, Inc. 555 Technology Square Cambridge, Massachusetts 02139	NAS1-15359
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

Langley Technical Monitor: J. Larry Spencer
Final Report

16. Abstract

This two volume report addresses the problem of data and power distribution in advanced transport aircraft in support of the NASA Energy Efficient Transport Program. Advanced aircraft design concepts are employing active control techniques to achieve significant increases in aircraft performance and energy efficiency. The concepts depend, however, on the availability of control mechanisms, with their supporting communication and power systems, that can perform flight-crucial functions continuously. Traditional methods are likely to be inadequate for these requirements. The objective of this study is to develop the technology that will meet the challenge.

Volume I addresses several specific technology issues. Candidate data communication techniques are identified, including dedicated links, local buses, broadcast buses, multiplex buses, and mesh networks. The design methodology for mesh networks is then discussed, including network topology and node architecture. Several concepts of power distribution are reviewed, including current limiting and mesh networks for power. The technology issues of packaging, transmission media, and lightning are addressed, and, finally, the analysis tools developed to aid in the communication design process are described. There are special tools to analyze the reliability and connectivity of networks and more general reliability analysis tools for all types of systems.

Volume II directly addresses the application of communication structures to advanced transport aircraft.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
data communication multiplexing communication networks reliability analysis avionics	Unclassified - Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	226	

For sale by the National Technical Information Service, Springfield, Virginia 22161

TABLE OF CONTENTS

	Page
CHAPTER 1 - INTRODUCTION	1
1.1 Background	1
1.2 Objectives of the Study	2
1.3 Summary of Existing and Emerging Methods	3
1.4 Summary of Problems	6
1.5 Baseline Assumptions for the Study	10
CHAPTER 2 - DATA TRANSMISSION TECHNIQUES	12
2.1 Factors Underlying System Interconnection Concepts	12
2.2 Introduction to Signal Transmission Methods	22
2.3 Dedicated Links	29
2.4 Local Buses	33
2.5 Broadcast Buses	34
2.6 Standard Multiplex Buses	36
2.7 Variants of Multiplex Buses	44
2.8 Mesh Networks	46
CHAPTER 3 - MESH NETWORK DESIGN	53
3.1 Review of the Network Concept	53
3.2 Network Management Principles	56
3.3 Network Topology	70
3.4 Subscriber Assignments	93
3.5 Issues Concerning Node Architecture	97
3.6 Network Design Summary	104
CHAPTER 4 - POWER DISTRIBUTION	107
4.1 Hydraulic Power Distribution	108
4.2 Electric Power Distribution	109
4.3 The Substation Approach	110
4.4 Current Limiting	111
4.5 Mesh Networks for Power	115
4.6 Power Distribution Summary	120
CHAPTER 5 - TECHNOLOGY ISSUES	121
5.1 Integrated Circuits	121
5.2 Packaging of Dispersed Electronics	125
5.3 Transmission Media	133

		Page
5.4	Lightning Effects and Countermeasures	143
5.5	Software	149
5.6	Terminal Redundancy Design	150
5.7	A 1553-Compatible Network Node	151
CHAPTER 6 -	Analysis and Modeling Discussions	153
6.1	Network Reliability	153
6.2	Network Dispatch Probability	184
6.3	Network Connectivity	186
6.4	Bus Reliability and Dispatch Probability	196
6.5	Remote Power Control Reliability	199
6.6	Power Network Using Current Limiters	203
6.7	Reliability Analysis Tool	203
REFERENCES		226

LIST OF TABLES

		Page
Table 2.1.1-1	Levels of System Interconnections and Connections	13
Table 2.1.1-2	Individual Interconnections Cost as a Function of the Level of Interconnection	13
Table 2.2-1	Summary of Attributes of Communication Methods	24
Table 4.5-1	Power Overheads with Magnetic Limiter	119
Table 4.5-2	Percent Overheads for Various Magnetic Limiter Cases	119
Table 5.1-1	Reliability Equation Parameters	124
Table 5.2-1	Package Densities	132
Table 5.3-1	Mechanical and Physical Properties of Insulating Materials	137
Table 5.3-2	Chemical Properties of Insulation Materials	138
Table 5.3-3	Electrical Properties of Insulation Materials	139
Table 6.1-1	Probability of Getting Specific Isolation Patterns Around a Particular Node in a 3-Port Node Network	175
Table 6.1-2	Probability of Getting Specific Isolation Patterns Around a Particular Node in a 4-Port Node Network	176
Table 6.4-1	FMEA of 1553 Bus	199
Table 6.5-1	Remote Power Model Evaluation	202

LIST OF ILLUSTRATIONS

Figure		Page
2.3-1	Dedicated Links Per Signal	30
2.3-2	Dedicated Signal Interfaces	31
2.3-3	Dedicated Links Per Source - Destination Pair	32
2.4-1	Examples of Controller and Bus Arrangements	35
2.5-1	Broadcast Buses	37
2.6-1	Multplex Bus	41
2.7-1	Hierarchical Multiplex Buses	45
2.8-1	Mesh Networks	47
2.8-2	Network with Virtual Bus Shown	49
2.8-3	Alternate Virtual Bus Structure	49
2.8-4	Virtual Bus with "Y" Constructs	52
2.8-5	Multibus Network	52
2.8-6	Reconfigured Multibus Networks	52
3.1-1	Network Examples	54
3.2.1-1	Portion of a Network	57
3.2.1-2	Gateman Circuits in Node A	58
3.2.1-3	Growth Process	60
3.2.1-4	Growth Algorithm	62
3.2.1-5	Growth Pattern	63
3.3-1	Connectivity Example	71
3.3.1-1	Network with Connectivity of Two	72
3.3.1-2	Network with Connectivity of Three	72
3.3.1-3	Examples of Network Fragments	74
3.3.2-1	Multiple Path Examples	76
3.3.2-2	Dedicated Paths	77
3.3.2-3	Six-Port Node	78
3.3.3-1	Regular Geometries	80
3.3.3-2	Polyhedral Regular Geometries	81
3.3.3-3	Three-Dimensional Array	82
3.3.3-4	Regular Toroidal Pattern	83
3.3.3-5	Hexagonal Toroid	84
3.3.3-6	Private Channel Recovery - Four-Port Nodes	86
3.3.3-7	Private Channel Recovery - Three-Port Nodes	87
3.3.4-1	Injured Net Example	88
3.3.4-2	Growth Pattern from Central Node - Three Ports	90
3.3.4-3	Growth Pattern from Central Node - Four Ports	91
3.3.4-4	Semiregular Tree Network	92

Figure		Page
3.3.4-5	Variant of Semiregular Tree Network	92
3.3.4-6	Connections to a Toroidal Network	94
3.3.4-7	Inter-Toroid Connections	95
3.4-1	Force-Voting Actuator Node Assignments	96
3.4-2	Toroidal Connection of Triplex Subscribers	98
3.5.1-1	Shrinking Pulses	100
3.5.1-2	Leading Edge Pulse Translation	100
3.5.1-3	Leading Edge Pulses Repeated	100
4.4-1	Saturable-Reactor Current Limiter	113
4.4-2	Typical Saturation Characteristic for a S-R Current Limiter	114
4.5-1	Current Distribution in a Power Network	116
4.5-2	Switched Links	117
4.5-3	Current-Limited Links	117
4.5-4	One-Way Current Limiter Links	117
5.2-1	Rack Mounted Equipment	128
5.2-2	Remote Equipment	130
5.2-3	Node Contents	131
5.4-1	Diagrammatic Representation of Lightning Model	145
5.4-2	Considerations Regarding Circuit Design	148
5.7-1	1553 Network Node Concept	152
6.1-1	Schematic of Data Communication Node and Link	155
6.1-2	Node Isolation in Irregular Networks	156
6.1-3	Large Regular Network	158
6.1-4	Regular Closed Networks	159
6.1-5	Isolation of Innocent Nodes by Three-Link Breaks	161
6.1-6	Isolation of Innocent Nodes by Four-Link Breaks	163
6.1-7	Isolation of Innocent Nodes by Five-Link Breaks	164
6.1-8	Isolation of Innocent Nodes by Six-Link Breaks	165
6.1-9	Isolation of an Innocent Node by Four-Link Breaks	168
6.1-10	Isolation of an Innocent Node by Five-Link Breaks	169
6.1-11	Isolation of an Innocent Node by Six-Link Breaks	170
6.1-12	Minimum Spacing for Node and Link Failures	172
6.1-13	All Purpose Eight-Link Break Patterns Which Will Isolate a Given Node	177
6.1-14	Probability for Innocent Node Isolation for a 36 3-Port Node Network	180
6.1-15	Probability for Innocent Node Isolation for a 36 4-Port Node Network	181

Figure		Page
6.1-16	Failure Simulation Algorithm	182
6.1-17	Link vs. Node Failures	183
6.2-1	Buffer Zones	185
6.2-2	Plot of $1 - P_D$	187
6.3.2-1	Network Failure Program	189
6.3.2-2	Connectivity and Level Tables	191
6.3.2-3	Move Table	192
6.3.4-1	96-Node Network	197
6.3.4-2	66-Node Network	198
6.5-1	Remote Power Distribution Model	200
6.7-1	Graphical Equivalent of Basic Equation	205
6.7-2	Schematic Diagram of the Example System	211
6.7-3	Network Interconnections	211
6.7-4	Partitioning of System Elements	213
6.7-5	Equation Diagram Stage 1	215
6.7-6	Equation Diagram Stage 2	216
6.7-7	Equation Diagram Stage 3	217
6.7-8	Equation Diagram Stage 4	219
6.7-9	Equation Diagram Stage 5	220
6.7-10	Equation Diagram Stage 6	221
6.7-10	(Continued)	222
6.7-10	(Concluded)	223

CHAPTER 1

INTRODUCTION

1.1 Background

Airplane flight depends on the distribution of power and the communication of information within the vehicle. As aircraft systems increase in complexity for reasons of performance and safety, so also must the supporting data and power systems. The proliferation of signal and power wires carries penalties of excessive weight, installation cost, circuit cost, and certain forms of vulnerability. Alternatives, such as multiplex systems and remote power control, seem to address the proliferation problem, while presenting hazards of their own.

Human eyes, hands, and muscles were the original principal information and power elements in airplanes, and are still important, though not always sufficient. They are increasingly supplemented by hydraulic, electric, and electronic means. Pilots and passengers rely on hidden systems vulnerable to flaws and stress. The more sophisticated the systems become, the more fragile they seem to be. How, then, shall power distribution and data communication be handled in airplanes projected for the future in which lapses of correct control may not exceed several milliseconds in duration? On one hand, the proliferation engendered by extrapolation of present practice seems a reasonable price to pay for the preservation of the technology evolved by the aircraft industry over several decades of time. On the other hand, it is not clear that such extrapolations could ever meet present and future safety requirements in future airplanes that depend on constant automatic control for their moment-to-moment survival. It rather appears that multiplexing and remote power control will be the less hazardous approaches in cases where comprehensive redundancy is used to achieve extreme levels of reliability in complex systems.

The issues involved in on-board data communication and power distribution are of major importance in the development of technology

for future commercial transport airplanes, where energy efficiency is an overriding goal, subject to various requirements of performance, economy, and safety. The National Aeronautics and Space Administration is responsible for a research program entitled "Aircraft for Energy Efficiency - Energy-Efficient Transport," or ACEE-EET. One of the facets of this program, called ACT, Active Controls Technology, undertakes to develop the necessary tools with which to design full-time flight-crucial* controls. The scope of ACT extends to all aspects of control systems, including sensors, actuators, algorithms, computers, and, as described in this report, data communication and power distribution.

This report documents an investigation sponsored under ACT by the NASA Langley Research Center, primarily focussing on data communication structures, and secondarily treating power distribution structures. The title of the project is Systems Data Communications Structures.

1.2 Objectives of the Study

In brief, the objectives of the study can be summarized as follows:

1. Identify a common set of requirements
2. Define candidate data communication structures
3. Define candidate power distribution structures
4. Develop design theory and analysis tools as required
5. Generate tradeoff data
6. Investigate circuit technology issues.

1. Because much of the study is concerned with comparisons among various methods, a common set of requirements was generated to be used as a baseline for such comparisons. Some of the requirements are firm. Many are more or less casual extrapolations of present practice. Others are predicated on assumptions stemming from the Active Control Transport research program at NASA. These requirements are described in Volume 2 of this report.

---

*Flight crucial - the highest level of criticality: loss of function is catastrophic.

2. The candidate data communication structures can be classified into relatively few categories, i.e. dedicated, broadcast, and two-way multiplex. Each has its variants, depending on technology and function. The airplane data communication structures of today all consist of a variety of substructures with the result that systems are quite inhomogeneous. The evolutionary trend appears to be toward increasingly homogeneous systems and data communication structures.

3. Candidates for power distribution include various ways of utilizing remote-controlled electric circuit breakers and/or current limiting devices. The object is to create a system that can tolerate short circuits as well as open circuits. Hydraulic power was not treated, since there seems to be no significant alternative to present redundant hydraulic power distribution structures, other than abandoning hydraulics in favor of electrical actuators.

4. Design theory and analysis tool development was a requisite for the comparative study of various architectures and techniques. This was particularly so for networked forms of communication and distribution, for which little or no prior theory existed.

5. Tradeoff data has been generated for a number of different data communication structures that have suggested themselves for various phases in the evolution toward fully flight-crucial systems. It is presented in Volume 2.

6. Circuit technology issues underlie most of the evolutionary trends of communication and distribution architectures. Also, as technology itself evolves, the tradeoffs change. The purpose of evaluating technology here is to support the identification and evaluation of the various structures for which comparisons are made.

1.3 Summary of Existing and Emerging Methods

Most data communication on commercial transport airplanes today is accomplished using dedicated wires, one wire per signal, with additional wires where a signal has multiple destinations. Most of the signals involved are discrete or analog, including signals from DC analog devices, linear variable differential transformers (LVDT's),

synchros, and frequency analog devices.

Some of today's airplanes have used digital broadcast buses (ARINC 429), and this form of communication promises to be substantially more prevalent in the near future. Experience so far has been that the noise tolerance is much greater than that of analog signal transmission.

Contemporary power distribution is hierarchical, using redundant main buses plus a variety of subsidiary buses. Breakers are centralized in cockpits with a few exceptions. Protection is afforded by breakers and bus shedding. Batteries are used to provide essential power when generation capacity is lost, but they are viewed as hazardous cargo, and are used in as small numbers and sizes as possible. Problems arise from failures of generators, diodes, breakers, and so forth, as well as from wiring hazards.

Wiring is cabled for ease of installation and mechanical support. Its location and routing is one of the lowest priority considerations in airframe design, subject only to safety and validation considerations. For the most part this presents no problem, but there are occasional spots where wiring becomes awkward, such as in the wings and tail, and at the instrument panels. The major cost factors are installation, termination, and connectors. Cables are partitioned into sections for installation. Each break in the cable requires terminations and a connector pair. More breaks mean easier installation, but higher termination and connector costs. Wiring can be thought of as costing about one dollar per conductor-foot on the average. Total wire lengths for jet transports run from about a hundred thousand feet to somewhat under a million feet.

Wiring hazards exist in many forms. Most of the problems stem from environmental stress. Vibration is a major problem in certain locations such as engines, control surfaces, and landing gear. Damage can result from flexure, insulation cold flow, and/or abrasion. (Strengthening insulation can sometimes result in increasing abrasion damage, rather than decreasing it.) Corrosion from chemicals and moisture in the atmosphere is another hazard. Others are handling and repair, high temperature, and manufacturing defects.

In addition to the hazards listed above, airplanes are exposed to a number of hazards that can cause failure in a number of places at the same time. One such hazard is lightning strike. Another is damage, such as may result from engine burst, bird strikes, structural

failure, or terrorism. These events are relatively rare, but not so rare as to be classified as being "highly improbable."

Airplanes are designed to tolerate the effects of these hazards insofar as they can reasonably be tolerated. Philosophies of functional separation and channel separation are applied, with special treatment accorded to certain paths, signals, and functions.

The present generation of airplanes, such as the Boeing 767, are incorporating the new ARINC 700 series of standard avionics, which rely heavily on digital communication, notably the ARINC 429 broadcast bus. Digital transmission has intrinsic properties making multiplex use of transmission facilities little or no more costly than simplex use. Thus broadcast buses have the property of reducing the number of signal paths needed as compared with non-multiplex analog systems. Further reduction can be made by multidrop runs to multiple destinations, although there are limits to how much this can be done. Except for the use of digital broadcast buses, the data communication and power distribution in present generation airplanes is very little different from those of prior generations.

In the case of military aircraft, efforts have been made to establish a new data communication standard employing a form of multiplex transmission rather more ambitious than that of broadcast buses. The foremost example of such a standard is the MIL-STD-1553 multiplex bus, now in its B revision. This standard has evolved from a number of earlier versions proposed by different design teams, one of which was used to control electric power distribution in the B-1 airplane (E-MUX). This is a two-way form of busing. Each bus has numerous transmitters and receivers. Broadcast buses, by contrast, have one transmitter with multiple receivers. Two-way buses save substantially on wire as compared with broadcast buses, but they are vulnerable to a greater variety of hazards, and have not yet been used for any flight-critical functions. Variants of two-way buses may, however, be considered as candidates for future airplanes in flight-crucial roles.

Military aircraft power distribution systems have also been the subject of recent research and development. Remote solid-state breakers can be controlled by digital commands communicated over multiplex systems. Direct current is proposed to replace alternating current for generation and distribution.

Looking more to the future, one can anticipate increasing distribution and dispersion of electronics for purposes of multiplexed

data communication and remotely-controlled power distribution. Sensors and effectors are tending to become "smart," that is, to incorporate embedded digital computer control. This, along with the lure of fiber optics, creates an environment in which miniature electronics will crop up in diverse places. This will increase the burdens to be borne by data communication and power distribution, but it can also provide the tools with which to support them.

1.4 Summary of the Problem

Redundant flight control systems in some contemporary transport aircraft are given full authority for the order of 10^{-2} hours during autoland. Active control technology for future control-configured fly-by-wire airplanes will require full authority capability for ten hours, some three orders of magnitude beyond current practice. Those orders of magnitude may be achieved with moderate redundancy levels only if malfunctions can be contained, and hence prevented from propagating unchecked through the system. Surviving elements must then be able to be accessed, configured, and applied to the control function.

The problems of containing malfunctions and maintaining access to surviving elements fall heavily upon the structures which carry data signals and power among the various dispersed elements of the system. One might tend to think of redundant computers, sensors, and effectors as being the only significant constituents of fault-tolerant flight control systems, and perhaps think of system interconnections in the same light as ordinary componentry and packaging. In fact, however, system interconnection must be one of the principal architectural considerations in fault-tolerant systems. As such it presents serious challenges to system architecture and technology alike.

The objective of this program was to study interconnection technology for integrated, fault-tolerant aircraft electronic systems, encompassing flight control, guidance, and navigation, as well as flight management, monitoring, surveillance, and support. The study hypothesized an evolutionary development, through several aircraft generations, toward a full-time flight-crucial system with no external backup.

The motivation for a full-time flight-crucial system is simply that this affords the maximum latitude for the airplane architecture to capitalize on active control. The question is not whether full-time active control is necessary. It is rather a question of what benefit

it could be if it were made possible. Fuel economy can be served in several ways through control-configured vehicle (CCV) design. Other potential economies exist in traffic control routing strategies which depend on precise guidance. Workload relief and comprehensive contingency management capability have the potential to enhance safety. The list could grow indefinitely, some items requiring flight-crucial control, others requiring full-authority command, and still others requiring wholly-integrated management. In most or all of these cases, the prevention of system malfunction would be critical or crucial.

The motivation for assuming the absence of external backup forces the assumption of stringent requirements, and therefore leads one to seek the limits in fault-tolerant technology. It does not imply the belief that this is the inevitable outcome. It has its legitimate basis, however, in the sense that present-day system techniques do not extrapolate to a full-time flight-crucial capability, and that all contingencies for which a backup might be employed must be addressed by the redundant primary system.

The interconnection structure of an airplane must take account of the fact that certain sensors and effectors must be located in specific dispersed places, where the environment may well be extremely inhospitable. Contemporary systems address this problem by concentrating electronics in equipment bays in the fuselage, and locating mostly passive devices elsewhere. The interconnections between electronics and sensor/effector components is by some combination of dedicated wire, hydraulic, and pneumatic paths. Present systems are not full-time flight-crucial; and, accordingly, redundant channels are located in separate boxes in a common bay. In flight-crucial systems, damage tolerance considerations will probably require that separate redundant elements be located so that a single damage event will not destroy more than one such element. These systems are therefore visualized as being partitioned so as to occupy different bays for the sake of damage tolerance as well as for communication economy.

Data communication between one electronic element and another can usually be accomplished economically through the use of multiplex buses. Between an electronic element and a passive device, however, dedicated interconnections are required. Insofar as wire length and weight are concerned, it would be preferable to locate electronics near to passive devices to minimize dedicated paths at the expense of multiplex paths. That is, it takes less wire to disperse electronics than

it does to concentrate it. If wire length were the only consideration, all systems would henceforth be dispersed.

The extreme case of dispersion for damage tolerance and wire economy at the same time is a hypothetical "fully dispersed" system, in which each sensor and each effector contains one or more embedded electronic "microbays," housing the dedicated information and power handling devices for the specific sensor or effector component. A future transport airplane could contain the order of one hundred or more such "microbays." Aside from the revolutionary notions involved, the principal design risk stems from the possibility of high failure rates, and therefore high maintenance costs, occasioned by the hostile environments of many microbays. This is, nevertheless, an appropriate, albeit extreme, case for our study.

In between the contemporary and the extreme cases lie a number of alternatives ranging from two or three bays to a quasi-dispersed approach with intermediate environmental conditions. In each of these cases, the "bay" would have the function of providing a containment boundary for externally-induced malfunctions such as damage. In most of these cases, the bay must be designed to degrade gracefully when its contents, i.e. "boxes," sustain random failures. The fewer "bays" there are, the more this is true. In a two-bay structure, for example, there must be a highly dependable set of containment boundaries internal to each bay, since it will be necessary to tolerate some two to four random box faults without losing the services of "innocent" boxes in that bay. Only in the fully-dispersed extreme may the "bay" (a "microbay" in this case) be abandoned after a single random failure.

The data communication and power distribution requirements are primarily to provide "continuous" service in and between bays, and to and from all dispersed elements outside of bays. Service must be available despite the existence of any probable fault condition, and lapses in service due to recovery actions must be brief enough so as not to impinge on active control. A probable fault condition, for purposes of this argument, can be thought of as any fault condition whose probability exceeds 10^{-9} in any hour for a flight of ten hours, corresponding to an emerging FAA guideline.

The kinds of faults that must be considered are of three main classes: random faults in response to normal flaws and stresses, induced faults in response to environmental phenomena, and design lapses in hardware, firmware, and software. The manifestations, or

symptoms, of such faults may be intermittent or permanent. They may produce inconsistencies that implicate innocent entities. They may effect more than one entity at a time owing to correlation or lack of containment. They may also have no manifestation at all, which is known as fault latency. A latent fault is not necessarily harmless, because it can possibly team up with a subsequent fault to exceed the system's capacity to recover.

Communication and power are areas where it is particularly difficult to contain faults, owing to the widespread sharing of resources in these areas. Dedicated data communication links go far in the direction of fault containment, but they still present a large cross-section to damage hazards. Multiplex data communication, instead, presents opportunities for malfeasant modules to interfere with data transmissions among unfailed modules as well as presenting a large cross-section to damage. Power systems are vulnerable to over-current or over-voltage, such as may be caused by lightning-induced surges, as well as to short circuits.

Active-control airplanes require redundancy plus a certain degree of basic complexity. The result is potentially complex in terms of hardware and software unless the principle of simplicity can be adhered to. For one thing, the redundancy management of the system should be carried out transparent to, and independent of, the flight control. Otherwise the flight control program becomes burdened with many contingencies, to the extent that it jeopardizes the possibility of validation. For another, if the system is reconfigurable in any sense, it must be designed so as to be able to pass the configuration authority safely from one controller to another. These two considerations, i.e. transparency and multi-controller capability, can have important impacts on the design of communication and power structures.

Finally, since communication and power structures are interdependent with system structures, this study has encompassed avionics architecture at a fairly high level in addition to techniques for signal and power transmission. There is an interplay between the level of technology and the placement of system elements. As airplanes evolve toward fully flight-crucial active control, they are apt to evolve toward dispersed, multiplexed communication structures and remotely managed power structures.

1.5 Baseline Assumptions for the Study

The following assumptions are among those made for most or all of the study.

1. Evolution to the Active-Control Transport Airplane

Although data communication may take on more nearly crucial roles in near-term airplanes, it is unlikely that a single system will be granted a full-time flight-crucial role for some time. The near-term problem and the far-term problem are linked by evolution, and are both within the scope of the study.

2. Single System with Redundant Members

To assume separate, independent primary and backup systems is to evade the issue of redundancy management. Although backups may in fact be employed in actual systems, the primary system's specification should address the full safety requirement. The time required by a pilot to deliberate and switch is too great to be reliable for an active-control airplane.

3. Highly-Integrated System

One of the problems that needs to be dealt with in redundancy management is that of graceful takeover of command. When sensors and effectors fail, alternative sensors and effectors must be accessible to the controller. Traditionally, one failure has implied the loss of an entire redundant channel, which amounts to one quarter, one third, or one half of the subsystem concerned. This approach has been validated for the autoland case of 10^{-2} hour duration, but does not extend to the 10-hour flight-crucial case, where multiple faults must be assumed. Increasing the number of channels would be untenable from a cost standpoint. Instead, systems will have to rely on reconfiguration. This implies that all system elements are able to communicate with one another, in contrast to the channelized approach, where pains are taken to prevent interaction between channels except at carefully designated points, such as actuators.

The realization of such an integrated approach calls for a multiplex form of data communication between reconfigurable elements. Not to do so would impose a requirement for each signal to be separately interfaced to each of the several control sites, with high cabling and interfacing costs. Multiplex data communication is vulnerable to a broader class of faults than dedicated communication, however. Special

measures can and must be taken to contain the effects of faults for system survival.

Dedicated communication is not entirely supplanted in multiplex systems, though it may be more or less important according to the specific system architecture. For example, servo amplifiers may receive multiplexed inputs from flight control computers while having dedicated links to actuator valves and LVDT signals.

4. Fault-Tolerant Computer Control

The reconfiguration of a distributed redundant control system presents an awkward problem for computer hardware and software. Fault-tolerant computers, such as those currently being developed by the NASA Langley Research Center [1,2,3], afford a graceful means of handling the problem of multiple control sites in a multiplex environment. By assuming the existence of a fault-tolerant computer, it is possible to avoid the pitfalls of tailoring the data communications system to the problems of computer redundancy management as opposed to the broader needs of the system.

5. Maintenance Postponement Requirements

Airlines consider it undesirable for an airplane to be nondispatchable due to single faults. A redundant system should therefore be dispatchable despite the existence of faulty elements. If the function of the system is full-time flight-critical, it should be able to tolerate multiple faults, so that the airplane can continue its normal schedule until it arrives at a convenient maintenance base. This can place important constraints on system design, especially for intercontinental carriers. Thus this requirement should be interpreted with an eye toward reasonableness.

6. Damage Probability at a Single Specific Spot May Be Negligible

The probability of damage somewhere aboard an airplane is not negligible, nor is the probability that some part of the communication system or the power system will be affected by damage. The probability of damage to a fault-tolerant computer centrally located in an interior avionics bay will, however, be considered to be negligible.

CHAPTER 2

DATA TRANSMISSION TECHNIQUES

2.1 Factors Underlying System Interconnection Concepts

2.1.1 Traditional Levels of Connections and Interconnections

Systems are generally divided up into functional subsystems which are then further broken down into smaller functional units which are made up of electronic components. Some of these components are themselves quite complex, such as a large scale intergrated circuit or a hybrid circuit.

With each level of the system there is associated at least one means of electrical interconnection. For example, a module might be made up of a printed circuit board, which interconnects simple and complex components which are connected by soldering to it. A group of printed circuit cards might be inserted into connectors to a wire wrap field which interconnects them. Thus a system is made up of levels of interconnections, and these levels are connected to their neighbor levels by means of other connections. Table 2.1.1-1 shows an example of how four levels of system interconnections might be accomplished.

The connections at the first level are simply the ohmic contacts between the evaporated aluminum conductor and the silicon device. The connections at the second level are solder joints between the components and the printed circuit card. (Actually these connections may be more complex, e.g. a packaged semiconductor may have internal wire bonds which are also connections). At the third level, the connections would be solder joints and connector contacts. At the fourth level, they would be wire wraps and connector crimps.

There are some important comparisons to be made about levels of interconnect in a system. Perhaps foremost is the considerable difference in cost of the interconnections and connections at different levels. Table 2.1.1-2 shows a range of costs for various interconnection levels. Second is reliability. A level-four cable

TABLE 2.1.1-1

LEVELS OF SYSTEM INTERCONNECTIONS AND CONNECTIONS

INTERCONNECT LEVEL	INTERCONNECT	TO MAKE	WITH INTERCONNECTIONS	AND CONNECTIONS
1	Silicon Components	LSI	Evaporated and Patterned Aluminum	Ohmic contact to components in chip
2	LSI & other less complex components	Circuit module	Multilayer printed circuit board	Solder
3	Circuit	Subsystems	Wire wrap field	Module connectors
4	Subsystems	System	Cable	Cable connectors

TABLE 2.1.1-2

INDIVIDUAL INTERCONNECTION COST AS A FUNCTION OF THE LEVEL OF INTERCONNECTION

INTERCONNECTION LEVEL	COST PER INTERCONNECTION
Level 1	$ .00001 - $ .0001
Level 2	.01 - .10
Level 3	.10 - 1.00
Level 4	1.00 - 10.00 or more

connection may be connected to a line driver, so that this signal must pass through all the levels of interconnection in the system. This is inevitably a less reliable signal path than one that passes through only the first one, two, or three levels of interconnection. Another area of comparison is in the size and weight of higher level interconnections, which increases dramatically at higher levels.

In order to reduce cost, weight, and size, and to improve reliability, all reasonable efforts need to be made to reduce the number of higher level interconnections in the system.

The interconnections in a system for which the circuits have been specified cannot be reduced in number. However, the number of interconnections which are made at higher levels can be reduced by various means:

1. Increasing the number of components in a module can reduce the number of connections between modules.
2. Combining subsystems in the same box will substitute level-three interconnections for level-four interconnections.
3. Improved partitioning can minimize connections between modules and between subsystems.

In addition, the system designer can often help by changes in the design such as multiplexing signals. He can also select more highly integrated silicon components, which has the effect of making more of the interconnections in the system at the lowest level.

The incentive for minimizing higher level interconnections in a system becomes even more acute in systems which have considerable redundancy. Thus minimization of higher level interconnections becomes an architectural consideration. The telephone industry has always tried to minimize the number of long communication lines as a means of cost reduction. The ability to build transmitters and receivers with silicon integrated circuits for less money and with increasing reliability makes multiplexing techniques a valuable alternative for short distances as well as long.

The achievement of a highly reliable communication system requires techniques which reduce the fourth level of interconnections in the system and yet increase the number of optional paths in the system.

2.1.2 Distance Considerations

The cost of level-four interconnections as shown in Table 2.1.1-2 is unbounded as long as the communication distance is unbounded. When short distances are involved, the cost is dominated by connectors and interface circuits, and the predominant tradeoffs concern serial vs. parallel transmission formats. Bandwidth is seldom a problem over short distances, since channels can usually be added, up to a point, at low cost. Long distance might be defined for practical purposes as a case where channel capacity is expensive. In the MX inertial measurement unit, for example, the information path from the inner sphere to the outer sphere is expensive, though the physical distance is only a fraction of an inch. Disregarding such anomalies, we can reasonably cite distance as the significant parameter affecting channel cost. In the case of high channel cost, bandwidth becomes a treasured resource, and system designers tend to incorporate complex circuitry to maximize bandwidth utilization, thereby minimizing cost, at least as they perceive it.

The transport airplane is a microcosm in which both "short" and "long" distances exist. It is necessary to consider various topologies, protocols, and formats, plus tradeoffs of interface complexity with numbers of channels.

Within a single bay, numerous level-four connections exist among boxes, where the distance is short. Given the state of the art in box connectors and back planes, it hardly matters whether the number of box connections is ten or a hundred, other than, perhaps, the cost of interface circuit boards. If no damage environment or reconfiguration requirement existed, there would be little incentive for multiplexing or for high-bandwidth channels inside a single bay (intra-bay).

Going between a bay and a remote sensor or effector, distances can be substantial, so that an incentive exists, in principle, to use multiplexing insofar as it is feasible. In practice to date, however, the remote environment is relatively harsh, so that dedicated wire is preferred over multiplex-demultiplex electronics. The magnitude of the multiplexing incentive can be appreciated by considering the 800,000-odd conductor-feet and 4000-odd connector pairs in the Boeing 747. Multiplexing has much to offer in the future, and in all probability, advantage will be taken of it for long runs; but the environmental problem must be dealt with first.

Connections between bays are in the medium-to-long category, with numerous signals involved. No harsh environmental problem exists in this case, so that multiplexing is reasonable, and is, in fact, already used, as in ARINC 429 broadcast buses. The magnitude of the distance involved is such that it is reasonable to run several buses of moderate (100 Kilobaud) bandwidth, as opposed to a single bus of high bandwidth. It hardly matters, then, whether the number of broadcast buses leaving a bay is one or ten, other than, perhaps, the cost of the bus interfaces. Truly long-distance psychology applies if a single bus is conceived to serve the entire airplane. The cost of the single channel with suitable redundancy is high enough so that it does seem to matter considerably whether one channel or two are used. High bandwidth therefore becomes a dominant force of the design.

Digital communication makes practical the concept of "store-and-forward," in which signals migrate from source to destination via waypoints where the format, protocol, technology, or almost any other communication parameter may change. Because of the real-time character of avionic systems, one parameter that must not be altered indefinitely is time latency. Store-and-forward is tantamount to the creation of connection levels higher than the fourth. Hierarchical multiplexed systems are based on the notion of store-and-forward, with a limited number of remote multiplex-demultiplex centers interconnected by a serial channel, where each center translates between the multiplex signal form and one or several local forms.

2.1.3 Bandwidth Considerations

Probably the highest bandwidth channel in an airplane is the parallel internal bus of a computer, which can communicate the order of 10^9 bits per second over distances of the order of 1/10 meter. A single bit line of the bus communicates the order of 10^7 bits per second at TTL logic levels. Transmission errors are virtually nonexistent.

Longer signal distances present problems of reflections, attenuation, and interference, the solutions to which tend to reduce the effective channel bandwidth. Reflections emanate from improper terminations and channel impedance changes which are almost impossible to avoid in the extreme. Access to buses can be a particular problem owing to the necessity to place acccess ports at a nonzero distance from the bus.

Attenuation stems from impedance mismatches, as well as series resistance, which may even be inserted purposefully for short circuit protection. The uneven application of attenuation can cause different transmissions on the same bus line to be detected with different amplitudes at one port, while the situation at another port may be different from the first. Bus receivers must be designed with automatic gain control to cope with the resultant dynamic ranges of the signals they see.

Interference protection requires secure signal paths such as twisted shielded pair or coaxial cable. The attenuation due to interfacing with such cable must be compensated by gain, which tends to erode bandwidth.

2.1.4 Cost Factors

The dedicated wire cables in transport aircraft today are expensive to acquire, install, test, and fly. The fixed cost is of the order of hundreds of thousands of dollars per airplane, while the variable cost is elusive. It is equivalent to a few passenger seats in weight, and reflects a maintenance cost due to the replacement of a percent or more of the wiring each year.

Multiplex channels weigh less, and use less wire and fewer connectors. Installation costs will depend on the nature of the installation. Testing is likely to be inexpensive. Interfaces may be expensive, however, for the sake of bandwidth, although advanced integrated circuitry technology makes it possible to build multiplex interfaces within the cost, space, volume, and reliability constraints implied in an aircraft system.

Maintenance will be expensive to the extent that the more vulnerable wiring (e.g. wiring in wheel wells, on landing gear, engines, and in wings and tail) is "special," in the sense that coaxial cable and twisted shielded pair are special, requiring extraordinary care. Experience with coaxial cable in entertainment multiplex systems has been that the multiplex system has a greater life-cycle cost than dedicated wire systems did. This may be due to the fact that a great many connectors are involved in these systems, and that the seat environment for connectors and electronics can be hostile. Early experience with MIL-STD-1553 serial multiplex bus systems, however, has been favorable.

It should be noted that whereas multiplexing has the potential to save cost, experience shows that it will not automatically do so.

2.1.5 Toplogical Considerations

Communication system requirements are not fully characterized by distance, bandwidth, and number of signals. It matters where the signals originate and terminate. A natural hierarchy has requirements that differ from those of, say, a telephone system, in which each of N subscribers has random access to all of the (N-1) other subscribers. In the hierarchy, messages are channeled, and batch transmission techniques (e.g. multiplex) can be used.

Transport airplanes are less like telephone systems than they are hierarchical, but a great many things happen in parallel, so that a simple hierarchical model is inadequate. The greater the degree of functional integration, the more this is true. Section 2.1.8 discusses this subject further. In highly integrated systems, sensor information is pooled among functions, and function outputs may be shared among other functions, as well. This situation is supportable in a hierarchical system structure, but only if communication bandwidth resources are plentiful. The more dispersed the system is, the more costly this may be.

2.1.6 Technology Factors

The urgency to push bandwidth in long channels stresses technology in various ways. Power gain is one source of stress. It requires more signal power to transmit at higher bandwidth in a given channel. Power gain requires "real estate" in semiconductors, board area, heat exchange, power supplies, and rack space. It also tends to mean more connections at a higher level, since high-power circuits are not amenable to large-scale integration. It means more sources of concentrated high temperature, as well.

A more sinister form of stress occasioned by bandwidth enhancement is in the statistical error rate of the channel. The occurrence of an error always compromises a system. Coding or replicated redundancy may be required in order to contain the effects of the error, in which the system is burdened with extra equipment, with its extra failure rate and its extra overheads, including redundany management. A particularly burdensome impact of errors is felt whenever system recovery is not transparent to applications software.

High bandwidth transmission channels tend to be vulnerable to component aging and damage. This is another consequence of reducing comfortable tolerances to gain useful bandwidth.

2.1.7 Reliability Considerations

Reliability in the generic sense of the word denotes longevity of some sort, all forms of which are desirable. It is not possible, however, to enhance all longevity forms simultaneously. Therefore it is important to know the consequences and costs of the various forms. A simplified way to think of the subject is that airplanes operate in cycles: flights, days, and overhaul periods; and different forms of longevity apply to each.

During the flight of an active-control transport, a body of critical equipment must have an extremely remote probability of failing with any catastrophic consequences. Those failures that do occur must be detectable, and their effects containable, to a very high probability. Lapses of communication or of power should not exceed times of the order of milliseconds. This implies adequate redundancy of a kind amenable to in-flight requirements for essentially continuous control.

On the ground, the airplane sits at a gate between flight legs with a short turnaround time between arrival and departure. The penalty for delay and/or cancellation can be substantial, although in no wise comparable to the penalty for catastrophic in-flight failure. Economic viability requires that the need for at-gate maintenance during the day be encountered seldom, and then that the maintenance action be simple, such as swapping a common LRU. It is important that smaller airports not need to be stocked heavily with spares. Forgiveness exists, however, in that a design tradeoff must be made between dispatch probability and acquisition cost, so that occasional violations are expected of the rule that no maintenance be made during daily stopovers.

Overnight maintenance of LRU's of any kind may be considered normal, provided that the total number of maintenance actions is consistent with industry expectations. Communication and power links embedded in the airframe structure may not be repairable in an overnight time frame, however, and therefore should be repaired only at the time of overhaul. Again, exceptions are anticipated on a statistical basis, according to design tradeoffs.

2.1.8 Highly Integrated Avionics System Considerations

An unintegrated avionics system can be defined as the aggregation of elements (sensors, processors, and effectors) which mechanize a particular set of flight-related functions (e.g., navigation, flight control, displays, and controls). Associated with each of these functions is a particular subset of the system elements. If these functions are mechanized as autonomous subsystems, then the subsets are made disjoint, and the interconnectivity problem at the system level is reduced to the interconnection of relatively few numbers of relatively large aggregates of functionally related elements.

In a highly integrated avionics system, on the other hand, the subsets of system elements associated with the various avionics system functions need no longer be disjoint. Indeed, in the terminology of set theory, integration represents an effort to ensure that the total set is as small as possible by allowing the various subsets to intersect or share elements whenever possible. Thus, the problem becomes one of interconnecting relatively numerous small aggregates of system elements.

The interconnectivity problem at the system level for a highly integrated avionics system is thus seen to be fundamentally different from the interconnectivity problem in systems consisting largely of autonomous subsystems. This difference creates new requirements to be met by the communication structure. It is from the reduction in size and component numbers, the total set size, that potential advantages flow.

What are these advantages? First, a significant reduction in weight, volume, and power consumption of avionics systems can be achieved through the multifunctional use of system elements. Multifunctional use consists of using a single set of sensors to satisfy a number of different requirements for a particular kind of measurement, of using a pool of shared information-processing resources to satisfy diverse processing requirements, or of using a small number of effectors in combination to effect a wide variety of control modes. Second, since the addition of a single component may in effect add redundancy to several functions, making each more reliable, it is possible to purchase increased reliability economically. Both of these advantages are potent in terms of satisfying pressing requirements for reduced size and weight, and for substantial increases in reliability.

What new requirements are placed on the communications structure? Foremost among these new requirements are those demanding ultra-reliability, high bandwidth, and support of many data sources and sinks.

Why is ultra-high reliability now required where it was not before? Previous unintegrated systems employed autonomous subsystems. Even where individual subsystems might be flight-critical, care was taken to minimize or eliminate the flight-safety implications of subsystem-to-subsystem communication failures. While some data were exchanged, which allowed subsystems to optimize their performance, degraded modes or contingency control within the subsystems provided safe control alternatives, even if inter-subsystem communications were to fail. In short, most previous designs were working toward reliability goals, logistics costs, and operational convenience and availability goals.

In contrast, failure of the communications within an integrated system has immediate safety implications. Collapse of the communication structure could lead to the loss of the aircraft.

Why must the integrated avionics communications system handle increased data traffic? Previous communications systems designed to handle inter-subsystem data traffic did not see any of the subsystem internal data traffic. For example, the high-bandwidth traffic between the inertial instruments and the navigation computer is not visible external to that subsystem. In a fully integrated system, each of the inertial instruments is a shared resource, and the data traffic between them and the navigation autopilot must be supported. Some of the data traffic within the new avionics architecture is from one source to one target; some is from many sources to a single target; and some is from one source to many targets. Depending on the exact volume and nature of each of these data exchanges, the new architecture must provide dedicated paths, two-way buses, broadcast buses and a quasi-hierarchical aggregate of all of these elements. Additionally, it must provide the necessary redundancy and robustness so that the communication structure can survive the random faults, data-terminal failures, and physical damage that cannot be purged from its environment. It must provide all of this with adequate reliability and minimum complexity and flexibility.

Finally, it is clear that when the integrated avionics system is compared to more conventional designs, the numbers of communicating data terminals have increased greatly. Thus, the communication system

is dealing with a multiplexing problem made more complex by an increased number of data sources and sinks. In effect, the integration of the avionics system, through multifunction use of elements and pooling of resources, allows significant reductions in numbers of sensors, displays, processors, actuators, etc. These reductions come at the expense of higher connectivity and reliability requirements for data communications. The tradeoff is highly favorable for the integrated system because of significant recent advances in electronic technology, which have enormously reduced the cost/capability ratio of the required data-communications facilities.

2.2 Introduction to Signal Transmission Methods

Signal transmission, within the context of this study, is required for the following types of locations:

- Within a complex component, e.g. a fault-tolerant computer
- Within a bay to interconnect boxes
- Between bays
- From bays to passive sensors or effectors, e.g., LVDT's, solenoids
- From bays to active effectors, e.g. displays
- From active sensors to bays.

Some of the signals involved will be less than flight-crucial. Our purpose here, however, is properly served by ignoring these. Equivalently, we may assume for the present that all of these signals are flight-crucial. This implies requirements on the system as a whole, as follows:

- Continuity of service
- Transparency
- Maintenance postponement
- Graceful degradation
- Migration of authority.

The first requirement is for "continuous" availability, where the quotation marks signify that lapses may occur for periods of the order of a few milliseconds without necessarily impairing system operation. Second is a requirement for "transparency" in redundancy management, where neither the crew nor the application software is fully responsible for malfunction recovery owing to time-scale and

complexity problems. Third is a requirement for sufficient redundancy to allow the postponement of maintenance to a convenient time and place. Fourth, no random or induced malfunction of any one element may totally impair the system. This may be said to be a requirement for graceful degradation. Fifth, and finally, as regards the signal transmission media, it must be possible to support the migration of authority and responsibility among the various bays and boxes in response to malfunction-induced reconfigurations.

The various signal propagation methods can be distinguished by several attributes. One is the degree to which a given interconnection is dedicated to a single signal, a single component, or a single function. A second is the degree to which a single link is localized, as opposed to being system-wide as in a data highway. This second attribute might be mistaken for a repetition of the first, but is necessary to distinguish the case where local links collaborate to form a nondedicated network. A third attribute concerns the degree of bidirectional path capability. This is called "simplex" for a one-way link, "half-duplex" for alternate use of one channel in each of two directions, and "full-duplex" for dual channel simultaneous bidirectional capability. A fourth attribute concerns the degree to which redundancy may be realized by non-replicative means, such as by encoding information or finding an alternate path in a mesh. A fifth attribute is the degree to which the method can interface directly with passive devices without the use of an active electronic interface. The sixth and final attribute is the degree to which the method is compatible with fiber optics.

This study has essentially confined itself to seven methods as follows:

1. Dedicated links, one signal per channel
2. Dedicated serial bus, point-to-point multiplex
3. Parallel or serial local bus
4. Broadcast bus, one-way multiplex bus
5. Standard multiplex bus such as MIL-STD-1553
6. Variants of standard multiplex buses
7. Mesh network, point-to-point links carrying general data traffic

The principal uses and attributes of these methods are summarized in Table 2.2-1, which is the basis for the following preliminary discussion of the seven methods.

TABLE 2.2-1

SUMMARY OF ATTRIBUTES OF COMMUNICATION METHODS

	A. INTRA-BAY	B. INTER-BAY	C. BAY TO PASSIVE SENSOR OR EFFECTOR	D. BAY TO ACTIVE ELEMENT	E. TYPE OF DEDICATION	F. LINKAGE	G. BIDIRECTION	H. REDUNDANCY	I. FIBER OPTICS POTENTIAL
1. Dedicated Links	Poor. Proliferation of connectors & interfaces.	Poor. Heavy and bulky plus proliferation of connectors & interfaces.	Excellent. The only valid method for passive elements .	Poor. Inefficient use of channel capacity. Heavy.	Separate channel for each signal	Point to point or multidrop	No	By replication, as much as several-fold in some signals.	For discretes, conceivably, where point-to-point.
2. Dedicated Serial Bus (point-to-point)	Fair to good. Less proliferation of connectors and interfaces.	Good to excellent. Reduces weight and bulk.	Poor. Not applicable to passive elements.	Good to excellent. Efficient.	One serial channel for each distinct source-destination module pair.	Point-to-point	No	Implicit in module-to-module structure.	Good. Point-to-point application.
3. Parallel Local Bus (multidrop)	Excellent. High bandwidth, simple protocols, possible error correction.	Poor. Not applicable outside of bay.	Poor. See 3B ← See 2C †	Poor. See 3B ←	Shared by all modules in a bay	Multidrop	Yes	By replication.	Poor. Parallel bus.
4. Serial Local Bus (multidrop)	Good. Modest bandwidth, medium complexity.	Poor. See 3B †	Poor. See 3B ← See 2C †	Poor See 3B ←	Shared by all modules in a bay.	Multidrop	Yes	By replication.	Poor to fair. Serial bus.
5. Broadcast Bus	Fair. Much like dedicated serial bus above, but permits multiple receivers. More economical but more vulnerable.	Good. See 5A ←	Poor. See 2C †	Good. See 5A ←	One or more serial channels for each source module.	Multidrop with one source.	No	Implicit in source module dedication structure.	Fair to good if receivers are few.
6. Multiplex Bus (MIL-STD-1553)	Fair to good. Like serial local bus but more complex.	Good to excellent. Fairly efficient, somewhat vulnerable.	Poor. See 2C †	Good to excellent. See 6B ←	Shared by all signals.	Multidrop with remote couplers.	Yes	By replication	Poor to fair. Serial bus.
7. Augmented multiplex Bus	Fair to good. See 6A †	Good to excellent. Less efficient but less vulnerable.	Poor. See 2C †	Excellent. See 7B ←	Shared by all signals	Multidrop with remote couplers.	Yes	By replication	Poor to fair. Serial bus.
8. Mesh Network	Fair to good. See 6A †	Good to excellent. Slightly less efficient than multiplex bus. Less vulnerable. Damage Tolerant.	Poor. See 2C †	Excellent. See 8B ←	Shared by all signals and/or multipath options	Point-to-point	Yes	Implicit in mesh	Good. Point-to-point

2.2.1 Dedicated Links, One Signal Per Channel

Dedicated linkage with one signal per channel has an important role wherever time-shared use of a channel is impractical. The method is conceptually simple, and its interfaces are usually simple also. One disadvantage in conventional dedicated systems is that the cost and weight due to links and interfaces can be excessive when the number of signals is large. In fault-tolerant systems, the disadvantages are especially severe with respect to graceful degradation with damage and the ability to accommodate reconfiguration. This is largely because of the large volume of signals with different sources and destinations.

Dedicated links in many instances can be eliminated through the use of multiplexing. In other instances, however, it is preferable to pay the cost of separate channels and interfaces. Control surface actuation is one such instance. When secondary actuators are local to control surfaces, dedicated wires carry electrical signals between these actuators and the servo amplifier boxes located in one or more bays remote from the actuators. The typical signals are either analog, representing LVDT positions and electro-hydraulic valves, or discrete, controlling shut-off valve solenoids. Another possibility for control surfaces would be to integrate the secondary actuator with the servo and use a hydraulic or mechanical link from the remote secondary to the primary actuator. Thus the dedicated link methods is not restricted to electrical phenomena. Pneumatic links are commonly used between pressure ports and air data computers.

Apart from its ability to carry signals to or from passive devices, dedicated linkage is classified here as generally being a poor choice compared to other methods that support time-shared communication. Redundancy must be implemented via full replication of the medium. Fiber optics is not applicable here other than to light-producing or light-actuated devices, none of which are being considered in this study. The only fiber optics links considered here have active devices at both ends, and belong to the second category.

2.2.2 Dedicated Serial Bus, Point-To-Point Multiplex

The second dedicated channel category differs from the first in that multiple signals sharing a common source/destination pair are routed over the same channel. This precludes any use of passive interfaces, and for all practical purposes requires digital transmis-

mion. Because of its efficient handling of multiple signal traffic, it is a reasonable candidate for traffic external to bays wherever multiple signals are involved. It can also be used within a bay between boxes. The use of multiplexing in this point-to-point medium makes fiber optics an excellent potential candidate technology for simplex (one-way) transmissions.

2.2.3 Local Bus

One of the contextual assumptions for this study is that it may be necessary for the system to withstand a completely damaged bay. It may therefore not be necessary, and perhaps not feasible, to make bays internally damage-tolerant. A serial or parallel local multiplex bus affords a very efficient intra-bay medium. It is usually not recommendable, however, for busing external to bays, owing to the excessive cost of damage-tolerant constructs using such buses as building blocks.

A local bus is a multidrop half-duplex medium, i.e. it is neither local to a single source-destination pair, nor dedicated to a limited function. This medium is one in which a degree of redundancy can be added by incorporation of code bits on extra channels, resulting in a moderate degree of fault tolerance which in some cases might be sufficient for its application. This depends on the allowable random failure rate for a single bay, which depends in turn on the number of bays, among other things.

Fiber optics implementations of such a medium are possible in principle, but their expense would not be warranted for an intra-bay application.

2.2.4 Broadcast Bus

A broadcast bus is a serial bus, in which multiple receiving parties can receive data from a single transmitter. This is no longer a local medium, but it is dedicated to the functional scope of the lone transmitter. The single-transmitter definition excludes bi-directional use of a single channel.

Fiber optics is a possible means of implementing a broadcast bus, although it would probably resemble closely a multiplicity of point-to-point links.

When broadcast buses are used in redundant systems, fault independence of the receivers may not be safely assumed. If a signal is to be routed to multiple destinations where receipt by at least one

destination is critical, then multiple dedicated serial buses are used. Moreover, special care must be taken that these buses do not short to one another in an undetectable way, since they would present a latent system hazard.

Broadcast buses are of special interest in this study, since they represent contemporary technology, are standardized for aircraft use, e.g. through ARINC 429, have performed well in early applications, and will have interface circuits implemented in large scale integrated form. They have important roles in contemporary designs, including the Boeing 757 and 767.

2.2.5 Standard Multiplex Bus

Standard multiplex buses, as exemplified by MIL-STD-1553B and its prior versions, are based on the concept of a single shared channel like the parallel bus discussed above. In this case, however, the physical distances are such as to place important constraints on the electrical realization. Transformer coupling, complex interfaces, and a non-negligible error rate testify to the design challenges that have to be overcome. The maximum number of parties served is 31.

Fiber optics realizations have been demonstrated on the bench, but are apparently not yet deployable owing to severe design, manufacturing, and installation problems. The problems are less severe where fewer parties are served.

A bus of this type is damage-vulnerable, so that any realistic system application would require replicated buses, installed in such a way as to remain far enough apart so that a single damage event is unlikely to affect more than one. The 1553 standard includes the necessary provisions to couple remotely into the buses. Dual redundancy is the most prevalent redundancy form in 1553 applications so far, but in a flight-critical application the likely form would be triplex or quadruplex.

This form of multiplex bus is vulnerable to a single party which fails to observe proper system "etiquette."

2.2.6 Variants of Standard Multiplex Buses

Various departures from the 1553 standard form have been proposed, primarily to overcome the physical limitations and vulnerabilities of 1553.

Hierarchical complexes of standard buses can in principle serve an unlimited number of parties. There are two drawbacks, however. The response delays will be substantially larger, and redundancy can become substantially more complex. Both of these problems can be solved, but require departures from the standard.

One departure from the standard, the French GINA bus [4] uses two channels for each transmission, one of which handles simplex commands from the control unit; the other channel is half duplex, and carries data.

A generic problem with any half-duplex multiplexing medium is the inability to deliver commands to remote interfaces when the sole channel is obscured by noise. Power switching of remote terminals has been proposed as a solution, which will indeed solve part of the problem if a trustworthy form of power control is included. A less tenuous solution, however, requires the ability to exert control on a party's transmission at a point external to the party itself. A full-duplex command-response bus with "smart" independent interfaces to every party would potentially be able to reduce single-event vulnerability to a very low level.

Redundancy in an augmented multiplex bus might be implemented in various ways depending on the types of variations employed. A hierarchical complex might not require outright replication, whereas a full-duplex variation probably would.

2.2.7 Mesh Networks

Mesh networks are multiplexing constructs which employ point-to-point serial full-duplex links to communicate among a group of parties. Each party, or "node", has a direct connection of this kind to three or more other parties, forming a mesh-like pattern if properly done. Each node is capable of repeating incoming data, i.e. transmitting it to one or more of the other nodes. The ARPAnet is a mesh network which transmits "packets" consisting of many words according to an adaptive routing algorithm that may send successive packets over different routes. A delay occurs at each node.

The mesh network conceived by T.B. Smith, [5] on the other hand, sets up paths which remain stable over relatively long periods, supporting bus-like protocols with negligible delay at each node. The Smith net has intrinsic redundancy which is capable of protecting against faults and damage by allowing reconfiguration despite one or

more noisy nodes.

2.3 Dedicated Links

The remaining sections of this chapter enlarge on the introductory comments just completed concerning the various methods of data communication.

Dedicated links have been the conventional carriers of data communication since its inception, principally in the form of analog data and discrete signals. Growth in system sophistication has been acompanied by proliferation of dedicated links to the order of ten thousand per aircraft. The ARINC packaging standards provide for up to several hundred pins per box. Wiring occupies enough volume so that in certain places, such as wings, tail, instrument panels, and wheel wells, there is often scarcely enough room to contain it all, much less to provide comfortable spatial separation for redundant signals.

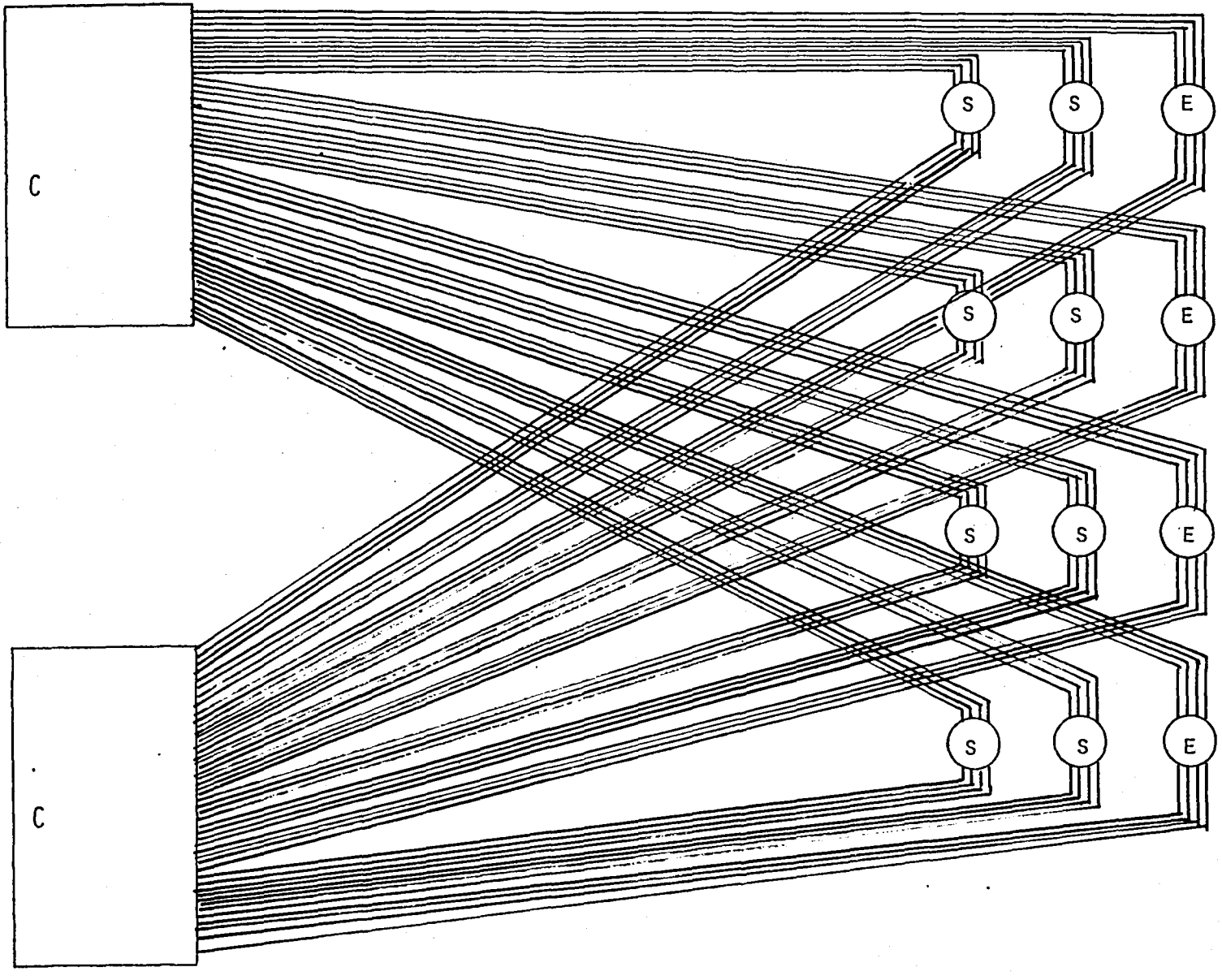
Figure 2.3-1 illustrates the concept of dedicated links carrying dedicated signals in a fictitious system employing dual controllers, eight sensors, and four effectors (e.g., actuators). Each sensor and each effector are shown as having four signals. It is immaterial here as to which direction a given signal flows. Figure 2.3-2 indicates that any given sensor or effector may involve bidirectional information flow.

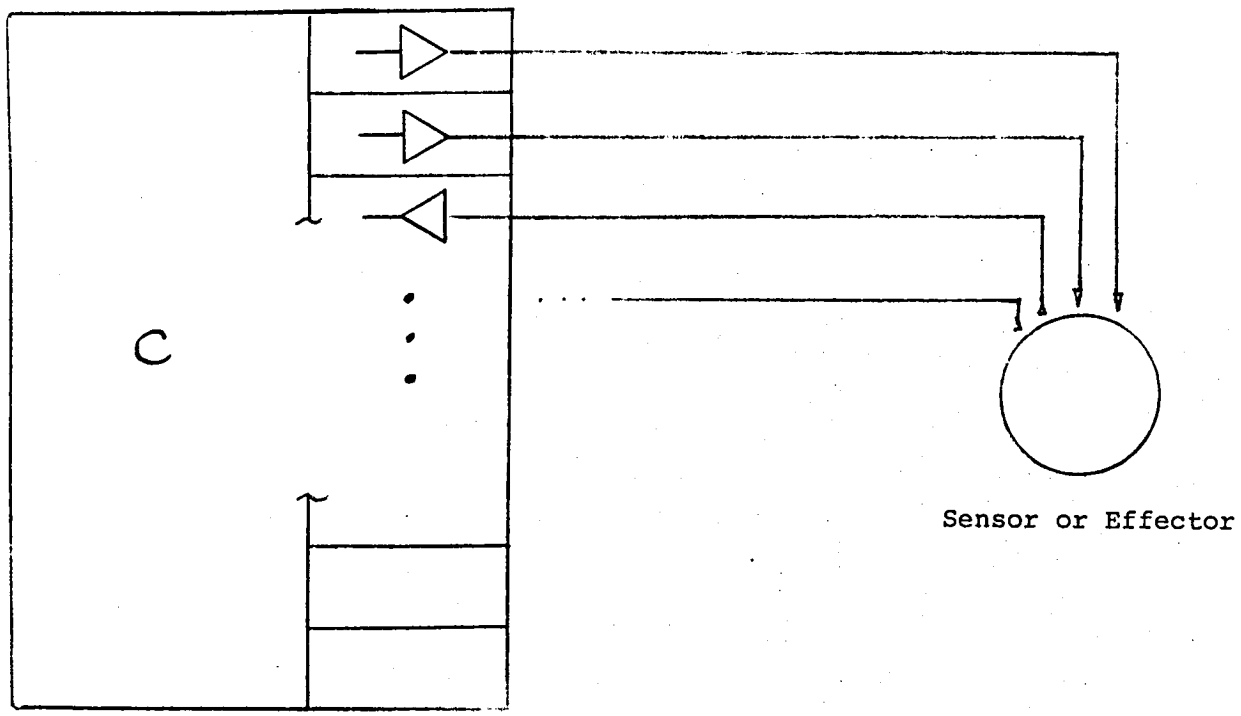
By multiplexing the signals that share a common source-destination pair, the system of Figure 2.3-1 is changed to resemble Figure 2.3-3, in which the figure is drawn assuming that two-way multiplexing is used. This represents the irreducible minimum configuration of dedicated links, where the former figure represented the maximal configuration.

Aside from their familiarity, dedicated links have two main advantages over more general multiplex forms. The first is that they provide ample bandwidth without stressing technology. Their cost is largely in wire volume rather than in sophisticated electronics. The second is that the malfunction of any one link can not deprive the system of more than the one or few signals it carries.

The principal disadvantage of dedicated links is their cost, particularly the maximal configuration. In a sense, they pay heavily for their advantages cited above. There are other disadvantages, however.

Figure 2.3-1. Dedicated Links Per Signal.

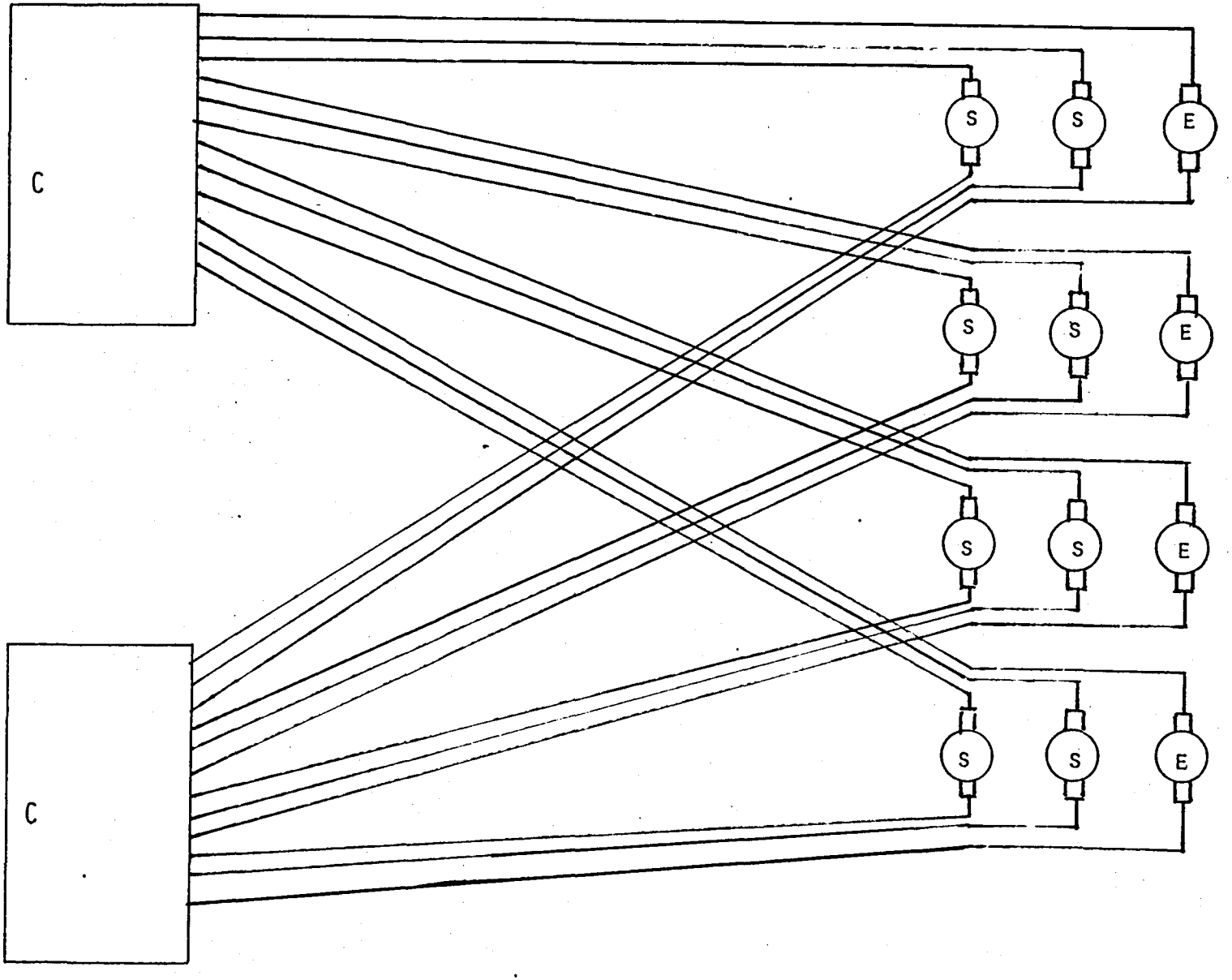




C: Controller

Figure 2.3-2. Dedicated Signal Interfaces.

Figure 2.3-3. Dedicated Links Per Source-Destination Pair.



One such disadvantage is that the controller is specific to the configuration, rather than being universal. The addition of new signals and/or source-destination pairs requires new hardware, wire runs, connector changes, and/or new links. Meanwhile, the controller is burdened with a large, awkward array of interfaces.

Another disadvantage is that even though the propagation of malfunction effects is bounded, all links and notably the longer ones, have large exposures to both faults and damage, owing to the interfaces, connectors, and wire runs. A credible damage event could sever enough wires so that neither of two controllers could communicate with a sufficient subset of the sensors and effectors to maintain flight in an active-control transport.

2.4 Local Buses

In the category of "short distance" data transmission, a large amount of data may be transmitted within a single bay. In general, such intra-bay transmissions may be made at relatively low cost by back-panel links. Dedicated back-panel links are inexpensive, but they still suffer from being design-specific and awkward. Multiplexed communication, on the other hand, can provide a graceful and general means of effecting intra-bay data transfer.

Both serial and parallel formats can be appropriate for intra-bay communication. ARINC 429 multiplex broadcast busing is used now for intra-bay traffic as well as inter-bay traffic, as is discussed in the next section. The present discussion is primarily concerned with two-way, half-duplex, multiplex busing for intra-bay communication.

Two-way busing differs from dedicated signals and broadcast busing in that multiple transmitters are present in the same channel as well as multiple receivers. Protocols become important as means for resolving contention for trnasmission access to the bus.

As a general rule, local buses can afford multiple channels or parallel transfer formats as means to accommodate high bandwidth. The reason is that intra-bay transfers are relatively well-sheltered, both electrically and damage-wise, which means that interfaces can be simple and economical. Meanwhile the total wire volume does not grow appreciably, because runs are short.

One interesting exception to the rule occurs in the design of certain fault-tolerant computer systems, such as the FTMP computer [3].

In the FTMP, the minimal serial bus set among processors comprises twenty channels, two hundred transmitters, and two hundred receivers. Clearly, the cost of parallel buses of, say, 16 bits would be prohibitive, as it would involve 320 channels and 3200 transmitters and receivers. Instead, the FTMP uses high-bandwidth channels of up to eight million bits per second each.

The technology for local buses has matured in numerous instances, including the DEC Unibus ®, the IEEE 488 instrument bus, and the Intel Multibus ®. Such buses are primarily used under control of a single computer, although some of them are arranged so that they can support multiple computers. Emerging standards will likely be influential in the design of any local buses for aircraft.

In a redundant system application, it might be desirable to use redundant buses and redundant controllers arranged so that any controller is capable of controlling any or all of the buses. This kind of arrangement suffers from a single-point failure mode, however, where one controller fails in such a way as to interfere with traffic on all of the buses. This could be tolerable in cases where the fault propagation boundary is set at the level of the entire bay, i.e., loss of the bay is tolerable provided that it is sufficiently improbable that more than one bay is lost in flight.

In order to set the fault propagation boundary at a lower level, no one controller would be able to transmit on all buses, which would mean in most cases that there would be one bus per controller.

Two examples are shown in Figure 2.4-1. In both examples, the subscriber components have redundant interfaces. In the top example, an active failure of either C1 or C2 will fault the redundant bus. In the bottom example, a passive or active failure in one controller and one bus will fault the redundant bus. Each is immune to the other failure. Depending on which of these failure modes has a higher probability than the other, one of these two approaches will be preferred, unless the probabilities are both so low as to make the choice irrelevant to safety.

2.5 Broadcast Buses

Broadcast buses are digital analogies to analog signal channels. The transmitting terminal, instead of maintaining an analog voltage, repetitively transmits a serial digital value. In either case, there can be one or more listeners. The broadcast bus has advantages in that

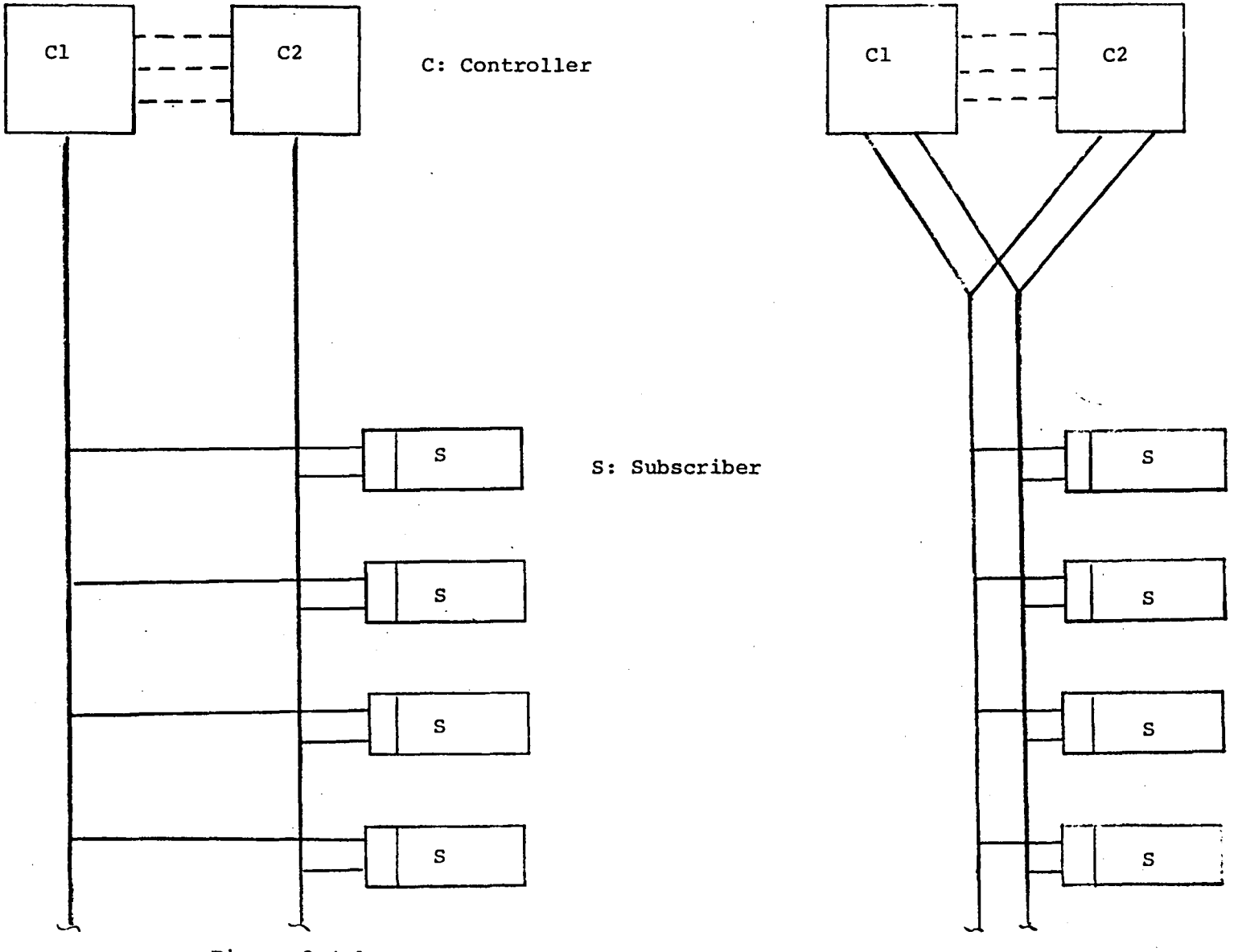


Figure 2.4-1. Examples of Controller and Bus Arrangements.

the channel can be time-shared by multiplexing among several digital signals. Each signal can be identified by any of several means. In the ARINC 429 broadcast bus, each value is accompanied by an identification tag to indicate to the receivers what signal value is being sent. Protocol is almost non-existent. The sole transmitting terminal simply emits data with identifier tags, much as a stock ticker system emits market quotations to local receivers.

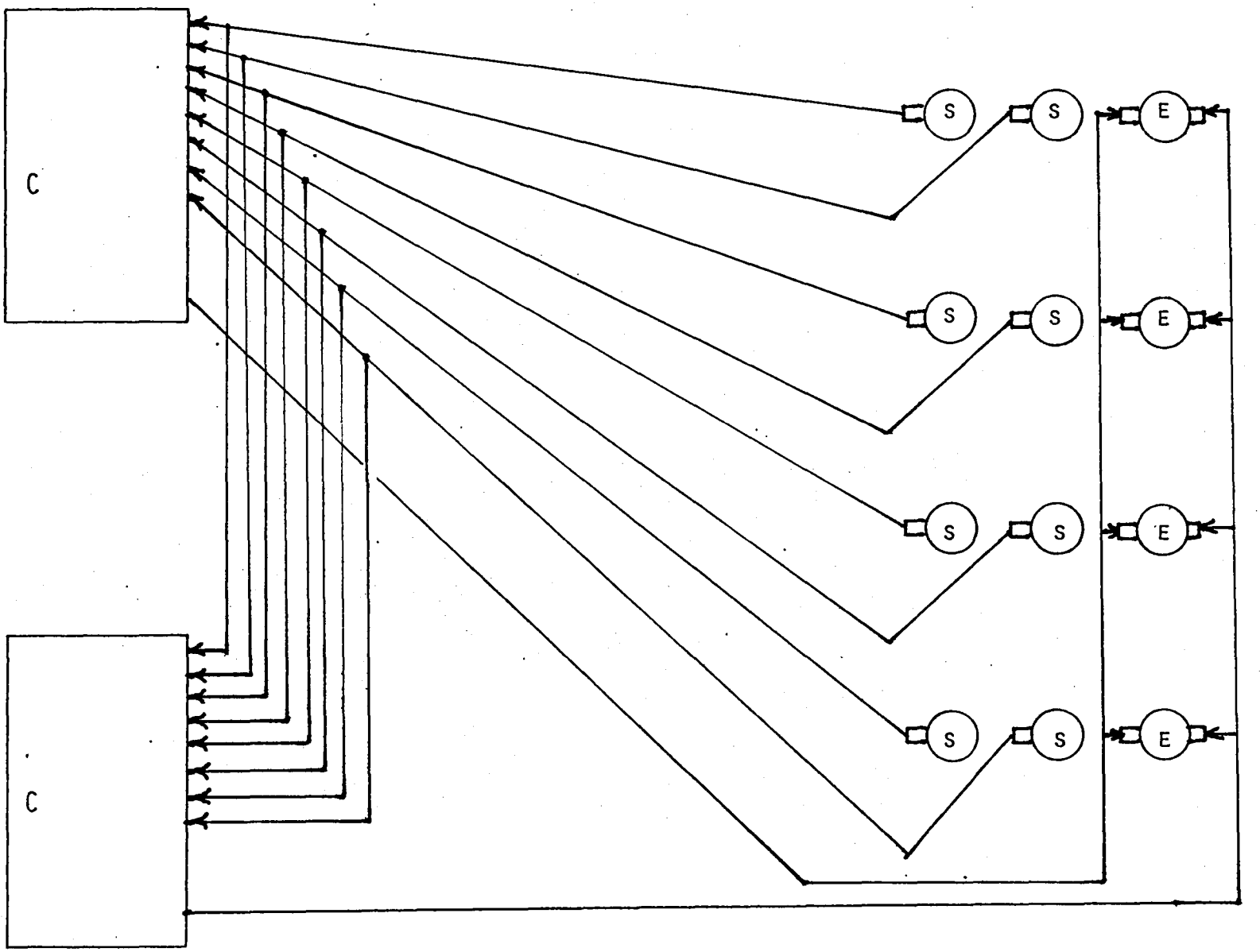
To the extent that data signals have multiple destinations, broadcast buses are more economical than dedicated multiplex links. More to the point, one-way dedicated multiplex links can be thought of as broadcast buses with single receivers. (Two-way dedicated multiplex links are not equivalent to broadcast buses). Thus the economical considerations of broadcast buses depend heavily on the data migration patterns of the system. Meanwhile multiple destinations present a potential system hazard, as may be seen in Figure 2.5-1. When all effectors are linked by broadcast buses, a damage event at any one of them that causes both inputs to short circuit will result in system failure. A similar statement holds for either of the controllers, but this may be considered to be far less likely, as there will be fewer controllers, which are located in safer places. In any event, critical effector signals require independent links rather than broadcast buses, plus separate buffered interfaces so that electrical accidents on one link will not affect another. It could conceivably be necessary for similar treatment of critical sensor signals, but this would depend on details of the particular system.

Most broadcast buses used to date have operated at a rate of 100K bits/sec or less. Bandwidth has not been a particular problem, since the number of broadcast buses in the system may be on the order of one hundred, no one of which carries an inordinate share of the data load. The technology needed to support such low data rates is resonably simple, since reflections do not present a significant problem. Low data rates present a latent hazard possibility unless care is taken to prevent the shorting together of two isolated links carrying identical data. Should this happen, both links would appear to be working correctly, whereas the intended isolation would be absent, leaving the system vulnerable to a short circuit in one of the links.

2.6 Standard Multiplex Buses

In principle, it would be possible to eliminate most of the linkages and interfaces in an airplane, using a single two-way channel

Figure 2.5-1. Broadcast Buses.



time-shared among all subscribers (i.e., all boxes). Each subscriber would have a single interface to the bus. Of course, the bus would have to be reliable enough and have a high enough bandwidth. In practice, neither of these requirements is is readily achieved consistent with the premise. Nevertheless the fundamental appeal of the concept is strong enough to leave a good deal of room for compromise.

The potential benefits of two-way multiplexing stem from the nature of the interface as well as from the minimal channel volume. Since all boxes have a single bus interface, a high degree of standardization can be maintained, promoting equipment commonality, lowering design costs, and favoring competitive procurements. Again, realization of the potential is difficult, but worth the effort.

The sharing of a single two-way channel is a cooperative endeavor. Any subscriber can potentially pollute the channel by spurious transmissions. The only remedies for such a threat are either to make subscribers purge themselves, or else to adopt secondary control channels to override subscriber autonomy. One such approach is to suspend electric power distribution to the individual subscribers, one at a time, until the offending unit is found, and to leave its power off. This approach requires that power distribution be properly controlled, and not misused by a failed controller. This in itself may require a second communication channel, which would violate the single-channel premise.

Given proper cooperative behavior in a bus channel by all subscribers, the channel can be shared by any of several algorithms called "protocols." One type of protocol defines fixed time slots assigned to the respective subscribers. Say there are S subscribers. Then S consecutive time slots of individual duration T make up a frame of duration ST. Each subscriber transmits during the time slot assigned to it. Its slot number is wired in, and it counts slots from a synchronization mark at the beginning of the frame. The channel must either be synchronous, or else have a mark to denote each slot boundary. A possible variation on this scheme would allow more slots per frame, assigned on a basis of need. Another would permit variable length slots. Protocols of this sort are called "time division, multiple access" or TDMA protocols.

A second major category of protocols is called "command-response," which operates on a speak-when-spoken-to basis.

Unlike the TDMA category, command-response requires the definition of a specific unit at any given time to be the system controller. The controller identity might be passed from one subscriber to another like a baton, or it might never change.

Another major protocol category is called contention. Whenever the bus is not busy, any subscriber is eligible to bid for access to the bus. Several variants of this method have been used. One variant calls for the contender to begin its transmission directly, and to detect possible interference from another contender using error detecting codes. In case of interference, each contender waits a different length of time before retrying. Alternatively, each contender can wait a different time before beginning its transmission after the previous transmission in order to lessen the probability of interference in the first place. Another approach is to have each contender synchronously transmit a priority word while listening to the bus. If the contender hears a higher priority than its own, it drops out of contention. Otherwise, it has won the contention.

All of the protocols for single two way buses share the attribute that all communications are heard by all subscribers. There are no private messages, as there may be in dedicated links and networks. Potential fault and damage vulnerabilities include shorted buses and terminals, open buses and termination impedances, and stuck or active transmitters. One form of bus called a "lossy" bus was devised to reduce some of these susceptibilities by placing series resistance in the bus line at each terminal. This creates a dynamic range problem, i.e. a problem for receivers of having to be able to decode signals of diverse amplitudes over a substantial range, depending on the number of terminals. Meanwhile, by constructing the bus with multiple paths, the channel can be made immune to most open and short circuits and stuck transmitters. It is not immune to active transmitters, however.

Perhaps the best-known example of a serial two-way multiplex bus is the existing military bus standard, MIL-STD-1553B. This standard, its predecessor A version, and the various applications of 1553, represent a well-accepted architectural framework. The 1553 bus is a partially-lossy bus, but without multiple paths. This compromise provides immunity to terminal shorts and opens but not bus shorts or opens. At the same time the dynamic range requirement is only moderately difficult to achieve. A stuck transmitter has roughly the same effect as a shorted terminal. Active terminal faults are dealt with

by incorporating watchdog timers on transmitters.

The 1553 standard calls for remote access to the bus using stubs of up to approximately six meters (20 feet) in length. If short stubs or no stubs were to be used, a single damage event could conceivably sever both or all members of a redundant bus system. Transformers and resistors are used to couple terminals to stubs and stubs to the bus (with certain exceptions). There is a limit to the number of terminals supportable by such a system. In 1553 it is defined to be 30. Figure 2.6-1 illustrates such a bus.

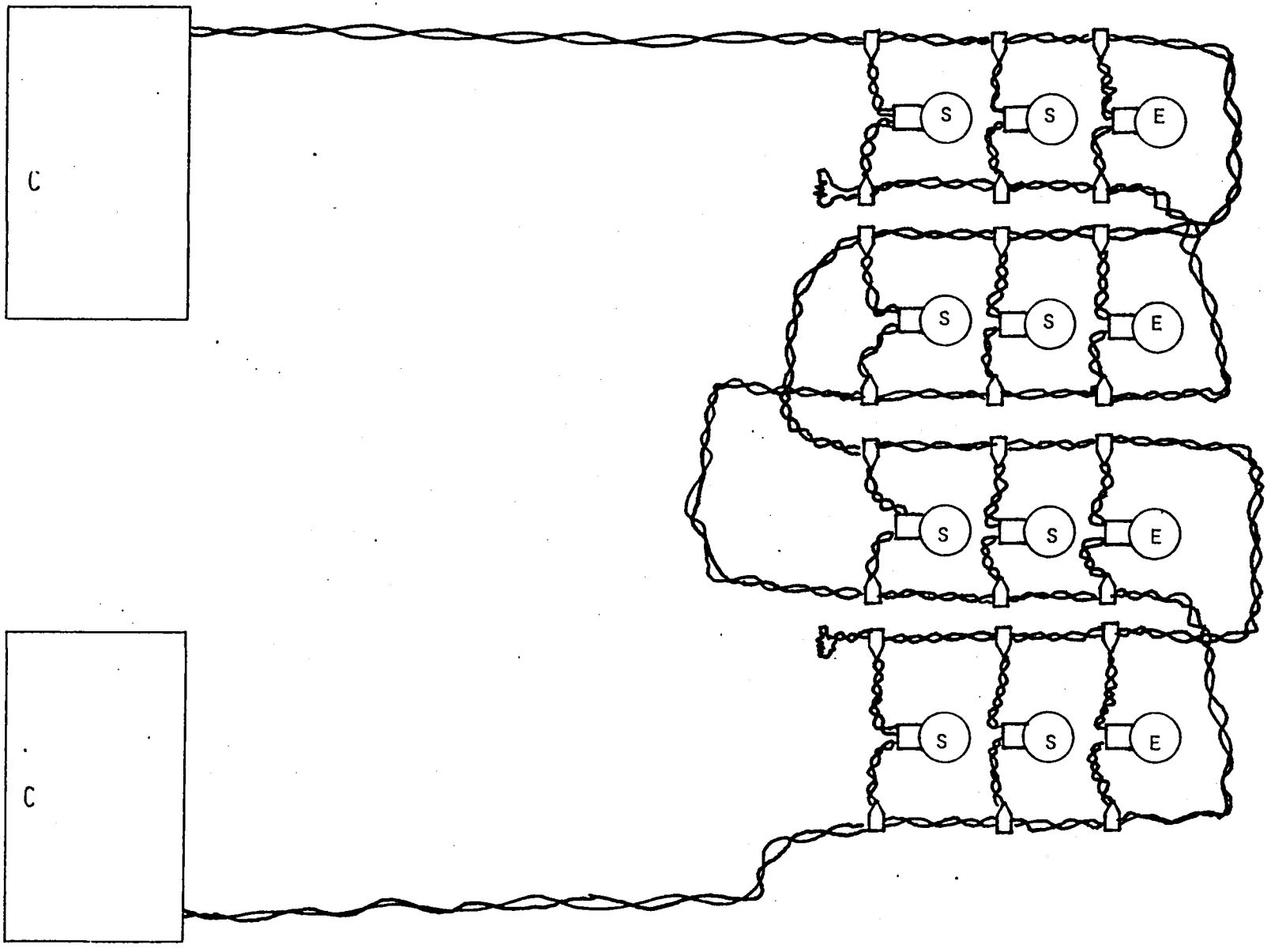
The protocol in 1553 addresses thirty terminals in a command-response manner. The B revision allows migration of the controller, while earlier versions do not.

This existing standard has shown itself to be reasonably resilient to pressure from its various applications. Some of this pressure has created confusion as to connectors, wave forms, specific meaning of various mode and submode commands, and other growing-pain incompatibilities that have diluted the benefits which might have been expected from 1553. Nevertheless, the hybrid microcircuit and large-scale integration (LSI) implementations have proceeded, and it is on these developments that economic viability and practicality will be based. The investments required to rival 1553 microcircuit and component developments all but preclude the development of an unrelated competitive standard for a similar architecture. The incompatibilities will be solved in the 1553 applications, and many new applications will be able to live reasonably comfortably within this agreed-upon architecture. However, the MIL-STD-1553 architecture is not infinitely expandable or elastic, and there will arise new technological demands which cannot be met. New solutions will have to be found.

The most significant shortcomings of 1553 within the context of a fully integrated avionics system are its inability to interconnect many data terminals, its vulnerability to physical damage, and an inability to assure that a single terminal will not bring down all attached buses due to erroneous transmissions.

The problem of being able to handle only a limited number of terminals (fewer than 31) has its roots at two sources. First, the twisted-pair, transmission, line-termination, and terminal-coupling techniques chosen can not tolerate many more than 30 terminals. Secondly, the protocol allows address space for no more than 31 remote terminals.

Figure 2.6-1. Multiplex Bus.



Historically, these limitations were the result of an architectural concept that viewed remote terminals as fairly large unrelated aggregations of sensors or actuators. Since each remote terminal handled many sensors or actuators, the terminal limit did not seem to constrain the system significantly.

To realize fully the advantages of integration, however, it is important that the number of individual sensors or actuators handled by a remote terminal be kept small. If this is not done, the failure of a single remote terminal can result in the loss of an excessive portion of the system's resources.

This problem has been partially attacked by the use of hierarchical buses. In its most conventional application, several subsystems might be joined by one 1553 bus (or dual bus), and within each subsystem a 1553 bus (or dual bus) is used to interconnect subsystem components. This solution parallels conventional architectures of separate autonomous subsystems. However, it is sensitive to failure modes which would make all the sensors or actuators of an entire subbus unavailable, due to failure of the terminal connecting that subbus to its supervisor bus. It also fails to address the case where it is indeed desirable to organize many data terminals onto one bus. This latter case more truly represents the natural organization of a highly integrated system, where one sensor must be used by several functions, rather than by just one subsystem.

It is possible by appropriate use of repeaters or bus buffers to eliminate the electrical constraint on numbers of terminals which can be interconnected, and still maintain functional compatibility with 1553. It is not possible to eliminate the protocol constraint without some modification to 1553.

The bus's vulnerability to damage is a result of the fact that damage to any portion of the bus can disable the entire bus, and that the bus is distributed widely, thus presenting a rather broad cross-sectional area to potential damage. Since a bus with more than one shorted stub is also likely to be disabled, this cross-sectional area to damage must include the stubs and portions of the remote terminals. This vulnerability provides a mechanism whereby fairly local damage can impact distant equipment. Damage to the wing could disable elevator control, for example. Any design which is truly flight-critical must include damage- and fault-containment mechanisms of the airframe structure itself.

The final serious weakness of 1553 is the relatively ineffective mechanism for preventing faulty terminals from talking out of turn and disabling the bus. The preventive mechanisms which are included are partly effective to the extent that this mode of failure is not likely to be a serious maintenance, operational, or diagnostic problem. However, the uncovered failure modes which could result in a "babbling" terminal are adequate to present a serious safety threat. Examples of such failures have already appeared in the field; one in particular resulted in the loss of an entire dual-bus system due to a single fault. It is probably in this particular aspect of the 1553 design that the difference between designing to maintenance and operational goals and designing to flight-critical standards becomes most evident. The basic reliability of the dual 1553 bus design is such that operational and availability impacts on an aircraft due to data-interconnect malfunction should be minimal. The 1553 bus is sound, easily maintained, and unlikely to cause aborted missions or other operational difficulties. It represents a significant and dramatic improvement over previous practice. However, when the effect of a communications failure is magnified from an operational aggravation (such as an aborted mission) to a loss of aircraft, the reliability constraints are increased significantly. Thus, while the cost of two mission-aborts per year per fleet of aircraft is almost invisible in the maintenance and operational costs associated with the fleet, the loss of two aircraft per year is highly visible, particularly if these losses are compounded with loss of life.

The sources of this vulnerability are many. Some of the dual-bus implementations are particularly vulnerable due to designed-in single-point failures. The primary defense, the watchdog timer on bus activity by a terminal, is ineffective against address decoder failures in the terminal, which cause it to respond to either the wrong address or to all commands. The interaction of a faulty terminal with broadcast modes, or the interactions between dual buses, present fairly simple mechanisms for disabling one or all buses of a redundant 1553 bus system. All of these mechanisms have likelihoods or probabilities associated with them which are insignificant if the only costs associated with them were maintenance actions and operational costs, but which are much too large if flight safety is involved.

These weaknesses can be overcome without breaking with functional compatibility with 1553. The same mechanisms used to overcome the

bus vulnerability to damage are also effective in overcoming the babbling terminal problem. A proposed solution is outlined in Section 2.8.

Additional weaknesses of MIL-STD-1553, such as inadequate encoding of the data and commands for error recovery and detection, are not serious enough that they could not be designed around or coped with.

2.7 Variants of Multiplex Buses

Two-way multiplexing presents problems due to its continuous channel, which couples all subscribers directly without the possibility of intervening reconfiguration.

Hierarchical arrays of multiplex buses have been suggested as a remedy. This provides a means of partitioning the communication channel into smaller fragments, which would permit reconfiguration to a certain degree. Figure 2.7-1 illustrates a redundant hierarchical bus array, in which subordinate controllers act as subscribers, or remote terminals, on the superior bus. This arrangement can increase the number of subscribers served, and can enhance survivability, to the extent that the loss of portions of the respective buses may be tolerable. The role of the subordinate controllers (C') in this scheme is simply to repeat messages from superior bus to subordinate bus, and vice versa.

The subordinate controllers introduce a delay in the transmission of each message, however, that could well be intolerable. It would at least intrude upon the standard protocols. In a single 1553 bus, for example, the controller expects a rapid response, within a few microseconds, to a command. Here, the subordinate controllers, C', would have to handle the response, perhaps buffering messages in either direction. In this case, the C' units have to become "smart," and either anticipate controller commands, or do repetitive reads and writes to service their subscribers. The net effect would be to loosen the reins of control. Messages would now need time tags to remove ambiguities introduced by the extra latency of the hierarchy.

The C' units might alternatively operate without buffering, simply amplifying signals in a bidirectional fashion, similar to telephone repeater amplifiers. The hierarchical bus system then becomes a network composed of bus segments as links and C' units as nodes. This leaves unsolved the problem of subscribers that transmit out of turn or incorrectly.

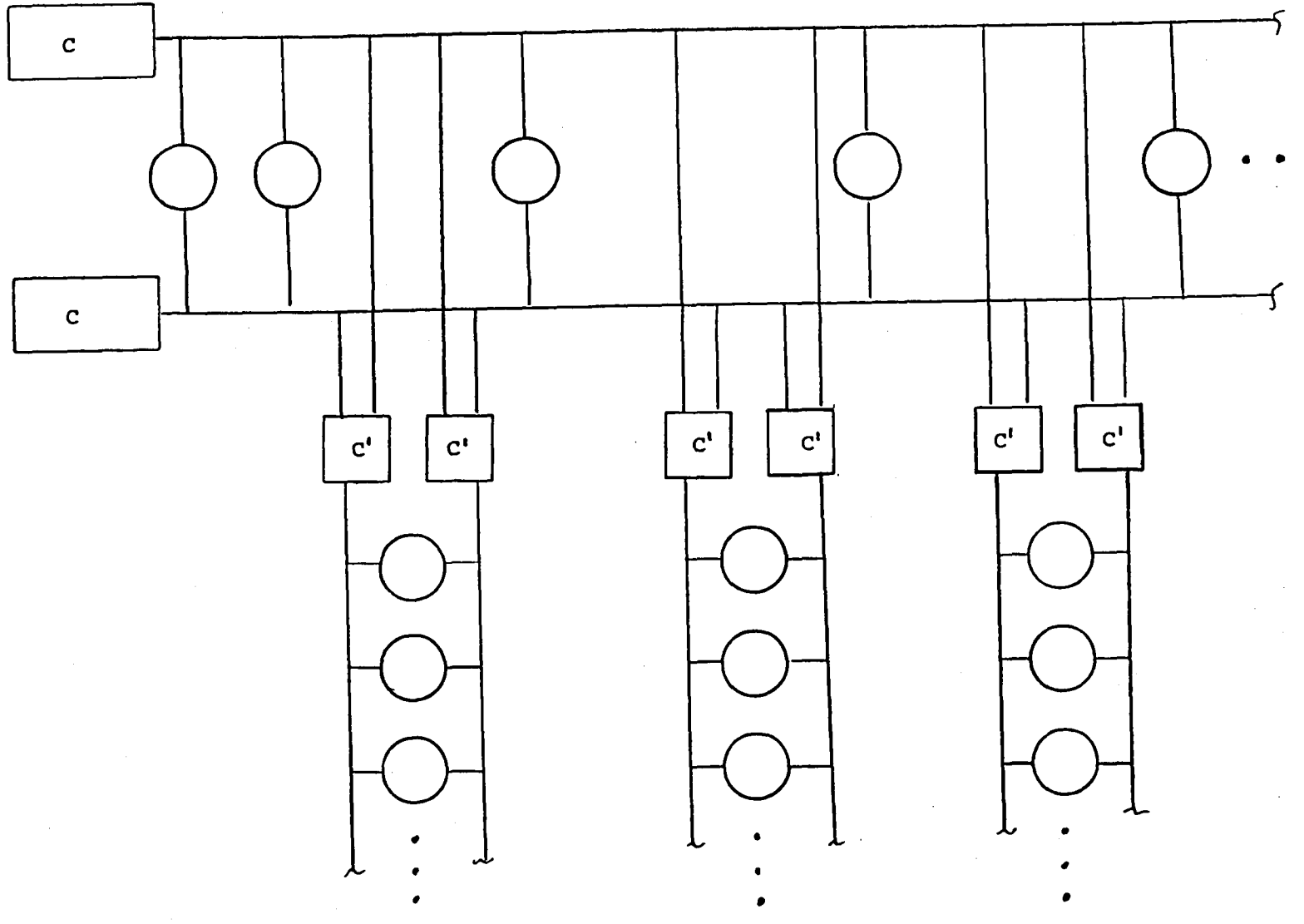


Figure 2.7-1. Hierarchical Multiplex Buses.

A second variation on the standard bus helps to solve the problem of uncontrolled babbling by an anomalous subscriber. This variation exchanges the half-duplex two-way medium for a full-duplex two-way medium, or else a hybrid arrangement still using two channels. The arrangement permits the controller to issue commands despite the pollution of the response channel by a subscriber. Commands thus issued may be used to invoke majority inputs to local controllers with which to selectively disable units until the culprit has been silenced. Meanwhile, the controller is the only unit that is equipped to transmit on the command channel.

A third variation uses subscribers as waypoints in a network or chain composed of point-to-point links. One version of this method is the mesh network, which is discussed in the next section and the next chapter. Another well-known topology of this sort is the ring network, in which a closed loop is formed. A typical ring network uses a logical token, passed from one subscriber to the next on the chain, to grant access to the communication channel. In order to prevent endless circulation of a transmission and consequent spurious waveforms, a subscriber holds the ring open while it is transmitting.

Rings may be designed with extra links so as to skip neighbors when necessary, and/or to be bidirectional. The topology of ring networks is strongly related to that of more generalized mesh networks only its layout and operation are more tightly constrained.

2.8 *Mesh Networks*

Mesh networks are major variants of standard multiplex systems. The topological principle of mesh networks is shown in Figure 2.8-1, in which subscribers contain, or adjoin, repeater and switching circuitry referred to as nodes, and where in this case each node has three ports. Each port interfaces one end of a link. All links are full-duplex, i.e. dual channel, so that commands can be sent to reconfigure the network despite the presence of anomalous transmissions from a subscriber or a node.

Meshes by definition consist of multiple circuit loops, which can potentially cause the problem alluded to concerning ring networks, where energy circulates to cause spurious waveforms. Therefore meshes, like rings, require restrictions on link connections in order to operate successfully. In this regard, mesh networks differ from lossy buses with multiple paths. The two may be topologically identical,

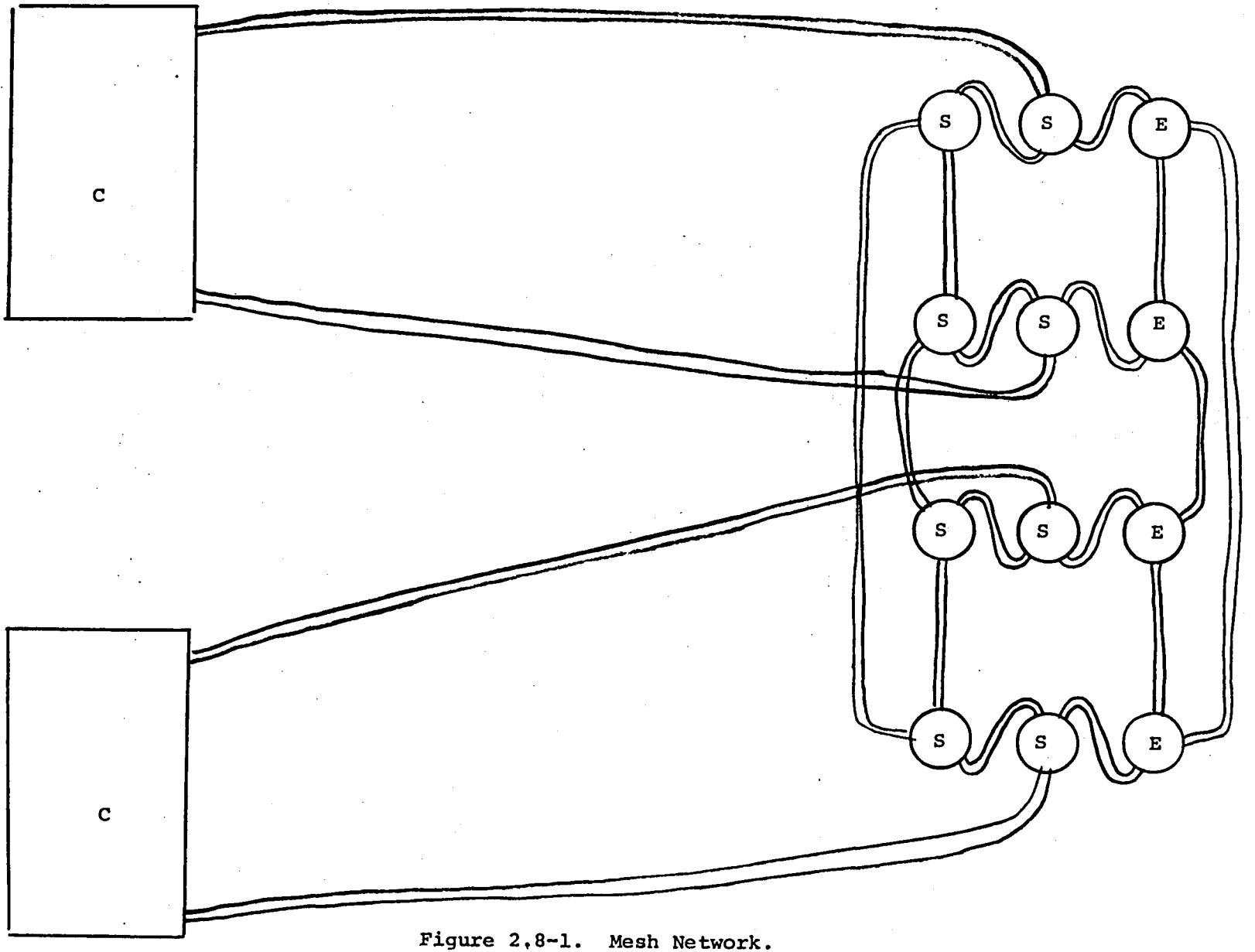


Figure 2.8-1. Mesh Network.

but operate in different ways. Mesh networks are software reconfigurable, where lossy buses are purely passive. Mesh networks can avoid the dynamic range problem cited earlier for lossy buses.

A well-known class of mesh networks is exemplified by ARPAnet, a transcontinental data network. Nodes of this kind of network buffer a quantity of data arriving at one port before deciding whether and where to transmit it through another. The time latency thus introduced is small by human interactive standards. This approach would be awkward for a flight control system, however, where transmission latency is less tolerable.

A family of mesh networks has been devised [5] in which the ports of each node are switched on and off by messages from the controller. The mesh is configured by this means in such a way that there are no loops. Data is then repeated with minimal delay and no buffering so that it arrives at all nodes nearly simultaneously, thereby emulating a standard multiplex bus. The most recent generation of this family is arranged so as to be compatible with subscribers using 1553 interfaces.

The architecture for this mesh network is a natural evolutionary step beyond 1553 practices. The constituent parts are bus segments (or links) and nodes which terminate and interconnect these links. A virtual bus can be created by activating circuitry within nodes, which effectively connects appropriate bus segments, one to another. This circuitry is analogous to relay closures which could actually create such a compound bus, but is implemented in solid-state devices. In its simplest incarnation, a single bus could be created by appropriately interconnecting multiple-bus segments to create one bus, which passes through each node. Figure 2.8-2 illustrates such a configuration. Active or utilized links are shown by solid lines, and inactive links are shown by dashed lines. Note that there are multiple options available as to how such a bus might be constructed from the available pieces, and that if damage or a fault should disable this bus, an equivalent bus could be constructed bypassing the damaged link. Figure 2.8-3 illustrates such an alternate configuration.

It is from this basic ability to reconfigure the bus routing that the high-survival characteristics of the network are derived. Note that once a bus has been created, it does indeed operate exactly as a true bus using standard bus protocols. Thus, there are no operational overheads associated with the operation of the virtual bus

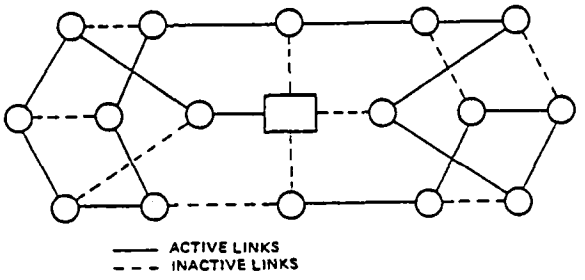


Figure 2.8-2. Network With Virtual Bus Shown.

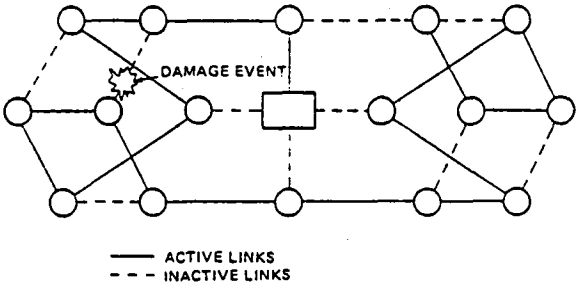


Figure 2.8-3. Alternate Virtual Bus Structure.

beyond those imposed by a standard bus and an initial setup or configuration procedure.

Damage containment and isolation of a remote terminal, which is disabling the bus, is now simple. First, each node is designed so that the interconnection circuitry provides isolation between bus segments. Electrical accidents are thereby blocked from propagating along the bus. At worst, such an accident can destroy only the isolation devices at the link terminations immediately surrounding the accident site. The logical impact of an accident, which is to disable the bus, can be overcome by reconfiguration. Because physical damage is confined to the immediate locale of damage, the success of reconfiguration is assured once the faulty components have been purged. Similarly a babbling remote terminal can be excised from the bus. Remote terminals can be attached at a node, or alternatively (but less desirably) along a bus segment using a 1553 stub arrangement. To excise a babbling terminal, the node to which the terminal interfaces, or the bus segment to which it is attached, can be dropped from the virtual bus. The system reconfigures around the faulty device.

The electrical constraint on numbers of terminals that can be interconnected is also eliminated. Since each node now uses active components to provide the electrical isolation between bus segments, the signaling waveform is regenerated at each node. An almost limitless number of terminals can be added without degrading the signal. This does not, of course, overcome protocol limits on the number of terminals, such as occurs in 1553.

To better place this architecture in technical prespective, it is interesting to observe that, except for the protocol limit on numbers of terminals, such a network could be built using 1553 technology for link and node electronics. Existing computers with nearly standard 1553 interfaces could be used to control the net, and any 1553 remote terminals could be attached to the net. Node devices, which are the unique new elements of the architecture, could be fabricated by capitalizing on 1553 microcircuit components.

In addition to overcoming the three primary weaknesses of 1553, inability to interconnect a large number of terminals, damage vulnerability, and vulnerability to a babbling terminal, the network enhances 1553 performance in other ways. While these benefits are secondary and not adequate to justify a change from standard 1553 practice, they are nevertheless significant.

First, unlike 1553 buses, it is possible for the virtual bus to "Y" or branch. Since nodes are active devices, the reflections and impedance mismatches, which preclude this in a standard 1553 bus, are not relevant. Thus, a virtual bus can look like a tree, much as shown in Figure 2.8-4. This considerably loosens topological and routing constraints.

Secondly, multiple buses can be active simultaneously, and the spare or inactive links constitute a shared redundancy pool, able to repair failures in either or both buses. By using this multibus capability, it is possible to set up several buses, possibly partitioning the system according to a natural hierarchy along with dedicated point-to-point paths to link terminals with high-bandwidth requirements. Redundancy is then available inexpensively in the form of a pool of unused links. Figure 2.8-5 illustrates a sample configuration with an active bus and an inactive bus, as well as a dedicated path between nodes A and B. Figure 2.8-6 illustrates an alternative configuration designed to overcome the local damage event which disabled node C.

Configuration control algorithms are quite simple for maintenance of one bus, and become more complex for multiple buses of different criticalities. Configuration-control information or commands are carried to the nodes over the links from a configuration controller at one of the nodes. The links between nodes are fully duplex, and each node continually monitors incoming data on all its links for configuration messages. Although the links are fully duplex (unlike 1553), the virtual bus normally operates as if it were half duplex (like 1553). When a node is commanded to interconnect bus segments, it causes any data arriving on the incoming half of a link to be repeated or retransmitted on the outgoing halves of the other interconnected links. Transmissions arriving on two arms of a "Y" interconnect are combined for retransmission on the third arm. Simultaneous arrivals would produce erroneous bus data on that arm, similar to the situation when two terminals on a bus are transmitting simultaneously. All incoming links are monitored for configuration commands before the data from that link are combined with the data from other links for retransmission. This assures that a node can always correctly receive any configuration commands if the bus fault is outboard of the terminal being addressed. The node immediately inboard of the fault can, therefore, receive the configuration commands necessary to disconnect the fault from the virtual bus.

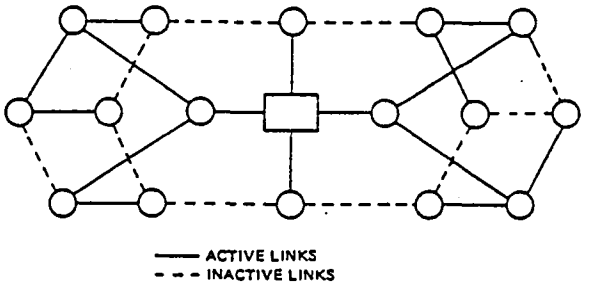


Figure 2.8-4. Virtual Bus With "Y" Constructs.

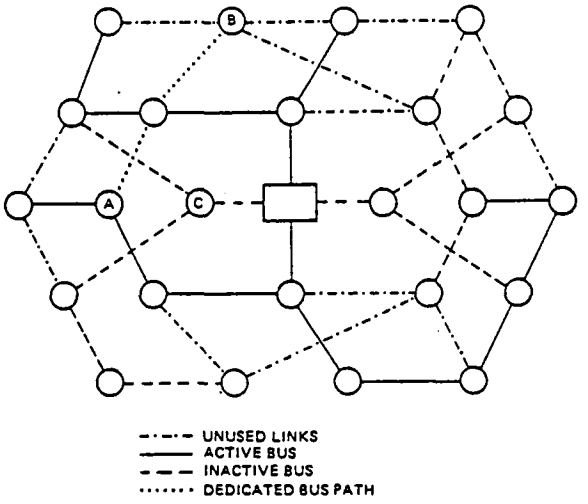


Figure 2.8-5. Multibus Network.

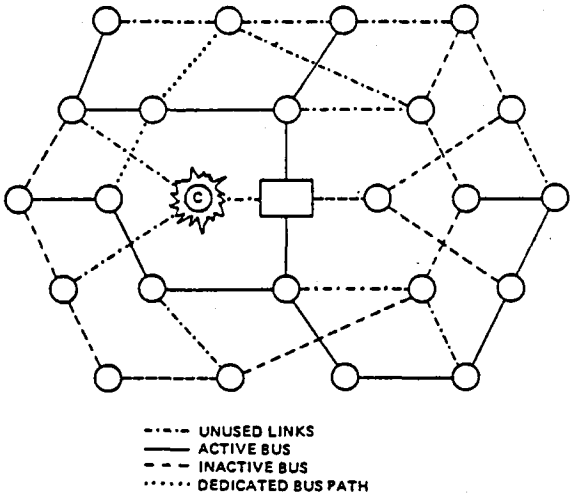


Figure 2.8-6. Reconfigured Multibus Network.

CHAPTER 3

MESH NETWORK DESIGN

The mesh network concept was introduced in the preceding chapter. In this chapter, the elements of design and management of large networks are identified and explored. The chapter begins with a reiteration of the concept.

3.1 Review of the Network Concept

Mesh networks are a means of implementing system-wide multiplexing channels using numerous interconnected channels of short dimension. Figure 3.1-1 shows an example of a mesh network with six nodes, each node serving a distinct subscriber. Two single-port controllers are shown, each of which is capable of managing the network in the absence of interference from the other. Full-duplex links run point-to-point between nodes. The number of links is half the number of link ports. In this case, there are six nodes with three ports apiece, plus two controllers with one port each, giving twenty link ports and ten links.

The network is designed to be configured by switch settings in the nodes. Links may be active, i.e. interconnected with others at the sites of the nodes where they join. Alternatively, they may be idle, i.e. isolated. To establish the multiplex data channel, the controller sends successive messages causing nodes to set switches such that each node has a single link bringing messages outbound from the controller. The other links interfacing with that node may each be directed further outbound or be disconnected. Active links form a tree that does not close on itself. Since each active link arrives at a node with outbound messages, the number of active links is equal to the number of nodes, irrespective of how the links are configured. The remaining links are idle. In the example of Figure 3.1-1, there would be six active, and four idle, links.

In the event of the failure of a link or a node, the network can be reconfigured by its controller by means of messages sent in

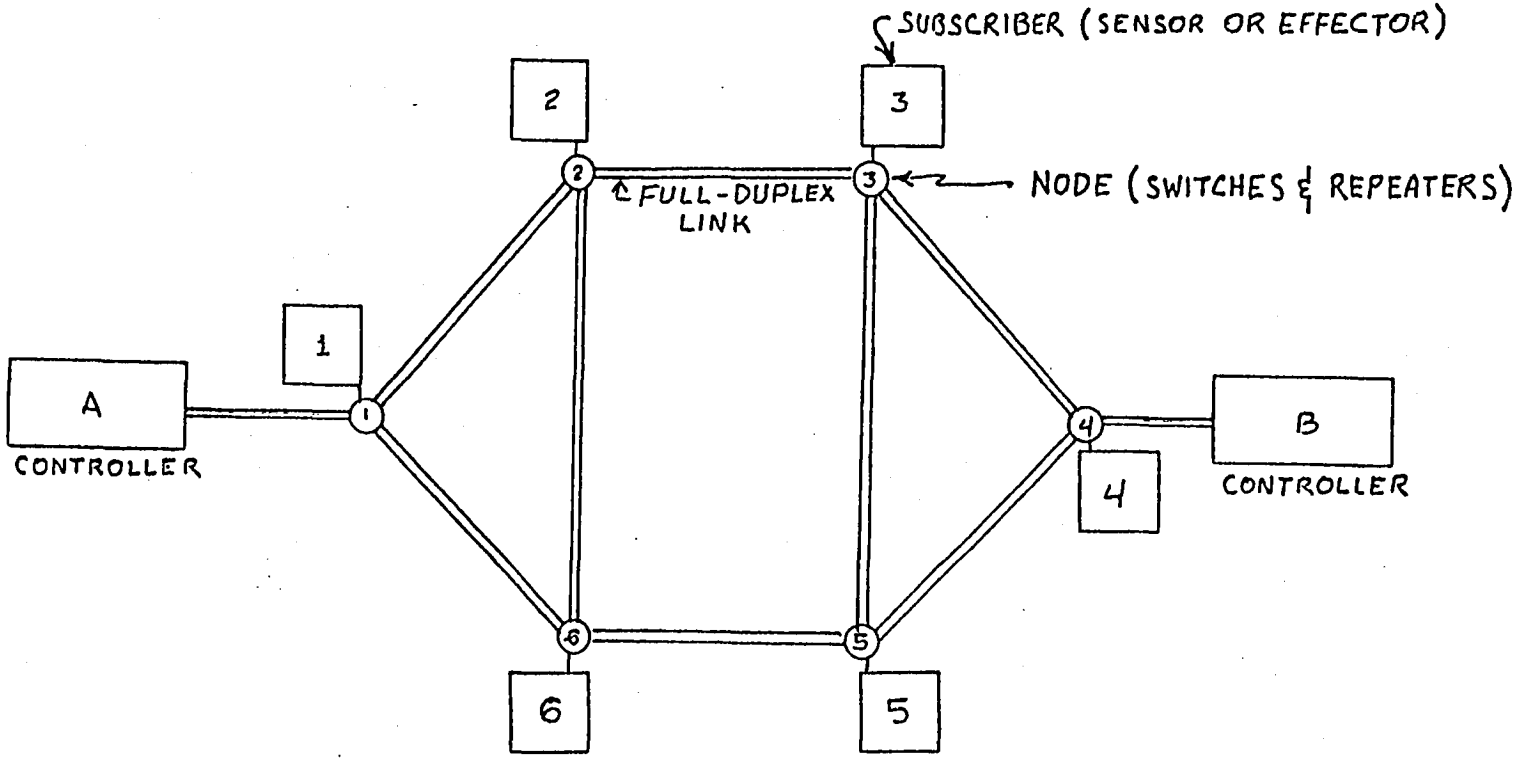


Figure 3.1-1. Network Example.

sequence to various nodes. If more than one failure occurs, the network may or may not be recoverable, depending on the number of failures and the topology of the network. This chapter treats the various design issues that bear on performance, economy, and safety of a mesh data network.

The philosophy underlying mesh network design is along the lines of self-defense. That is, the system does not depend on injured parties to fail gracefully. Rather each surviving party is made able to cope with problems forced upon it by its neighbors, whether the neighbor be injured or whether it be passing along erroneous data from another source. The network incorporates a substantial volume of electronics used for repeating and switching. Its intrinsic redundancy allows it to tolerate multiple failures, however. The degree of tolerance is such that the reliability can be greater than that for a standard multiplex bus, in spite of the fact that the network may use more electronics than the bus. The failures that occur in a network are more of a maintenance concern than a reliability or safety concern.

Networks rely heavily on active control for their survival, and are therefore dependent on software algorithms. This fact is apt to make their acceptance by the air transport industry slower than it might otherwise be. Once active-control airplane designs are accepted, however, this would not seem to be a factor. A more serious potential problem in networks is that a moderate number of passive faults may be able to split a network into two isolated fragments. This problem can be minimized by proper network design.

The redundancy of the network is interwoven with nominal elements, so that it can not be separately identified. The standard multiplex bus, on the other hand, uses a separate replication for redundancy. The bus does not need to reconfigure after a fault, as the replication is already in place and operating. The network, however, needs to reconfigure before it can resume active service. The time required for reconfiguration and recovery is an important concern for performance and safety. If reconfiguration can be accomplished in ten milliseconds or so, there need be little concern. Time-critical nodes can be located where they can be reached more quickly than non-critical nodes.

If reconfiguration time is much longer than ten milliseconds, then some form of error masking may have to be employed. Outright triplication of the network and its nodes for error masking is a

possibility that was envisioned when the network was first conceived. This approach is almost certain to be overly expensive, however. A less-expensive approach to masking would be to grow three or more separate trees in the same network, such that no more than one-third of the system would be affected by a single failure.

3.2 Network Management Principles

Given a properly-designed network, a controller has the task of making the network emulate a branching bus, to the extent that it is possible to do so with surviving protions of the network. Management algorithms are needed for some or all of the following functions:

- Grow
- Regrow
- Verify
- Test
- Dispatch
- Detect
- Diagnose
- Take-over
- Operate

3.2.1 Grow

When the network is in an arbitrary state, as it is at system turn-on, the controller must issue configuration commands before communications can be supported. The only nodes that can be reached for certain are the controller's immediate neighbors. The situation can be visualized with the help of Figure 3.2.1-1. The figure shows one port of a controller (rectangle) leading to one portion of a network. The controller can transmit to node A via link A-11, which is half of a full duplex link, i.e. A-11 only carries data to the node and A-12 only only carries data to the controller. Figure 3.2.1-2 shows node A in greater detail, showing boxes GA1, GA2 and GA3, called "gateman" circuits. Gateman circuits receive configuration commands arriving at their respective ports. When the controller sends a configuration command on A-11, the command is received by gateman GA1, irrespective of what may be happening on A-12 or any of the remaining links. The grow procedure calls for the controller to send a message to gateman GA1 to cause the configuration controller in node A to accept further commands via port A-1, and to reply with acknowledgement and status.

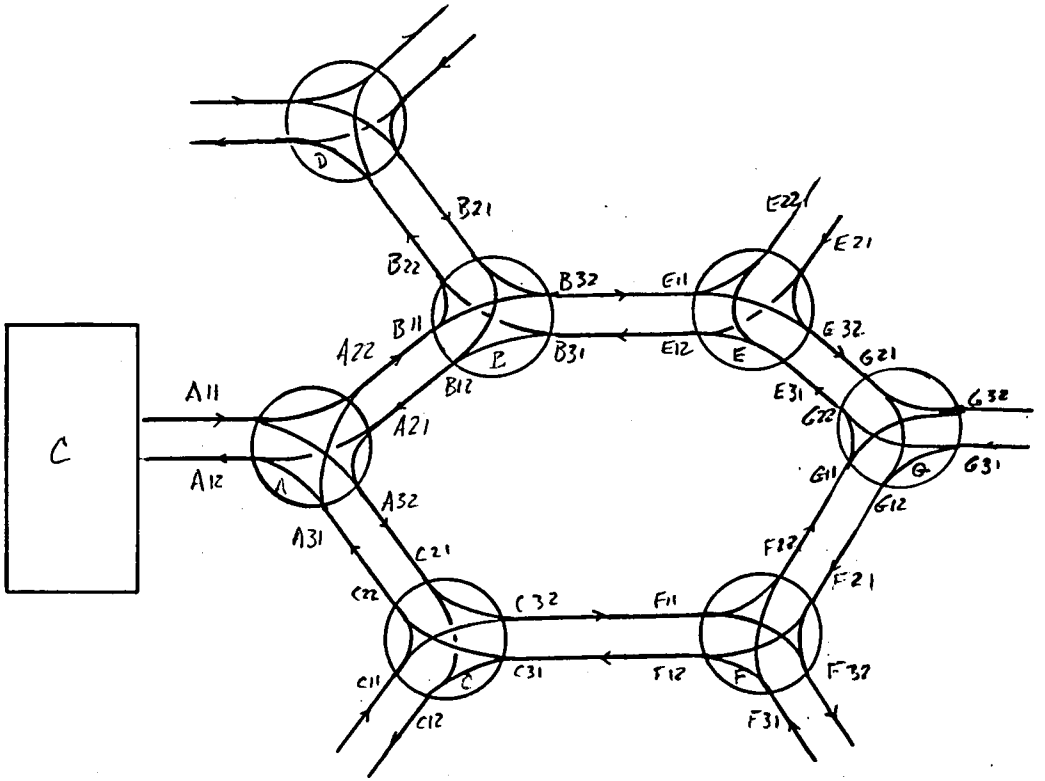


Figure 3.2.1-1. Portion of a Network.

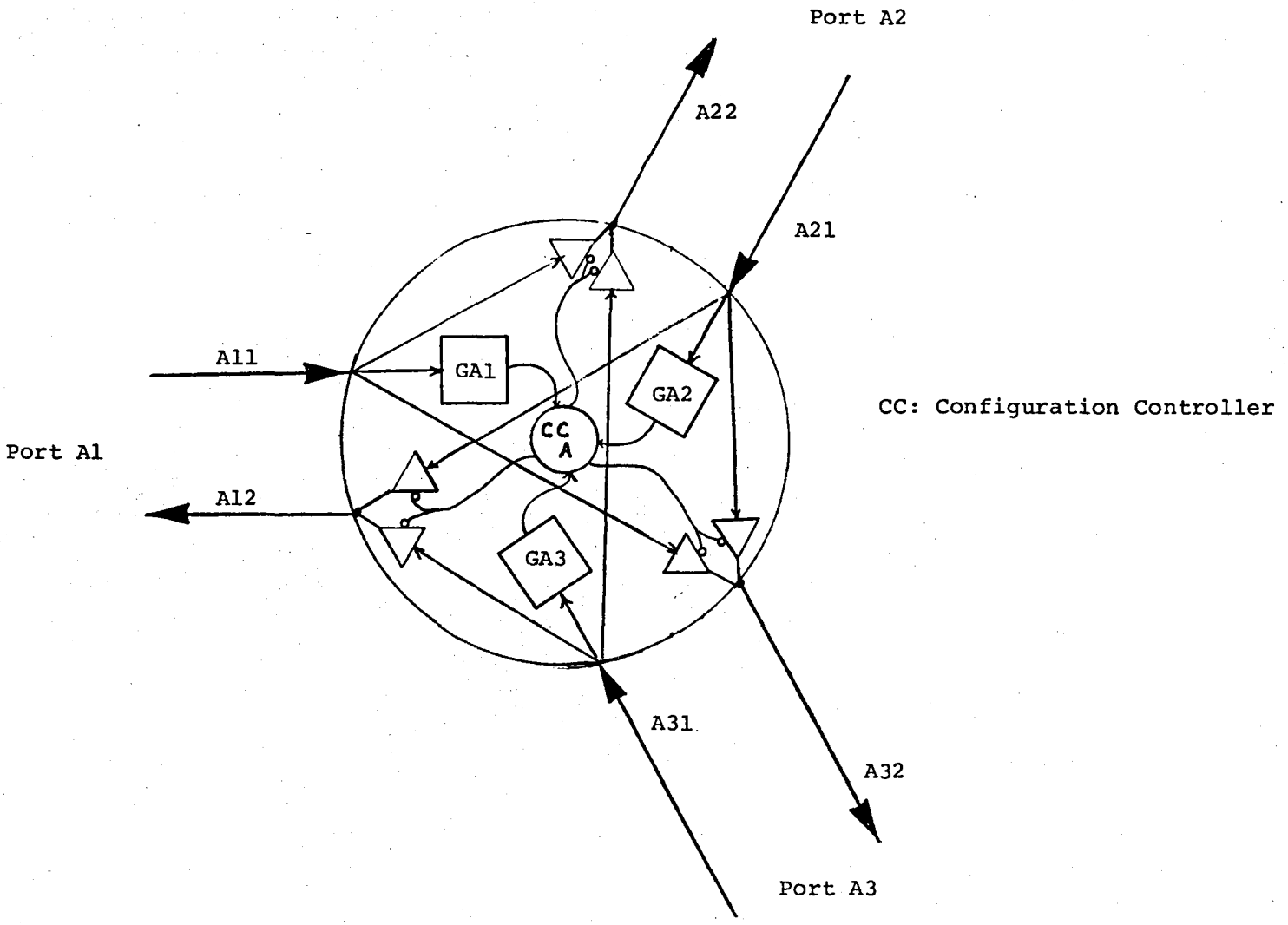


Figure 3.2.1-2. Gateman Circuits In Node A.

This done, the controller seeks to attach node B. It first commands node A to activate the port directed at B, which is port A-2. This is done by enabling repeaters from A-11 to A-22 and from A-21 to A-12. The controller can now reach gateman GB1 directly, and commands node B's attention via port B1.

Next, the controller will seek to attach node C in similar fashion. It first enables repeaters from A-11 to A-32 and from A-31 to A-12. Then it commands gateman GC2 to have node C listen to port C2.

Note that both nodes B and C can now hear the controller and vice versa, but B and C can not hear each other. If it is desired that they do so, then the controller can enable repeaters from A-31 to A-22 and from A-21 to A-33 in node A. This is a system design option, and is not necessary for all networks.

To continue the grow process, the controller attaches node D via B2, node E via B3, node F via C3, and node H via C1, as shown in Figure 3.2.1-3. The process continues in a similar way, attaching neighbor nodes of nodes already attached. Asterisks in Figure 3.2.1-3 indicate "growth points," which are node ports that are in a position to attach new nodes. Note in part (h) of the figure that nodes E and F are growth points toward the same new node, node G. As shown in part (i) of the figure, only one of the two is allowed to grow. The growth point from node F is eliminated when the link from node E to node G is established.

Since growth points define the periphery of the active tree, the definition of the grow algorithm is easily expressed in terms of growth points. A read-only file is maintained showing the physical structure of the network by listing the identity of the neighbor node and port for every node port in the system. Two variable-length writable files are then defined. One is the growth-point list. The other is the list of nodes that have been reached. The grow process begins with only the controller port(s) in the growth-point list, and a null list of reached nodes. The controller removes and reads the top growth point from the list and identifies its neighbor node and port. If the node has already been reached, as determined by the reached-node list, the growth point port is disabled, and the next growth point is taken from the growth-point list. If the neighbor node had not been reached before, the controller tries to establish contact with the neighbor node. If unsuccessful, the controller disables the

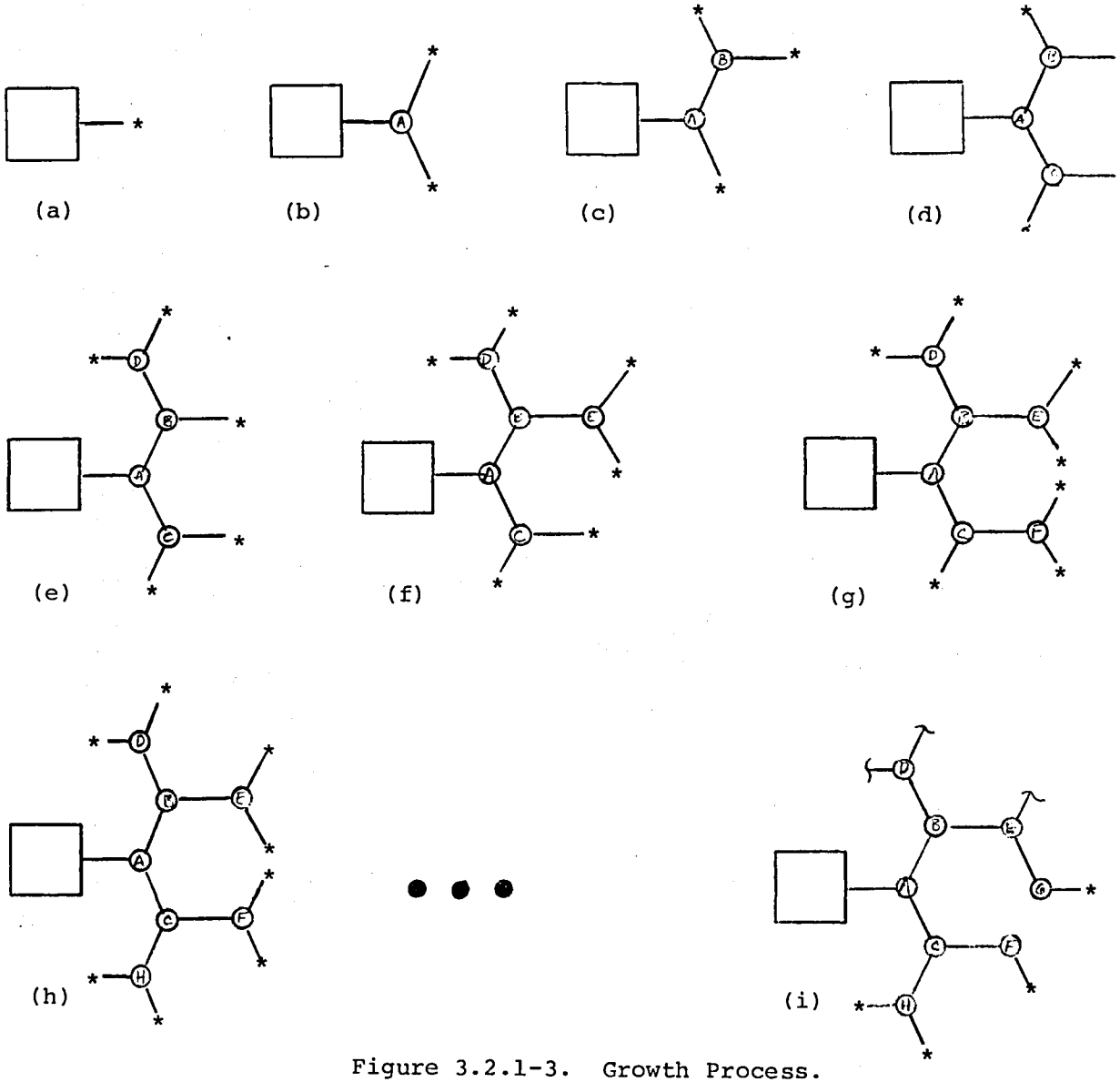


Figure 3.2.1-3. Growth Process.

growth-point port, and takes the next growth point from the growth-point list. The controller can also create a bad-port message at this time for maintenance data. Assuming that contact was successful, however, the identity of the new node is added to the reached-node list, and the identities of the other ports on the new node are added to the growth point list. The next growth point is then taken from the growth point list. When there are no more growth points left on the list, the growth is complete. This algorithm is shown in flow form in Figure 3.2.1-4.

The result of applying this algorithm to a regular hexagonal net with three ports per node is shown in Figure 3.2.1-5. The algorithm branches as much as is allowed by the network topology. This has the beneficial effect of minimizing signal latency. Regular patterns, however, do not permit a great deal of branching.

The grow algorithm can be shown to connect every node for which connection is possible. Imagine an unconnected node that is the neighbor of a connected node. When the connected node was reached, the port facing the unconnected node was identified and placed on the growth-point list. Therefore, when the growth-point list is exhausted, this port will have been tried. If no connection results, it will be because the node is not connectable by that port due to a fault. If the unconnected node is instead a neighbor of another unconnected node, no connection is possible via their mutual link until and unless one of the two nodes becomes connected.

The speed of the grow process is limited by two factors; computational speed and I/O bandwidth. The amount of memory required for minimal tables, lists, and code is relatively modest, and need not exceed 2K words or so. The computational speed bears on the time required to access the tables and lists as required, and to formulate the commands to be sent to the nodes. This is estimated to require something on the order of a hundred operations on a typical computer. On a computer with a speed of a million operations per second (1 MOPS) this would take a hundred microseconds per node, or ten milliseconds for a network of one hundred nodes. This would be marginally acceptable for recovery speed in an active-control transport if every fault were to require a complete grow operation.

With respect to I/O operation, each node will require on the order of 100 bits of I/O to command and verify its configuration. The MIL-STD-1553 bit rate is under a million useful bits per second, so

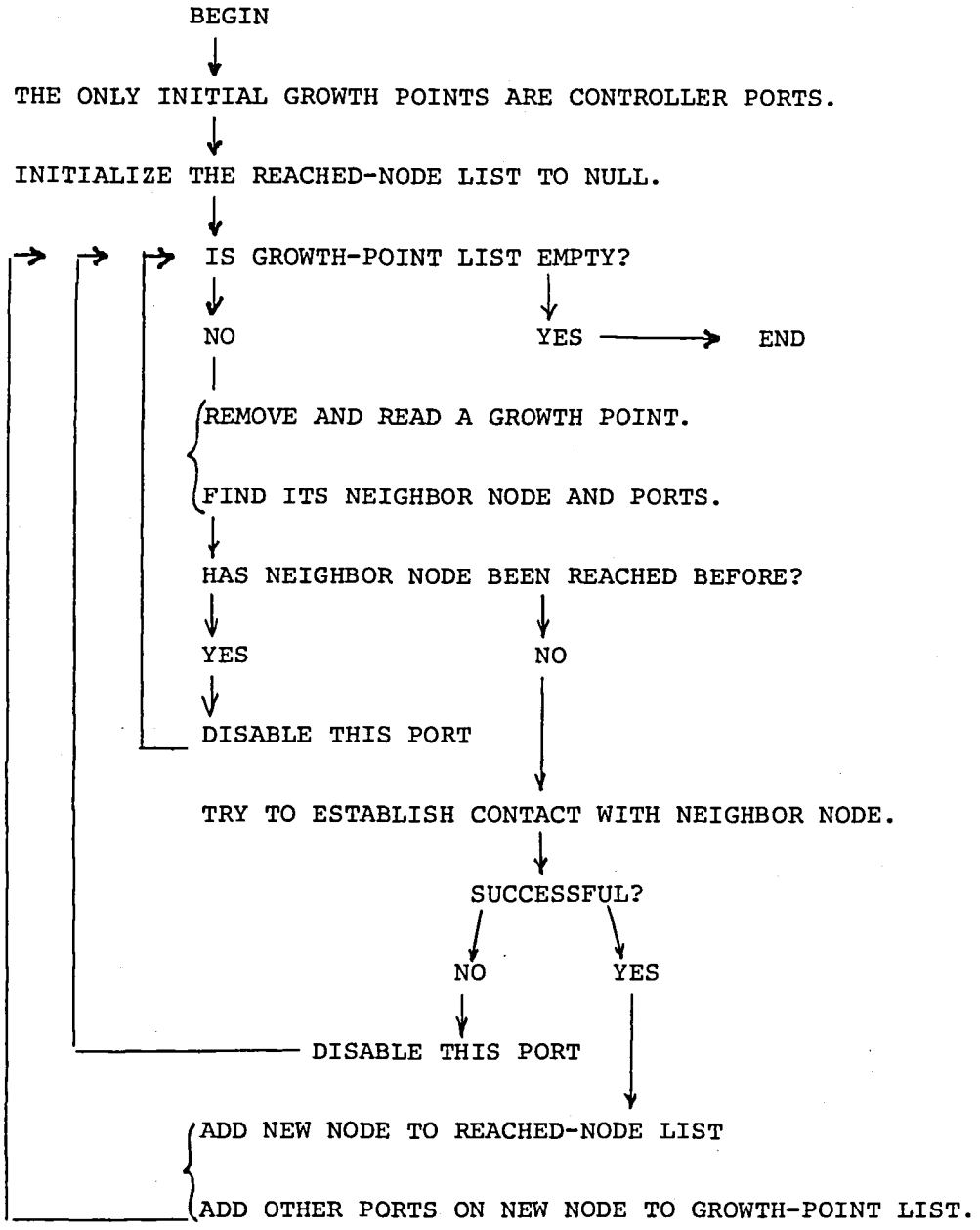


Figure 3.2.1-4. Growth Algorithm.

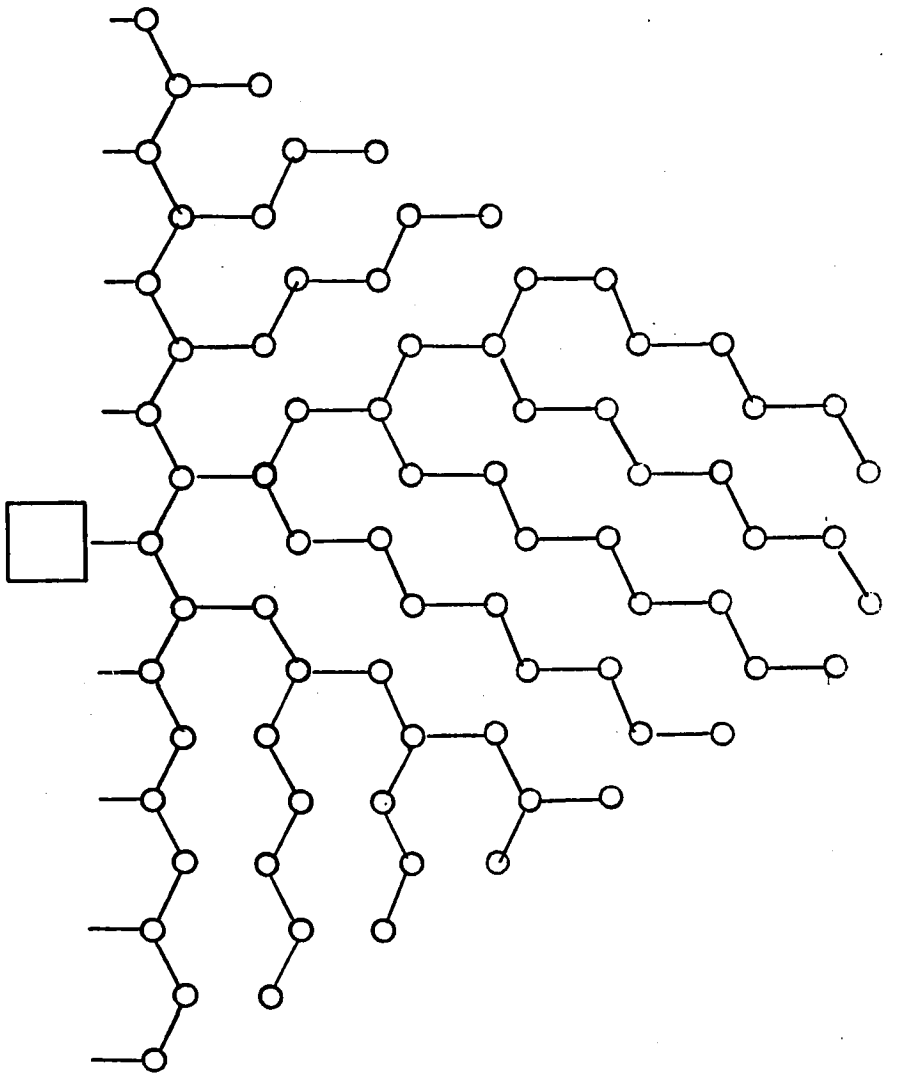


Figure 3.2.1-5. Growth Pattern.

that over a hundred microseconds of communication time are needed as well as, say, a hundred microseconds of computation time. These times would actually overlap to some extent, but an estimate of 200 microseconds per node might be realistic. Again, this is for the initial growth, and does not necessarily mean an unreasonable recovery time.

3.2.2 Regrow

When a faulty node or link is detected and diagnosed, its removal from the system may create a disconnected region of arbitrary size. If no record was kept of the growth pattern during the growth process, the regrow will have to be complete, starting from the controller. The process can be simplified, however, if complete records are kept showing the tree structure. A linked list would be a convenient representation. With such a list, the identities of the nodes in the disconnected region can be found easily. At the same time, the previously idle ports can be identified. These idle ports correspond to idle ports in the connected region. These latter ports may be placed in an initial growth-point list, and the grow algorithm run from this point.

If complete information is not kept, it is still possible to simplify regrow if all unused growth points are identified during the grow process and held in a list. This list can then be used as the initial growth-point list for regrow, and will reduce the time needed depending on the number of ports per node.

3.2.3 Verify

Verification of a growing network takes place incrementally as the grow process takes place. If 1553 protocols are used, each node sends back a status word when it has received its configuration command from the controller. Receipt of a proper status word by the controller does not, of course, guarantee that the node is completely valid. It does provide a high degree of confidence, however, that the node and all intervening nodes, ports, and links are operational. The ultimate test criterion of the network, as for any communication system, is the validity of the information transmitted during system operation between the subscribers and the controller. During normal operation, the subscribers will probably all be accessed sufficiently often to verify the network without further action. The information transmitted would presumably be subjected to acceptance tests such as comparisons, echo

checks, and/or consistency checks. Such testing simultaneously verifies the subscribers and the communications medium.

3.2.4 Test

Verification has denoted the assurance that a grown tree is operating correctly in a network. Test, by contrast, is a process employed to assure that the idle portions of the network are either functional or else previously known to have failed. The aim is to avoid situations where expected redundancy is in fact unavailable. The principal objects of the test process are idle links and ports, i.e. the links and ports not currently active in forming the communication tree. The simplest method of test, conceptually, is to modify the active tree in such a way as to replace active links with idle links. If this is done systematically, all of the links and ports will be rotated in and out of active service periodically.

Link rotation can be costly in terms of bandwidth in a large network, particularly since each link needs to be tested in two different directions so that each half-link can be proven to operate in both inboard and outboard senses. This is necessary in order to exercise all gateman circuits. A so-called "modify" algorithm attempts to select links for rotation so as to minimize the overall expense (in some sense) of a test cycle. Smith [5] has defined one such algorithm, making use of a tree map.

An alternative means of testing may be implemented whereby special test messages can be sent over idle links. One such approach would be to equip every node with a test message receiver. In order to conduct a test of an idle link, the controller would send a message to one of the link's two nodes commanding it to connect the link to that node's test message receiver. The controller would next command the other node to attach the other end of the link to the active tree. Now a test message would be directed to the first node's test message receiver, and a verification reply would be sent back. The link would then be disconnected again and tested in the reverse direction.

3.2.5 Dispatch

Before an airplane can be dispatched, it must be ascertained that its equipment is sufficient in number and correct in function. This is not to say that the vehicle must contain a fully operational complement, because one of the system requirements is for maintenance postponement. The complement must rather satisfy a minimum equipment list, where it is understood that to be counted as "present," an element must meet certain functional requirements.

In the case of a mesh network it is not easy to define a minimum data communication equipment list that is truly minimal. It is difficult, as discussed in Section 3.3, to measure the connectivity of a mesh network, and it is this connectivity that provides the redundancy upon which flight safety depends.

A pragmatic approach in this case is to define a dispatch criterion that is easy to measure. The ease of measurement is obtained at a certain expense, i.e. dispatch may be denied to a flightworthy system. The probability of such an undeserved denial, however, can be kept quite low without greatly complicating the dispatch criterion.

Consider a dispatch criterion that requires there to be no more than one fault in the network. This criterion is extremely easy to apply, but it is apt to deny dispatch too often. Taking only nodes into account, for a ten flight-hour day and a hundred nodes, one would expect to have two failed nodes on the same airplane once every 100 days if the nodes have MTBF's of 10,000 hours. On the other hand, if node MTBF's are 100,000 hours, the double-fault situation would occur only once per 10,000 days, which would be negligible.

Consider now a less stringent dispatch criterion that allows multiple faults. As long as the failed elements are independent of one another, and as long as they do not violate the minimum equipment list for subscribers, the safety would be no worse than for the single-fail criterion. One problem is to define what is meant by "independence." This definition becomes contingent on the specific network geometry, including the number of ports per node, and the connection topology. Suffice it to say, however, that for any specific network, it is reasonable to define independence in terms of minimum distance

among failed nodes. For certain regular topologies, such as a Cartesian grid, for example, the measurement of distance can be made trivial if the nodes are numbered and identified in a helpful fashion.

It would seem to be unlikely to have to tolerate more than three or four failures, in which case the distance criterion could be set at several nodes, say four, without creating more than a negligible probability of false denial of dispatch. Any attempt to tolerate higher numbers of failures would have to take into account the probability that numerous failures can sever a sizeable fragment from the network.

3.2.6 Detect and Diagnose

Because the mesh network requires active reconfiguration in order to tolerate faults, it is necessary to discover and locate rapidly any faulty situation that poses a threat to the communication system. In most cases, this is a trivial task, since most faults cause gross failure symptoms, such as lack of response to a command, incoherent "babble", or a violation of parity or framing constraints.

The most difficult kind of fault to detect would be one that produces no obvious symptom. An example would be a node that fails to connect its neighbor, yet answers coherently when the neighbor is polled, sending spurious data. This kind of fault could be generalized to the point where a single node postures as many nodes, to all of which it is supposed to be connected. Although this kind of failure mode can be rendered highly unlikely by proper design, it can not be ignored.

The ultimate acceptance criterion of any communication system is the reasonableness of the data it transmits. This will always require that a flight-crucial system be designed so as to be distrustful of sensor data received and effector data transmitted. Reasonableness testing requires control techniques that produce estimates of expected sensor behavior. When data that passes parity and framing checks produces disagreement and confusion in the reasonableness tests, notice can be served on the network controller to that effect.

This leads to the question of diagnosis. When a fault symptom is strongly correlated with a particular node, the network can be reconfigured so as to reach that node via a different port. This can be accomplished by declaring a failed link and executing a regrow. If the symptom disappears, the link remains declared as failed. If the

symptom persists, additional links are declared failed until all ports have been tried. Finally, the node will be cut off from the network, and if the symptom disappears, the node will be declared failed.

If the symptoms do not correlate with a particular node, it may be that a link or node fault is affecting a sizeable branch of the active tree. If all of the grow data was kept, it would be possible to interpret symptoms according to a map of the active tree. Suppose a node fails close to the controller, but the first symptom comes from a node far away, reached via the failed node. Attempts to reconfigure the symptom node may produce more fault symptoms. If not, other symptoms will arise in time. Now the diagnostic process would be invoked, using the map, where an attempt would be made to verify the active path from the controller to the nearest node for which a symptom has been received. This process will locate the faulty node, and it will be tested as in the preceding paragraph.

So far it has been tacitly assumed that the fault symptom is passive and consistent. An active, or babbling, fault causes disruption of the entire connected portion of a network, making it necessary to initiate a diagnostic process analogous to the grow process. The first step in this process is for the controller to command its immediate neighbor to disconnect its other ports one at a time, to see when the fault disappears. If it does not disappear, then the guilty node has been found. Otherwise, the process continues with successive neighboring nodes until the faulty node is reached.

The diagnosis of an intermittent fault condition requires more ambitious strategies than for consistent faults. The network must not be kept disconnected while waiting for a fault symptom to occur. Therefore all symptoms may be assumed to be obtained in a fully connected state. Diagnostic information must be obtained purely on the basis of changes in the active tree structure between symptom events. Use can be made of multiple ports in controllers for this purpose. By moving single nodes or clusters of nodes from one sub-tree to another, decisions can be made regarding the location of a fault, according to which controller port sees which symptom.

In order to simplify and standardize the diagnostic process, we might assume that the network will be managed by a multiport controller, and that normal growth will produce several independent active subtrees, one for each controller port. Symptoms, wherever or however obtained, may be assumed to point to one sub-tree at a time.

The controller would break up the suspect sub-tree and apportion its nodes among its original sub-tree and the other sub-trees. It may determine the culprit in the process of doing this. If not, the next symptom would be awaited. Eventually, the culprit would be found by successive fragmentation of the suspect sub-tree's node clusters. This strategy requires an intelligent management algorithm that maintains a map of all active trees and sub-trees. The same algorithm may also be given the job of periodic reconfiguration for test purposes.

3.2.7 Take-Over

The assumption is made in this report that a single fault-tolerant computer will possess the entire configuration authority for the airplane. This is not a necessary assumption, however, from the point of view of network management. A network is amenable to management from multiple distinct controllers, provided that they do not contend for control at the same time, i.e. that they fail passive.

Multiple controllers can operate either in a standby replacement mode or in a load-sharing mode. Algorithms required for multiple controller operation are the standard management algorithms discussed in this section, plus a reliable means for deciding when to assert control on the basis of perceived activity on the part of the other controller. The safety issue regarding control assertion is unsolved in general, however, and will always be difficult to evaluate in specific systems.

3.2.8 Operation

The final function to be discussed in this section is normal operation. A grown network has the logical attributes of a bus designed for command-response protocols. Depending on details of design, the subscribers may or may not be able to hear one another, but in any event they can all hear the controller and vice versa.

A controller with multiple ports normally broadcasts its messages from all ports at the same time, though not necessarily synchronously. Response from subscriber nodes are heard on a single port only. The controller may use its knowledge of the active tree assignments to access the response from the right port. Alternatively, it could form the logical or of responses from all ports, or else poll the ports one at a time, to obtain the response message.

In variant modes of operation, the different ports might operate in parallel, depending on the nature of the controller. This would be a

possible means of operating a "channelized" system in which different redundant sensor and effector channels are assigned to different controller ports. Substantial throughput advantages are possible here.

3.3 Network Topology

The attributes of a network are dependent upon the manner in which links join nodes. In particular, if a single link break separates the network into two disjoint fragments, then the reliability of the network is bounded by the reliability of the one link and the two nodes that it joins. One part of network design is to ensure that no such narrow neck exists unknown to the designer. It might be imagined that this poses no great problem, for a narrow neck would surely be obvious upon inspection, or so it might be thought. Figure 3.3-1 is offered as a means by which the reader can gauge the difficulty of finding a narrow neck in a network. The cutting of a single link separates this ten-node network into two five-node fragments. Trying to find such a narrow neck in a network of one or two hundred nodes would be vastly more difficult for a human to do by eye. It also turns out that to do the job by computer is difficult, more so than one might expect from the simple manner in which the problem can be defined; i.e., "where can the network be cut into two fragments with the fewest link cuts?" Just how difficult this computation can be is discussed in Chapter 6, where an algorithm is described for solving the problem. In this section, it is taken for granted that an arbitrary network topology is difficult to certify with respect to its connectivity. (Connectivity is defined as a measure of the number of cuts required to cleave the network in two.) This section is rather concerned with methods for laying out regular networks for which the connectivity is either obvious or easily found.

3.3.1 Connectivity

The connectivity of Figure 3.3-1 is equal to one, because the elimination of one link, specifically the link between nodes 2 and 5, separates the network into two five-node sub-networks. If the links were to be rearranged as shown in Figure 3.3.1-1, the connectivity would be equal to two. Clearly, the links from nodes 3 to 4 and 8 to 9 join two sub-networks together. A connectivity of three (with a one-node fragment) is shown in Figure 3.3.1-2. In both of these last two figures, the links have been arranged so as to minimize obscurity, which was purposely not the case in Figure 3.1-1.

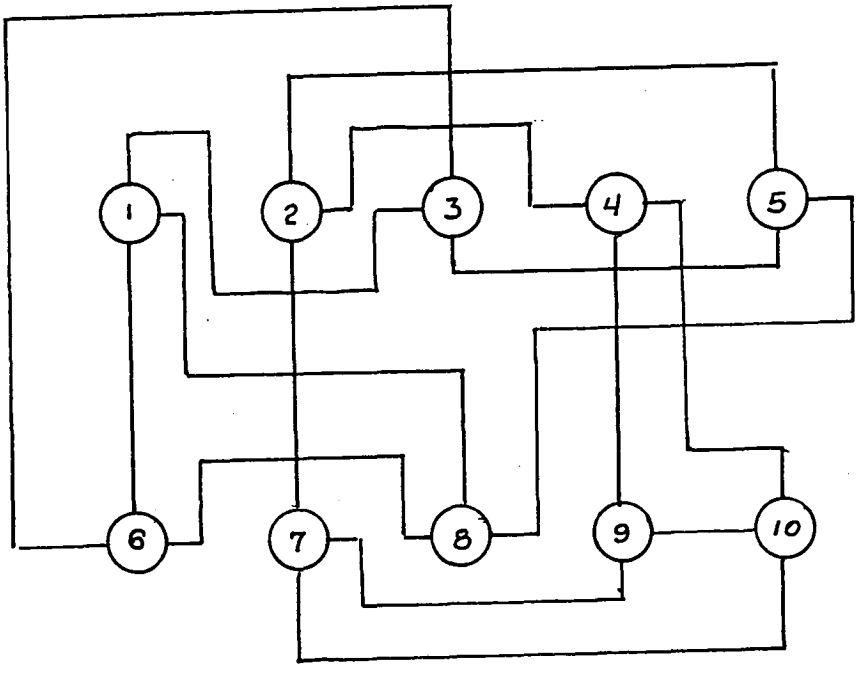


Figure 3.3-1. Connectivity Example.

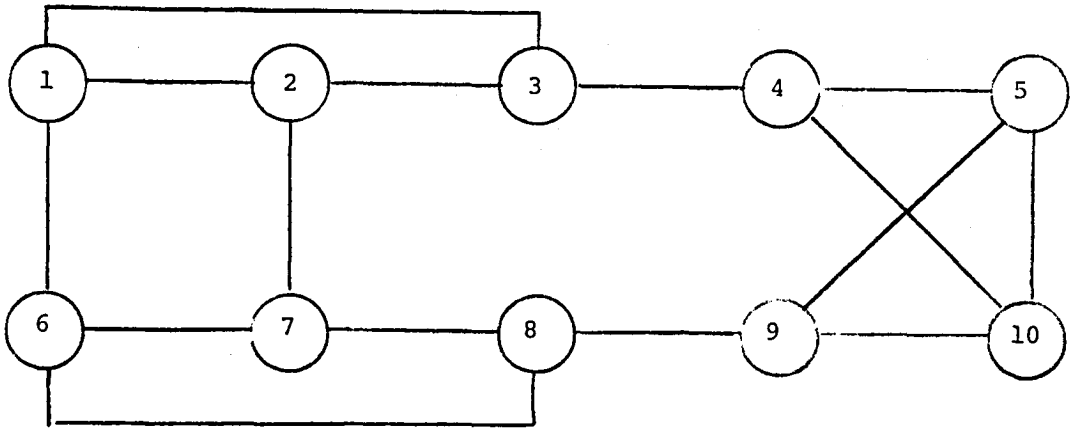


Figure 3.3.1-1. Network With Connectivity of Two.

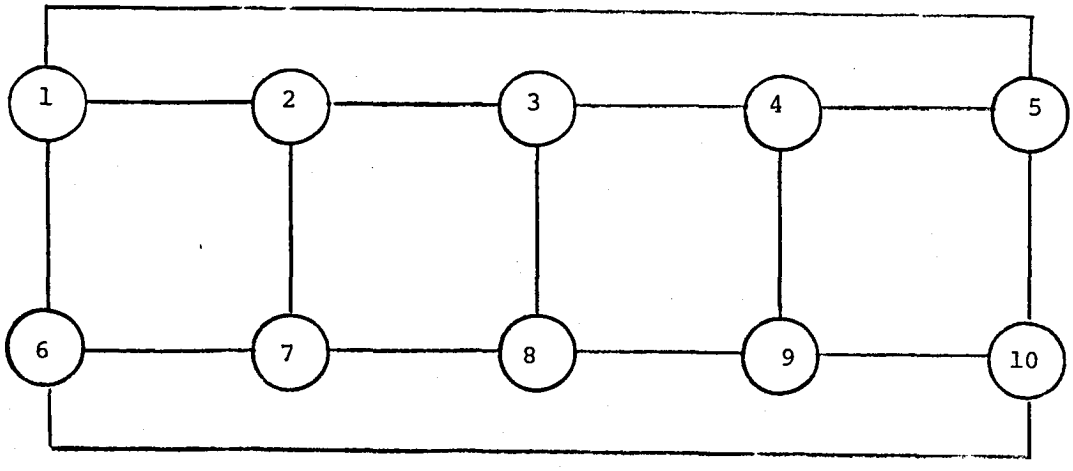


Figure 3.3.1-2. Network With Connectivity of Three.

In Figure 3.3.1-2, the connectivity is three, because a single node can be isolated by three link cuts. This is true for each of the ten nodes. To isolate more than one node at a time, it is necessary to make at least four cuts. This is a small example of a desirable property in certain larger networks, i.e. the number of cut links needed to isolate a group of nodes can be made to increase with the size of the group, up to a certain point.

As is evident from the example of the network in Figure 3.3.1-2, the connectivity of a network will never exceed the number of ports per node, or, more specifically, the minimum number of ports per node if different nodes have different numbers of ports. Connectivity in the strictest sense is therefore of concern only when it is lower than, rather than equal to, the minimum number of ports per node. As a design criterion, it is reasonable to require that the connectivity be equal to the minimum number of ports per node. Beyond this, other geometrical criteria may be applicable. Whereas it is desirable that the number of cuts to isolate a fragment increase with fragment size, this will not necessarily happen. Figure 3.3.1-3 shows several example fragments. Part (a) of the figure shows that four cuts will ordinarily be required to isolate two nodes. Parts (b) and (c) show that the number of cuts required to isolate three and four nodes can be five and six, respectvely, if no closures are made. If loop closures are made, however, as in part (d), (e), and (f), the number of cuts required to isolate will be less than before. The absence of tight loops does not imply high connectivity for the network as a whole, but it satisfies the second criterion wherein it becomes difficult to sever large fragments.

3.3.2 Multiple Paths

Before continuing the discussion of link geometry, it is appropriate to digress upon a point related to system bandwidth. The remedy for insufficient bandwidth is to have multiple channels, assuming that the single-channel bandwidth has been made as high as is practical. The question is often raised in a network (as it is in many other cases of system redundancy) as to whether idle linkage can be used in such a way as to provide increased performance, i.e. bandwidth, in the absence of faults. The answer is that, in principle, more than one tree can exist at a time. Moreover, it is possible for trees to intersect, so that subscribers can have access to more than one tree at a time. The problem is that it may or may not be feasible

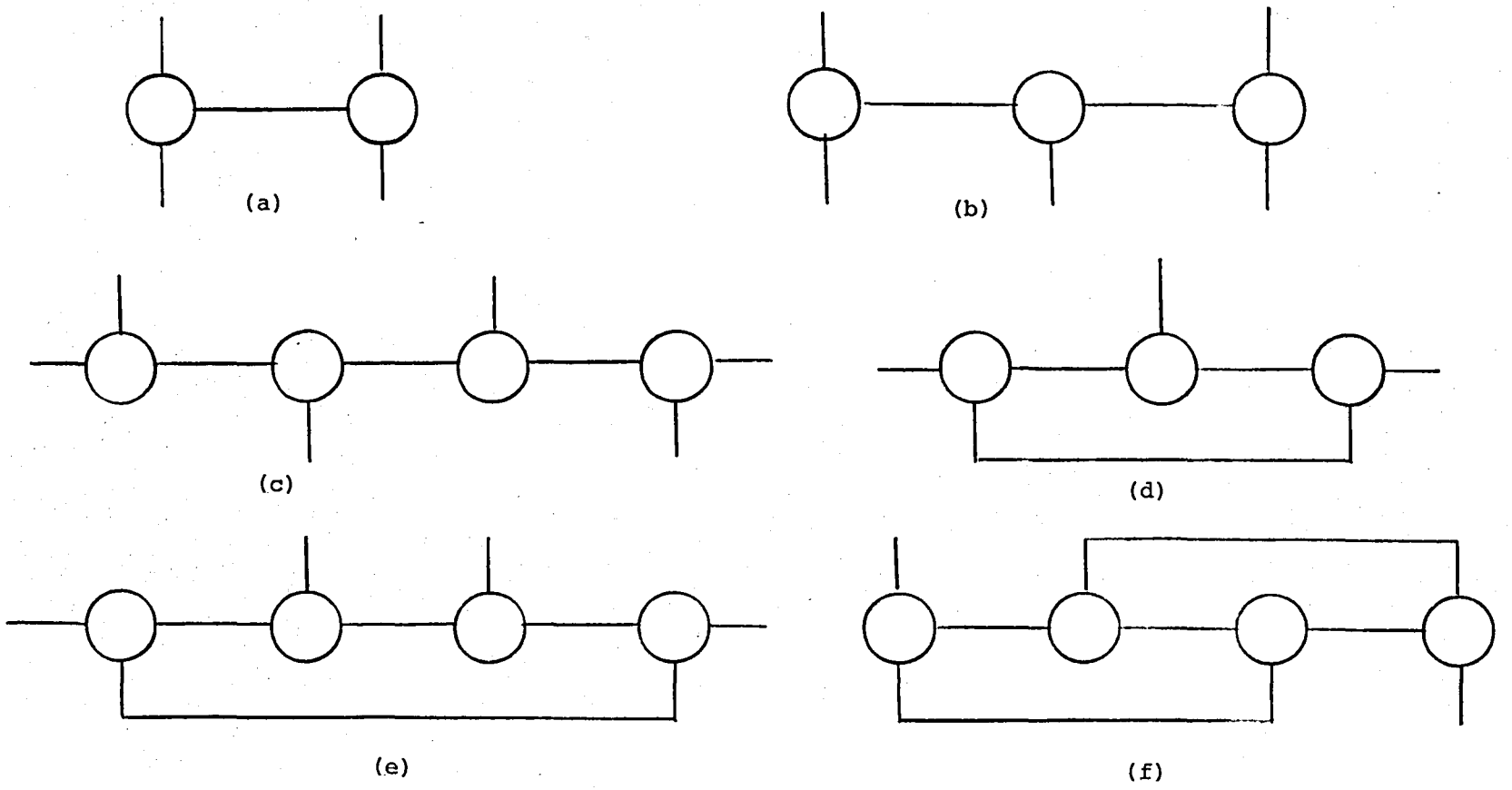


Figure 3.3.1-3. Examples of Network Fragments.

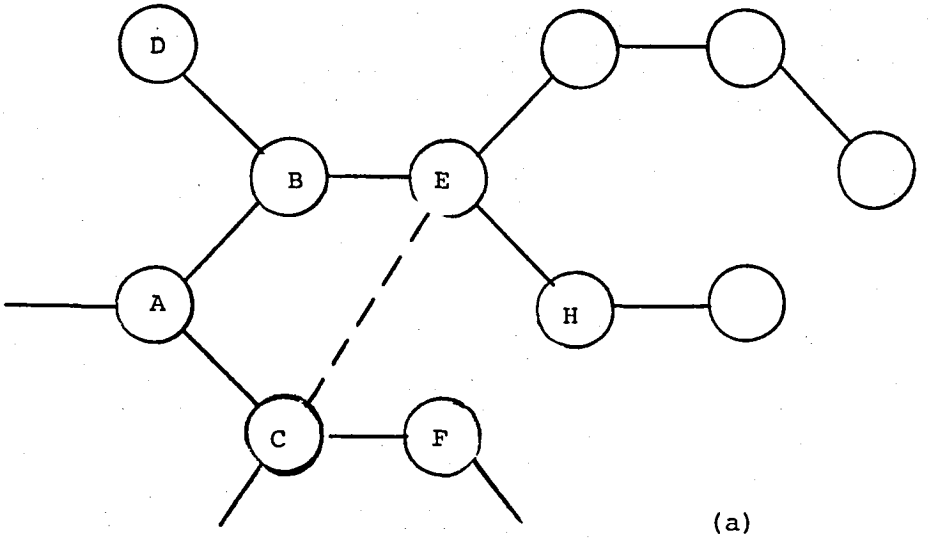
to support a desired degree of flexibility within reasonable economic bounds.

Suppose, for example, that two nodes communicate with one another at a high rate. They can be joined by a dedicated link, as shown in Figure 3.3.2-1 (a) in the form of a dashed line between nodes C and E. If this link should fail, however, this path would have to be replaced in such a way that the principal tree could still reach nodes C and E. The solution might be similar to the one shown in part (b) of the figure, where nodes F and H become waypoints without requiring extra linkage. This is made possible by placing nodes F and H at extremities of the tree as, for example, by reserving the growth points needed to support the dotted path.

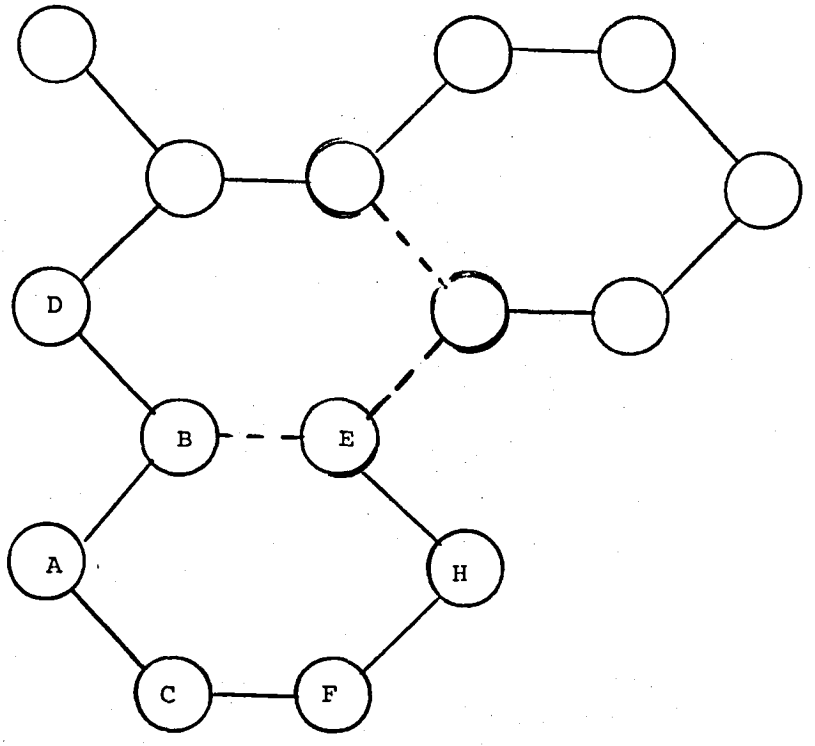
Figure 3.3.2-1 (b) is also suggestive of the possibility that the extra path might be a bus joining all four of nodes C, E, F, and H. This and numerous other things are possible, but they do put a strain on resources, especially if the extra communication channels cover large distances. Either sufficient dedicated links must be provided between distant nodes, or else the alternative path possibilities must be made sufficiently rich to support all paths under the appropriate fault hypotheses. Figure 3.3.2-2 shows in abstract form how dedicated linkage can be used as the primary means to establish multiple paths. In case of failure of the dedicated links, reliance is placed on the ability to grow replacement paths. This fails to utilize idle links for forming multiple paths, but it at least draws on idle links for redundancy, sharing them with the primary path in this regard.

If dedicated links are not used, then all paths must compete for linkage even in the absence of faults. Figure 3.3.2-3 shows a hypothetical node with six ports. Three are used to form a branch in the primary tree, two are used to form a waypoint in another path, and one is left for substitution in the event of failure. This is not to say that six ports per node are absolutely required for two-path operation, but if distance and flexibility are desired in the absence of dedicated paths, the number of ports per node may need to be significantly greater than for a single-path network.

Before leaving the subject of multiple paths, it has not been determined whether a grow algorithm exists for more than one path, that is guaranteed to find a multiple path solution if one exists. The single-path grow algorithm, of course, is guaranteed to find a tree to



(a)



(b)

Figure 3.3.2-1. Multiple Path Examples.

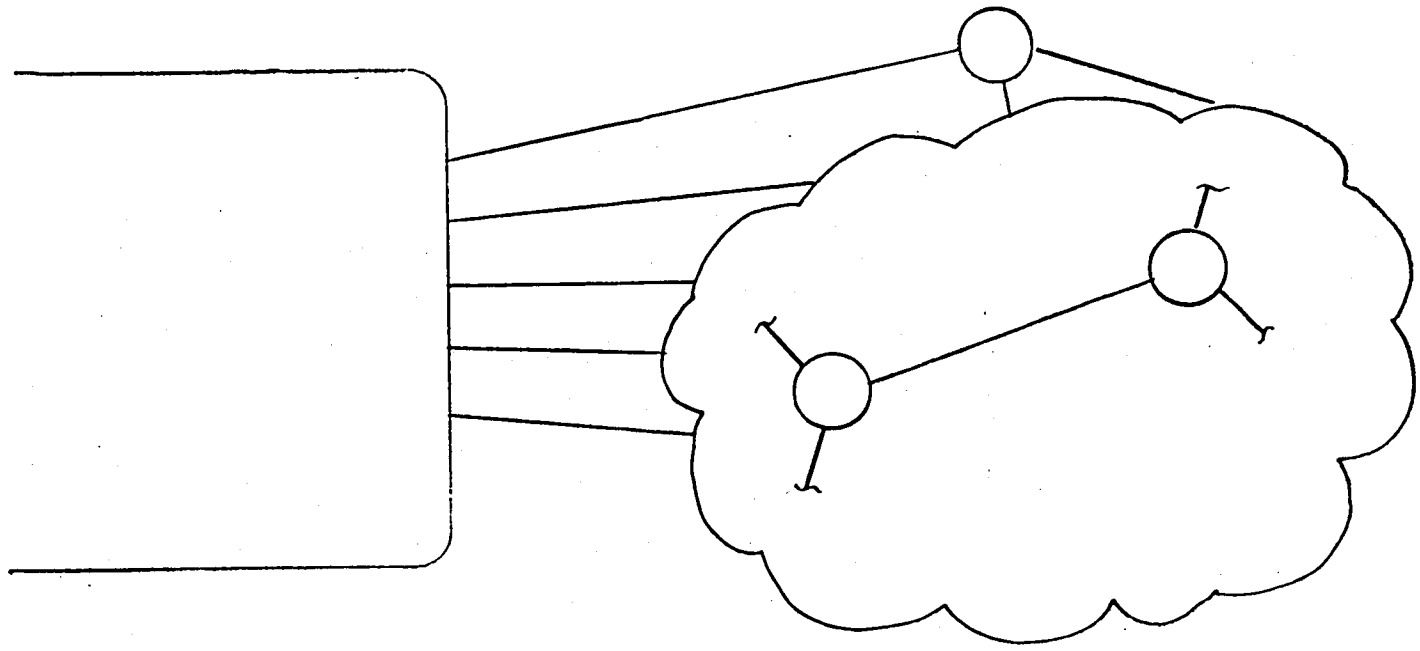


Figure 3.3.2-2. Dedicated Paths.

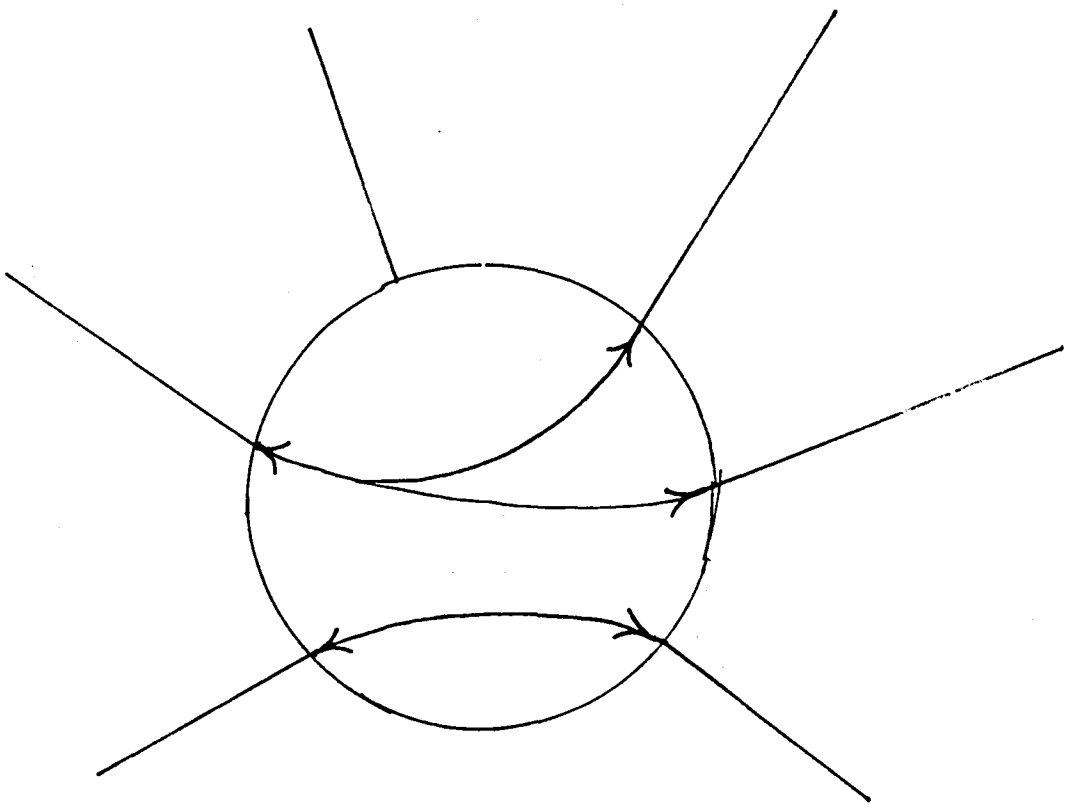


Figure 3.3.2-3. Six-Port Node.

all reachable nodes. If paths can be ordered in terms of critiality, then a pragmatic course would be to grow the most critical path first.

3.3.3 Regular Geometries

The network shown earlier in Figure 3.3.1-2 possesses the attribute of regularity; that is, a symmetry exists such that every node is connected in an identical fashion. If the network is rotated in any direction, it looks the same. Regular geometries are advantageous in that their connectivities can be made immediately obvious, by using simple geometrical constructs.

For three-port nodes, a simple regular geometry in one dimension is the one used in Figure 3.3.1-2. A two-dimensional regular geometry is based on hexagonal tiling, as shown in Figure 3.3.3-1 (a). Part (b) of the figure shows a regular two-dimensional tiling for four-port nodes, and part (c) for six-port nodes. A three-dimensional "tiling" for six-port nodes is shown in part (d). Note the tighter loops in (c) than (d). It is easier to isolate a fragment in (c).

The figures in Figure 3.3.3-1 are deficient in that they possess edge discontinuities, at least as shown here. In order to be truly regular, the figures must close on themselves in the next-higher dimension the way a toroid closes a rectagular sheet, for example. There are, of course, some regular polyhedra that form closed surfaces. One of the more interesting of these is the dodecahedron, as sketched in Figure 3.3.3-2 (a). Exactly twenty nodes of three links each can be arranged in a regular network this way. The icosahedron in part (b) of the figure is a complementary structure to the dodecahedron, formed by placing a face on every node of the dodecahedron and a node on every face. This leads to a network of twelve five-port nodes. Both of these patterns are, unfortunately, more interesting than useful.

Another interesting pattern is shown in Figure 3.3.3-3. This three dimensional array of three-port nodes has loops of length twelve, as opposed to the hexagons of Figure 3.3.3-1 with loop-length six. This pattern is not intrinsically closed, and must be folded at the edges into a "hypertoroid" to close it.

Toroidal patterns afford the means of making regular closed patterns of more or less arbitrary size. One such pattern is shown in Figure 3.3.3-4. This pattern can be rotated in either the left-right or the up-down dimension without altering the perceived pattern. The toridal array in Figure 3.3.3-5 uses three-port nodes. The tiling is

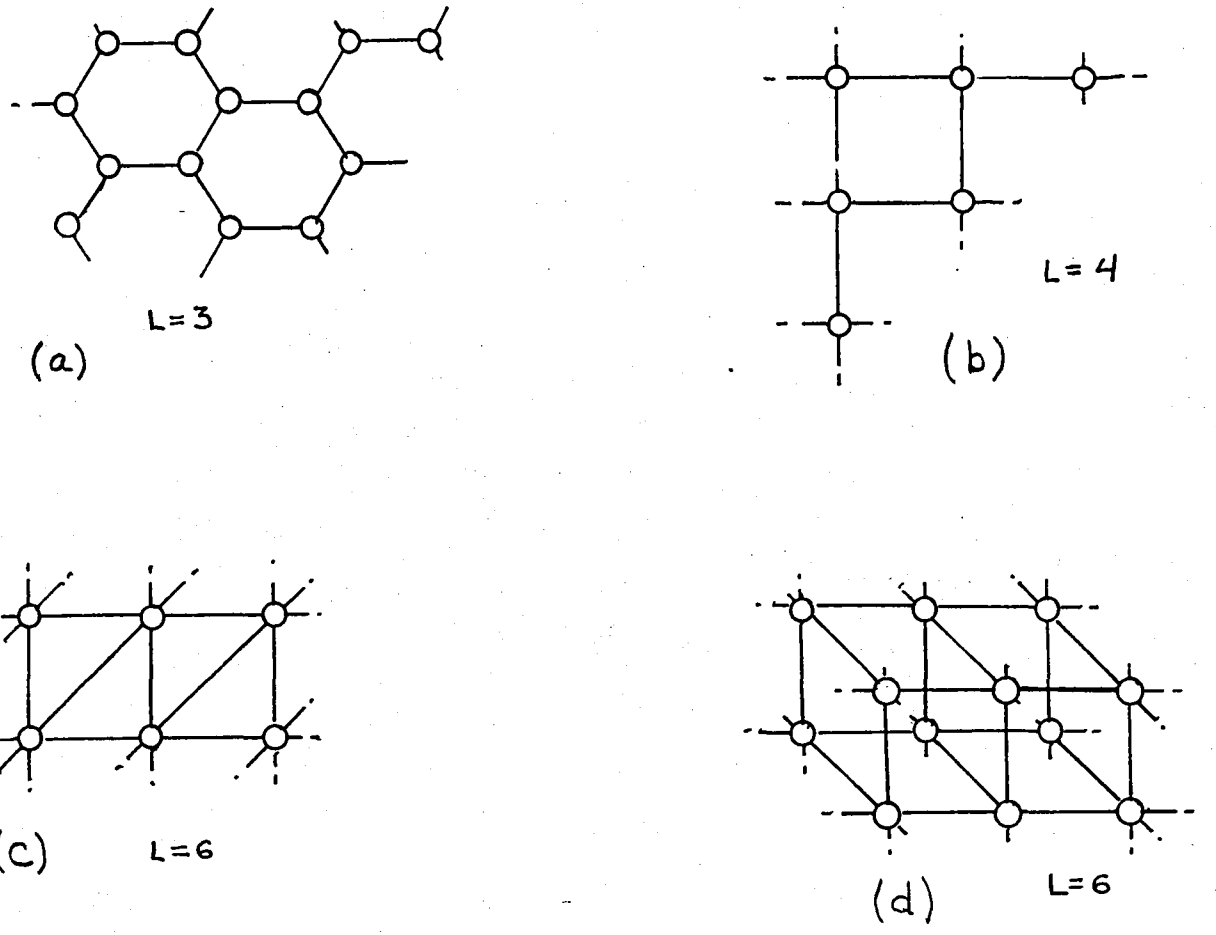
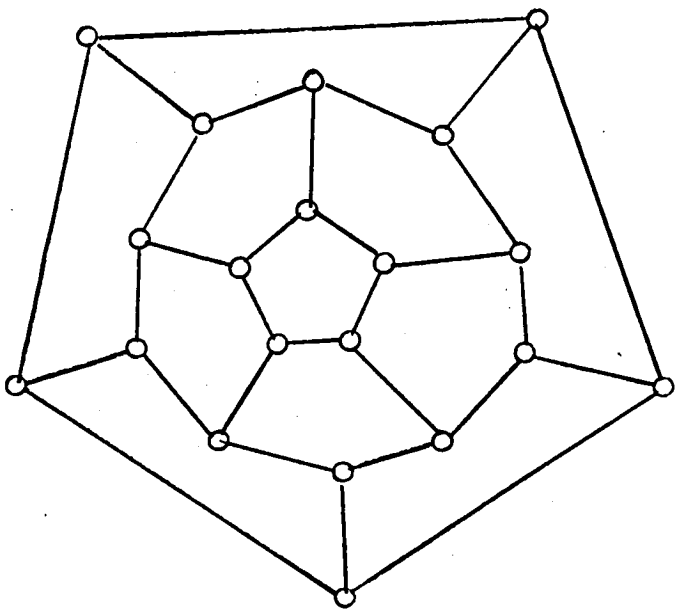
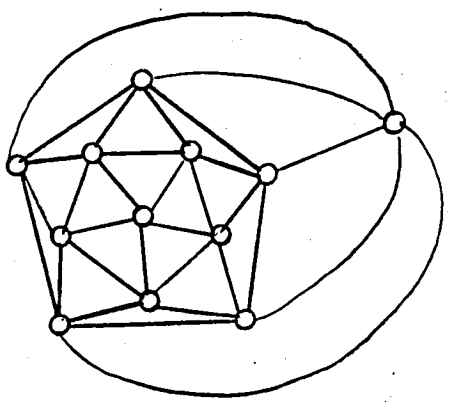


Figure 3.3.3-1. Regular Geometries.



(a) Dodecahedron
L=3 N=20



(b) Icosahedron
L=5 N=12

Figure 3.3.3-2. Polyhedral Regular Geometries.

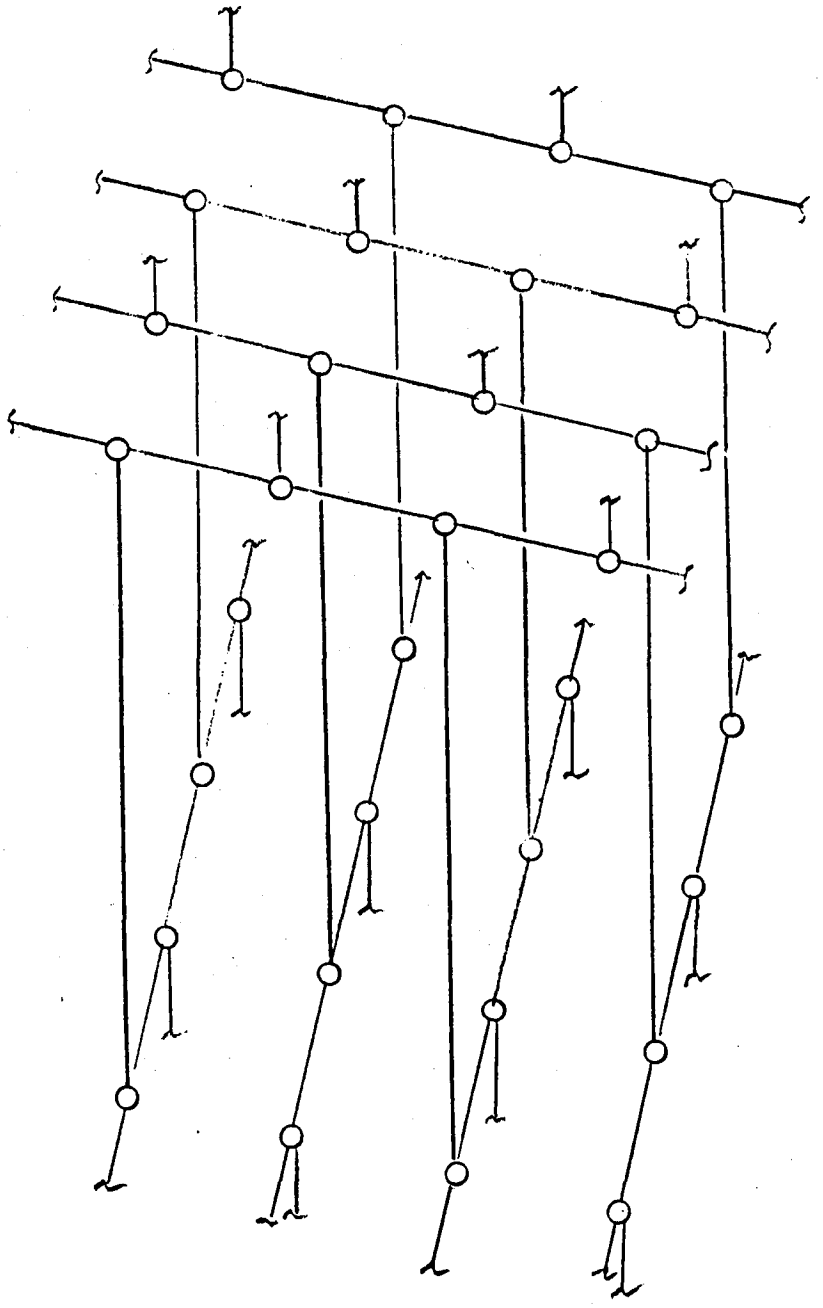


Figure 3.3.3-3. Three-Dimensional Array.

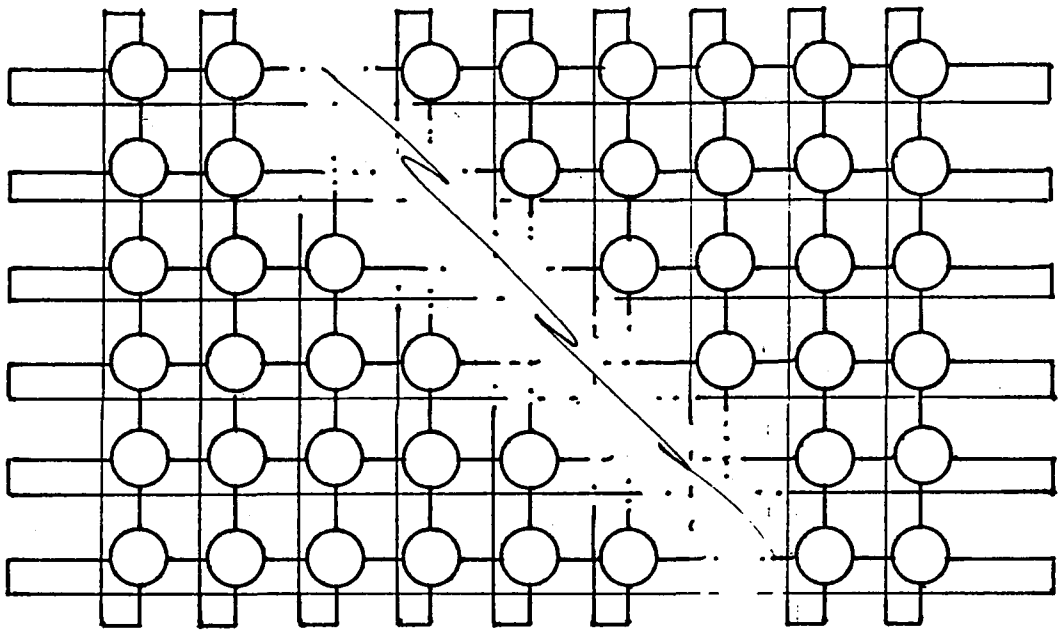


Figure 3.3.3-4. Regular Toroidal Pattern.

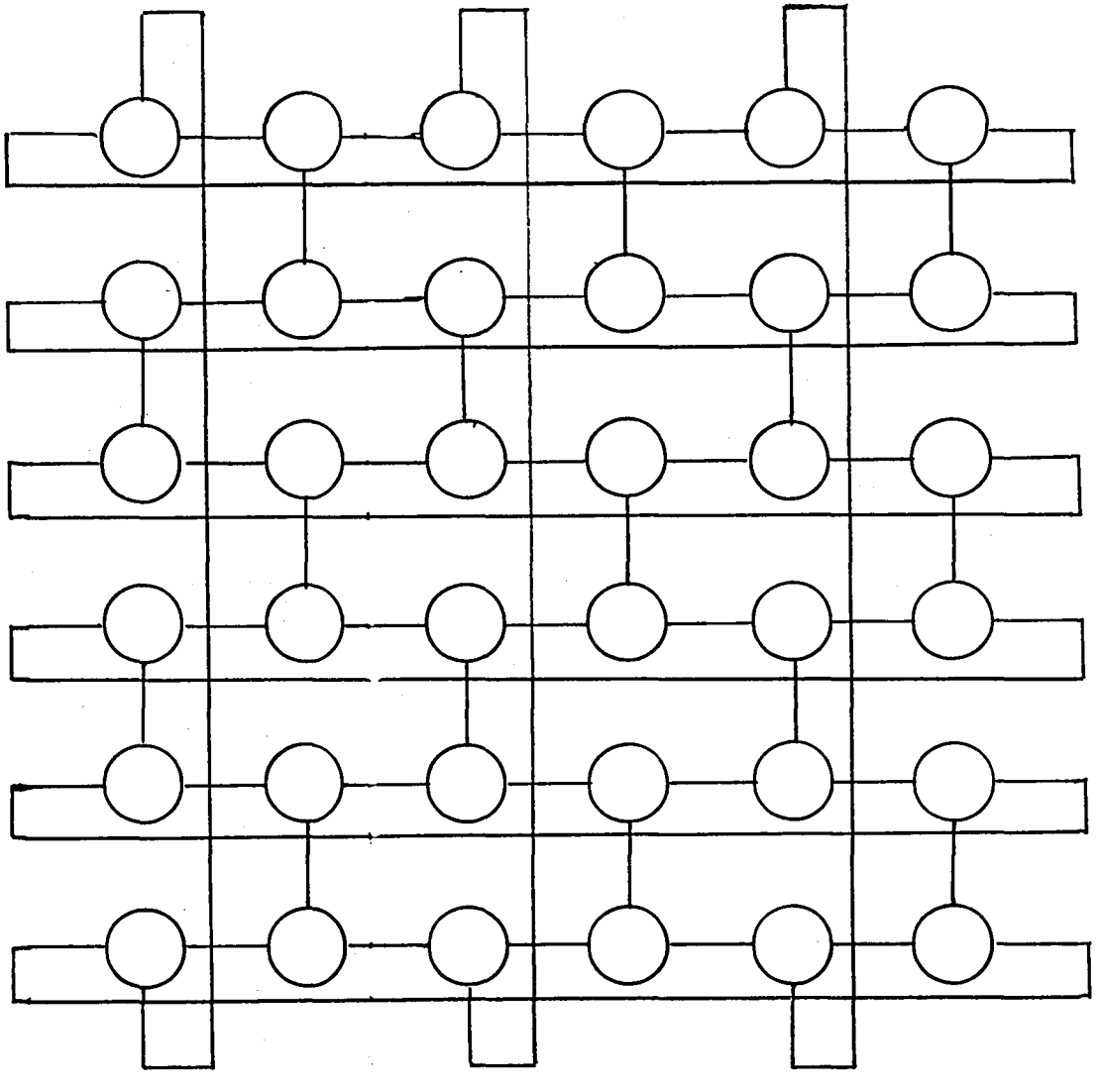


Figure 3.3.3-5. Hexagonal Toroid.

hexagonal, with the vertices constrained to lie in horizontal lines.

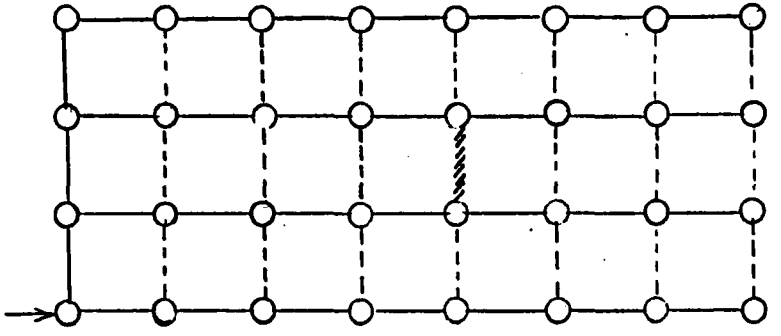
Figures 3.3.3-6 and 3.3.3-7 show how local multiple paths can be accommodated in square and hexagonal regular networks. The (b) parts of these two figures show how the private channel can be reconfigured following failure of the dedicated link. The principal channel is also reconfigured, of course. Some double link faults prevent successful reconfiguration, but most can be tolerated.

3.3.4 Semiregular Geometries

There are two principal reasons why networks that are slightly irregular are of interest. The first, and most important, is that a regular net becomes irregular the moment it is injured, i.e. contains a fault. Multiple faults can increase the degree of irregularity, even to a point where the connectivity is no longer easy to determine. Another reason for considering semiregular geometries is that the regular ones are not necessarily convenient or practical to implement in an airplane.

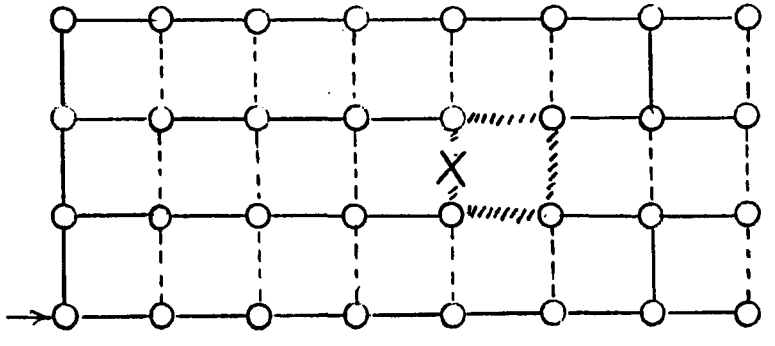
The impact of a faulty link or node on the soundness of a network is primarily local. Each neighbor node is threatened by the loss of one of its access ports. Additional faults in the vicinity could isolate one or more nodes. Suppose a three-port node fails in a hexagonally tiled network. Each of three neighbors is reduced to two ports. The probability of isolating one of these ports is predominantly determined by the probability that two additional faults occur, presumably in the other neighbor nodes of one of the threatened nodes. In Figure 3.3.4-1, nodes 7, 9, and 13 are threatened by the failure of node 8. If nodes 2 and 6 fail, then node 7 will be isolated. The consequences of this series of mishaps depend on which subscribers were assigned to the nodes. From the standpoint of network reliability alone, the loss of "innocent" node 7 in this case is considered to aggravate the situation already imperiled by the failure of three nodes. As a practical matter, the loss of the three nodes may already have been catastrophic. It also may not, depending on how nodes were assigned to subscribers.

At any rate, the joint probability of such an event occurring, given that the airplane was dispatched with node 8 failed, is roughly three times the square of the probability of a single node fault. For nodes with a 10,000 hour MTBF and a ten hour flight, the joint probability is of the order of 3×10^{-6}. If the node MTBF is 100,000 hours,



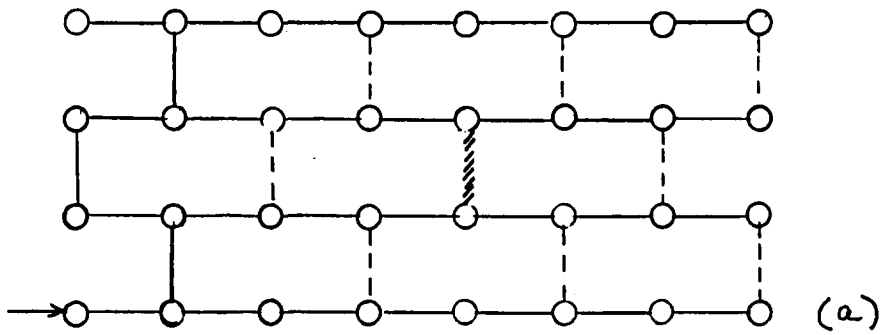
(a)

Private Channel



(b)

Figure 3.3.3-6. Private Channel Recovery - Four-Port Nodes.



Private Channel

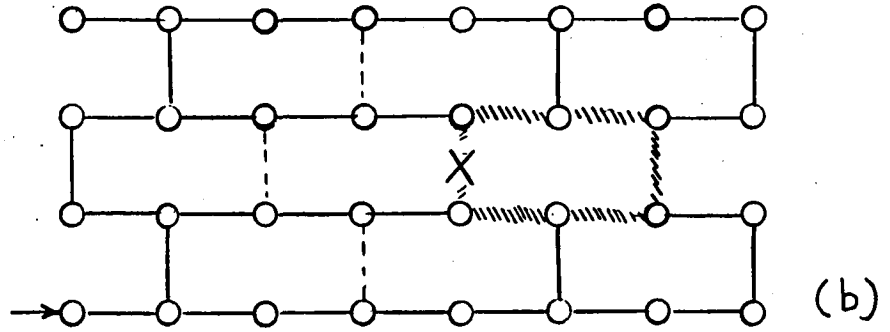


Figure 3.3.3-7. Private Channel Recovery - Three-Port Nodes.

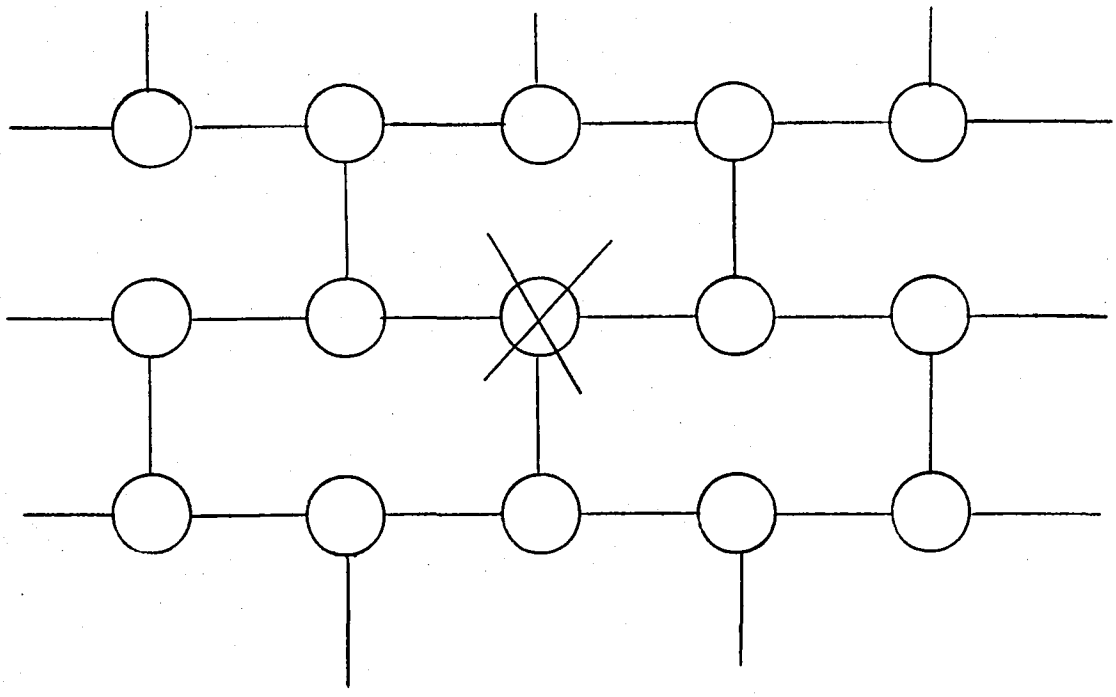


Figure 3.3.4-1. Injured Net Example.

the joint probability becomes $3x10^{-8}$, which is becoming close to an acceptable value for a catastrophic condition, given that the probability of having the initial fault is of the order of 10^{-2} or 10^{-1}. It may be desirable or necessary to use four-port nodes unless MTBF's can be made extremely high.

If the network contains two faults at dispatch time, it is possible that they might both threaten a single innocent node, making it moderately probable that the node would be isolated in flight. This situation is easily avoided as long as the dispatch criterion requires all functioning nodes to have at least two valid access ports. Consider, however, the case where nodes 8 and 1 in Figure 3.3.4-1 are initially failed. Now, if nodes 3 and 6 should also fail, both nodes 2 and 7 will be isolated. The initial faults were distance 3 apart in this case. Distance 3, therefore is probably insufficient to establish fault independence in a network of three-port nodes. Distance 4, on the other hand, is probably sufficient. For four-port nodes, distance 3 is probably sufficient. One can compute approximate probabilities that a network is dispatchable by estimating the joint probability that successive faults are at a sufficient distance from one another. This is discussed in Chapter 6.

Some proposed geometries are made semiregular from the start. One reason has to do with the awkwardness of assigning a toroidal regular net to an airplane. This can result, for example, in having tens of links between each wing and the fuselage, which is more than necessary, and costly. Another possible reason, not considered likely would be to structure the network along the lines of the desired growth tree in order to limit the distance of nodes from the controller. This would be more practical if the network never had failures. When provisions are made for reconfiguration, much of the desired attribute is lost.

To illustrate the last point, Figure 3.3.4-2 and 3.3.4-3 illustrate the result of applying the grow algorithm to regular networks of three-port and four-port nodes, respectively. Note the long runs of non-branching strings in each case. A semi-regular network, shown in Figure 3.3.4-4 is arranged so that thirty nodes are all within distance four of the top node using maximal branching. In case of a fault near the top, however, the longest distance can be as large as eight, following a complex reconfiguration. Figure 3.3.4-5 shows a modified version of the previous structure where provision is made to overcome such

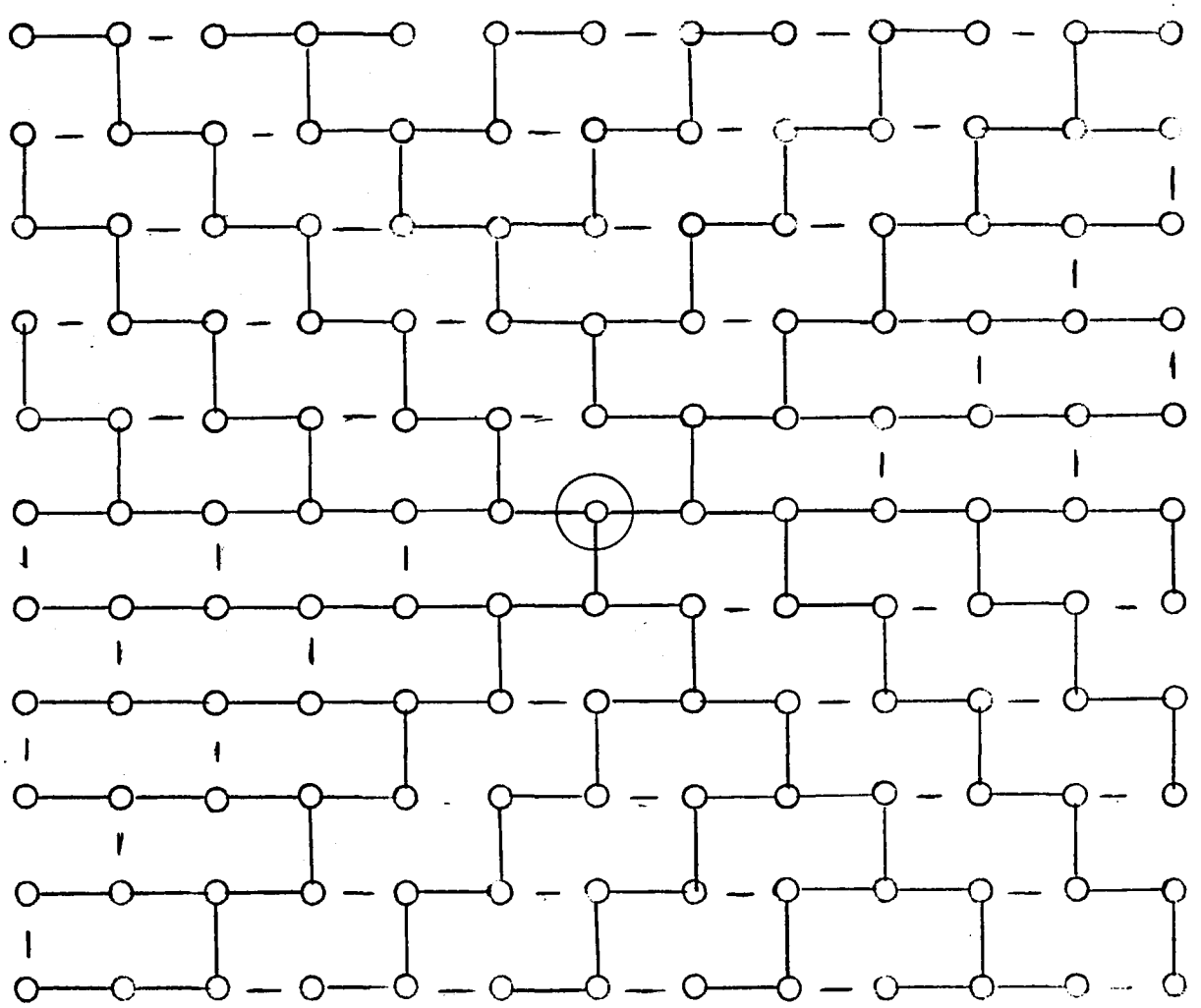


Figure 3.3.4-2. Growth Pattern From Central Node - Three-Ports.

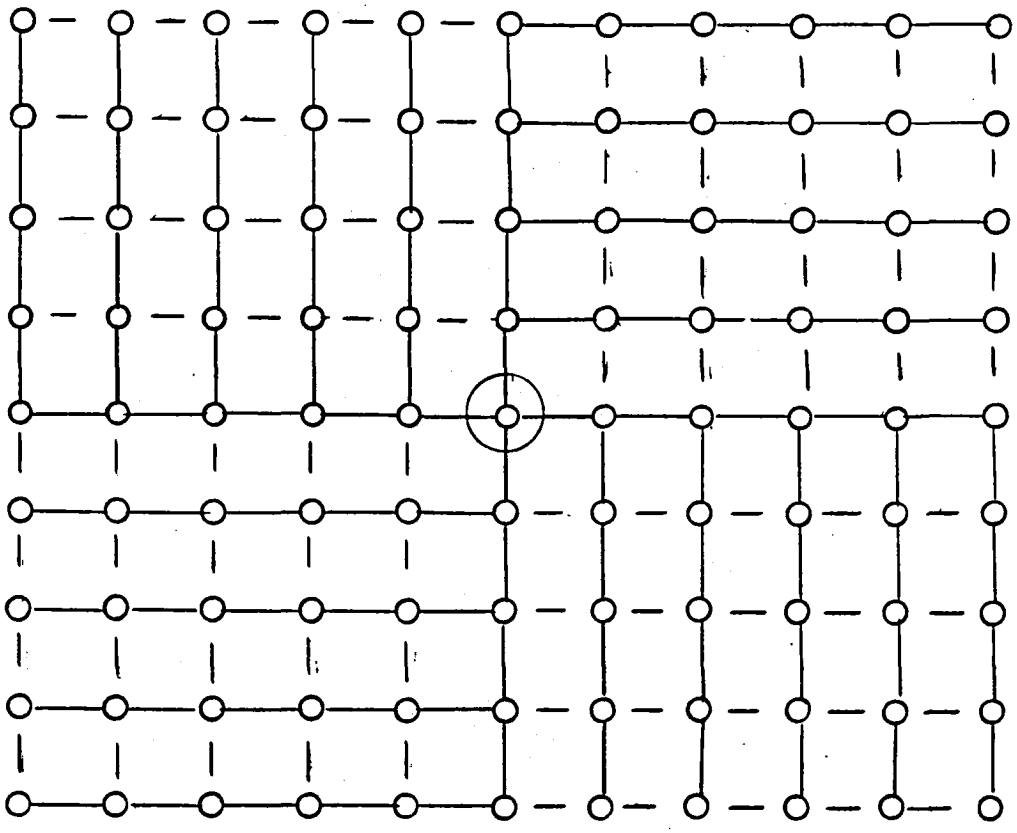


Figure 3.3.4-3. Growth Pattern From Central Node - Four-Ports.

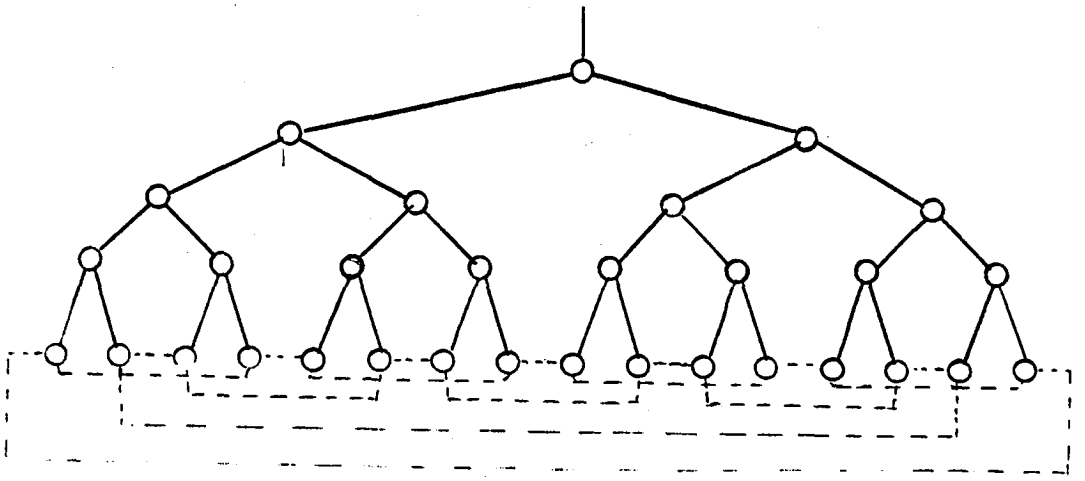


Figure 3.3.4-4. Semiregular Tree Network

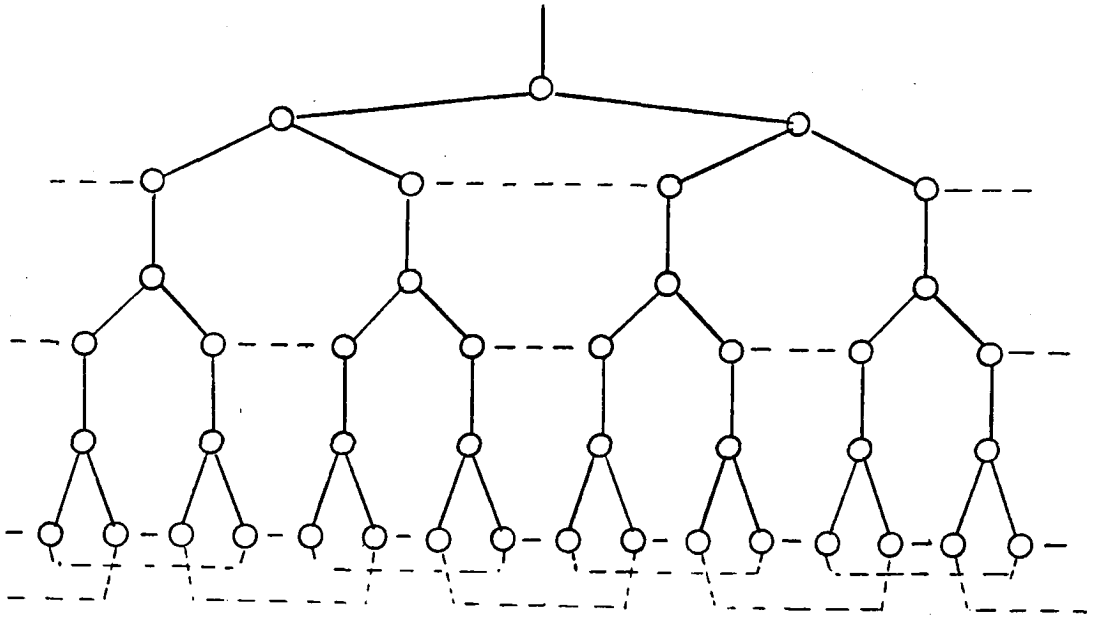


Figure 3.3.4-5. Variant of Semiregular Tree Network.

problems, losing much of the original advantage.

Semiregular tree structures are actually of minimal interest for data communication in active control transports. Not only do they not gracefully fit the airplane, but the attribute of short nodal distances is of secondary importance for the sizes and types of networks that would be used.

The most profitable approach for aircraft applications is a semiregular structure composed of interconnected regular structures. Thus the fuselage systems might be connected as one toroid, the wings another, and the tail another. Boxes in bays might form separate toroidal groups as well. To connect one toroid to another, certain mutually distant nodes of one toroid can be replaced by sets of external ports, as in Figure 3.3.4-6. Four such nodes create twelve ports in this case, where three-port nodes are used. Figure 3.3.4-7 indicates how toroids might be interconnected, joining the analogous port clusters in each of four toroids. Obviously, this is only one example, but it serves to show how regularity can be preserved within the individual toroids that make up the larger system.

3.4 Subscriber Assignments

Subscriber assignments to network nodes would preferably be a matter of convenience, presumably to minimize cabling. Two considerations, however, may interfere with assignments that are convenient. One consideration is reliability. The other is the possibility of exploiting parallel channels for purposes of masking faults.

Some redundant sensors and effectors are dispersed about the system (or can be if desired), whereas others are naturally located in the same vicinity. Examples of the latter are skewed inertial strapdown instruments and triplex force-voting actuators. When assigning such localized redundant elements to network nodes, each simplex entity would have its own node, forming a redundant group. Convenience would call for choosing these nodes from a continuous fragment of the network, as in Figure 3.4-1 (b). But if the airplane is dispatched with the B node failed, then both the A and C nodes are threatened by possible failure of their neighbor nodes. If the MTBF of the nodes is not very great, this may present probabilities of losing two or three channels that are too high for certification of the airplane. If this is the case, then draping a regular network onto an airplane system as in Figure 3.4-1 (a) could turn out to be a messy business, with a great

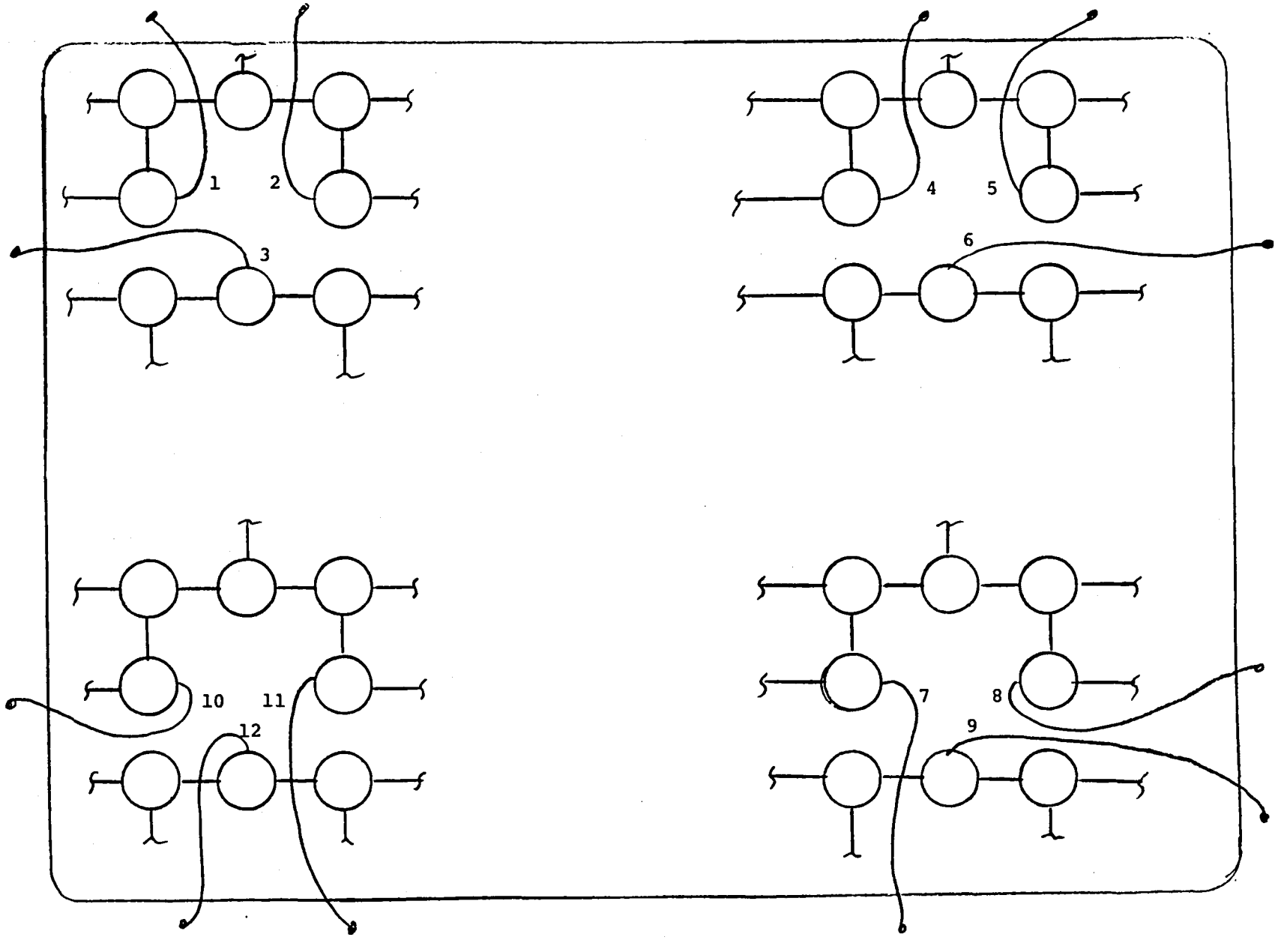


Figure 3.3.4-6. Connections to a Toroidal Network.

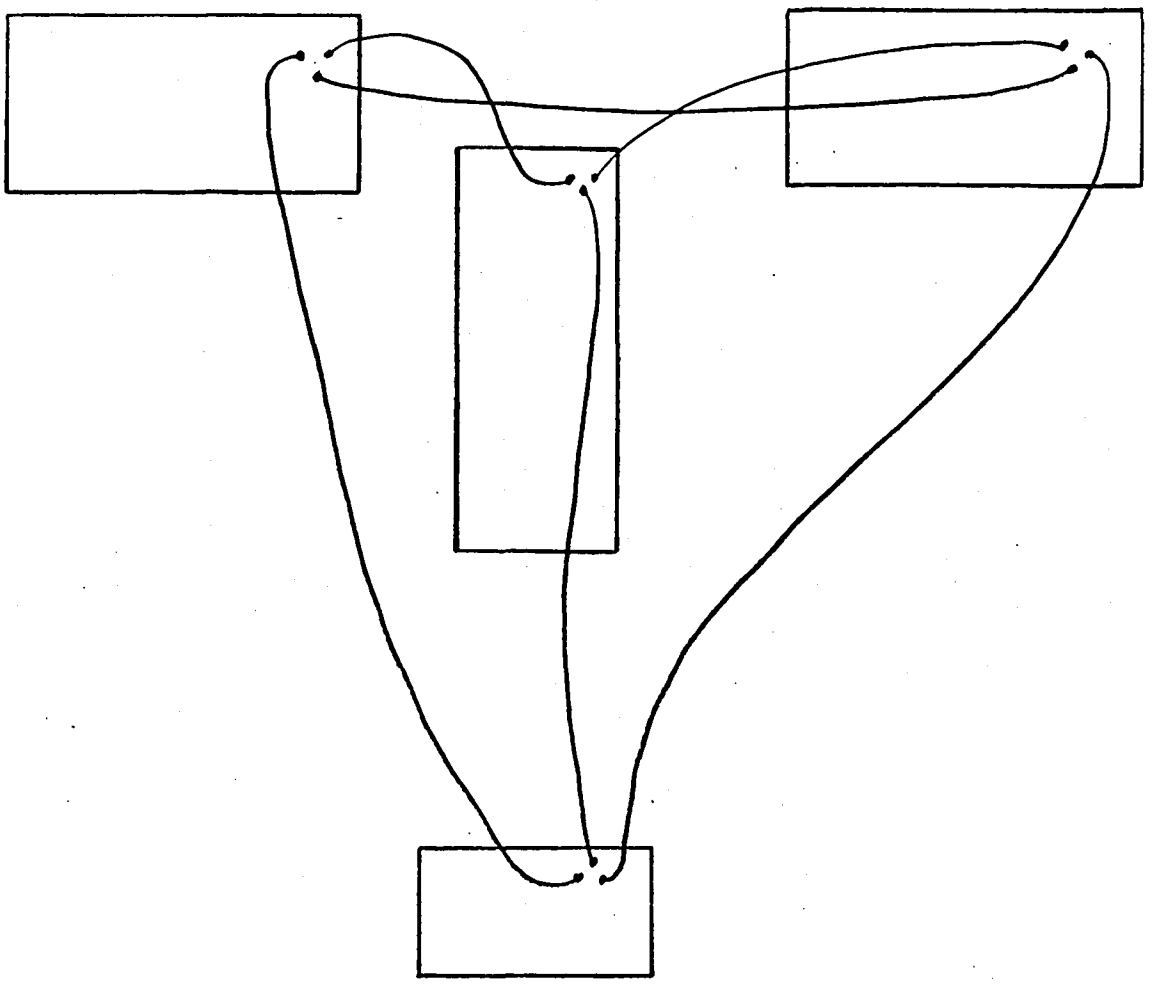


Figure 3.3.4-7. Inter-Toroid Connections.

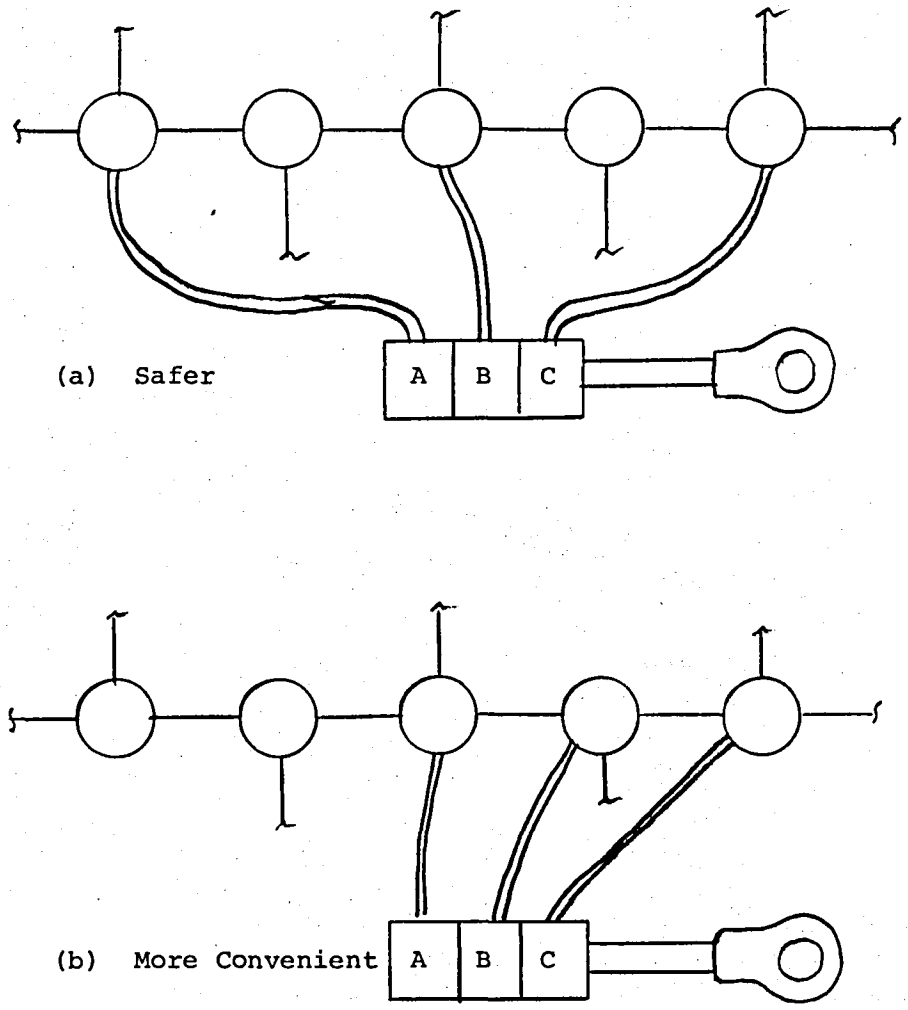


Figure 3.4-1. Force-Voting Actuator Node Assignments.

deal of extra linkage required. From this viewpoint as well as the maintenance viewpoint, it would be preferable, if not necessary, that the MTBF's of the nodes be of the order of 50,000 hours or higher for passive faults, and 300,000 hours or higher for active (e.g. hard-over) faults.

For multi-channel operation, the network needs to be draped so as to make it easy to grow multiple subtrees to handle separate channels. Figure 3.4-2 illustrates one possible toroidal net draped on a group of triple actuator nodes, as might be required in an advanced aeroelastic wing.

A final note on subscriber assignment: The stated assumption underlying much of the discussion thus far is that everything is critical, whereas in fact some subscribers will be non-critical, but rather present for economic benefit. The non-critical elements can be assigned for convenience, without any particular regard for the constraints discussed in this section.

3.5 Issues Concerning Node Architecture

In this section a number of issues are reviewed having to do with the design of network nodes. The nature of nodes has so far been described in abstract terms. Nodes contain repeaters, enabling switches, gateman circuits, and controllers. This much is clear from a description of the grow algorithm. Numerous variations are possible as to how these elements are mechanized, and how they interact.

In addition to the latitude available to the designer in realizing the basic functions of the node, the designer has other choices to make that affect the architecture of the node. Four such choices are addressed in this section.

- Interface Standardization
- Fiber Optics vs. Electrical Conductors
- Embedded vs. External Location
- Identification Options

3.5.1 Interface Standardization

Mesh networks have almost the same problem of interface standardization as multiplex buses do. This is to be expected since the network emulates a multiplex bus. The multiplex bus has interfaces at controller ports, remote couplers, and between remote couplers and

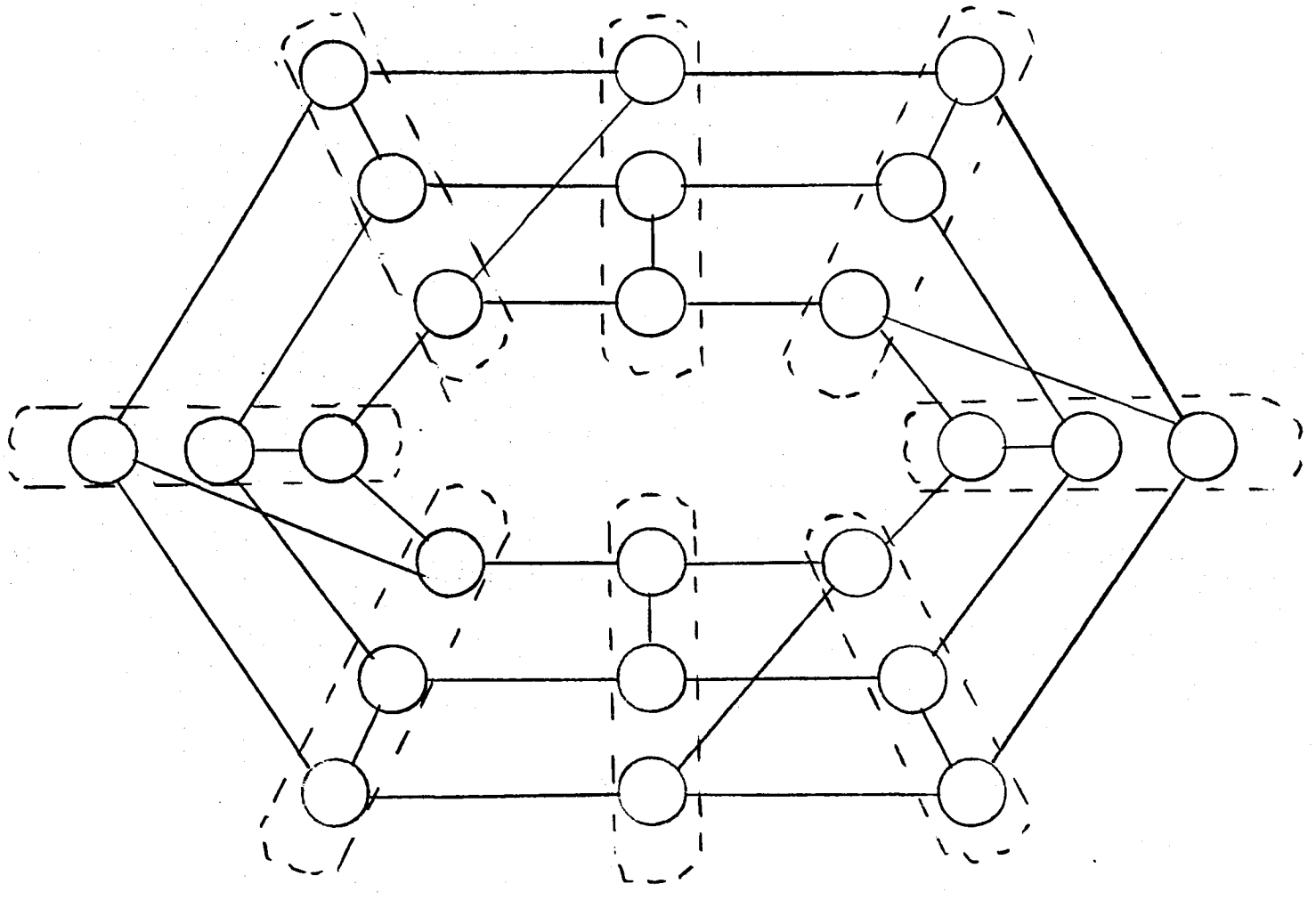


Figure 3.4-2. Toroidal Connection of Triplex Subscribers.

subscriber ports. Mesh networks, meanwhile, have interfaces at controller ports, node ports, and between nodes and subscriber ports. Standardization is equally important and equally difficult in the two methodologies.

The preferred approach would be for all nodes to be identical, although provision must be made for competitive manufacture. Competitive design to form, fit, and function specifications is another possibility. The interface to the subscriber port can be, and presumably should be, identical to that of a standard multiplex bus, such as 1553. This interface has successfully been implemented to form, fit, and function by numerous manufacturers. The node port and controller port interfaces would be essentially identical, and would be similar to multiplex bus interfaces, and also amenable to form-fit-function specification. The node control logic is the most complex part of a node specification, and it remains to be determined whether this kind of specification is practical for nodes.

A strong incentive exists for making the network compatible with 1553. The electrical interfaces of 1553 all use a serial Manchester biphase signal on twisted pair. A growing number of subscriber devices are being developed that are interfaced this way. The node-to-subscriber interface can be made this way as well. A problem arises, however, if the Manchester signals are passed through a series of repeaters like those in a mesh network. Repeaters contain amplifiers, whose rise and fall characteristics are not guaranteed to be identical. If several in a row have the same bias, pulses will grow or shrink, thus distorting the wave-form, quite possibly to the point of incoherence. Figure 3.5.1-1 shows the effect of successive repeaters in which rise time is slower than fall time.

One alternative to Manchester coding is to transmit short pulses demarcating leading edge positions of the equivalent Manchester code, as shown in Figure 3.5.1-2. It is possible to reconstruct the Manchester code from the pulse code using an accurate time counter circuit, based on the property of Manchester code that it always changes state at the data strobe instant half way through the bit period. When the leading edge pulses are repeated, their durations are ignored at each repetition. As shown in Figure 3.5.1-3, their separation is preserved despite biased repeater amplifiers.

The pulse code is the recommended standard for all of the node-port to node-port interfaces in a 1553-like mesh network. This means

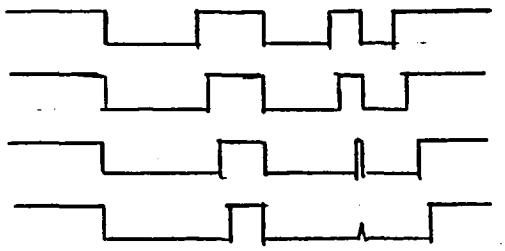


Figure 3.5.1-1. Shrinking Pulses

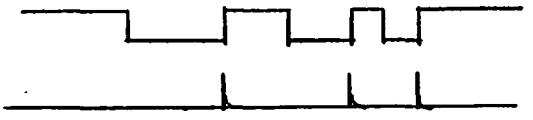


Figure 3.5.1-2. Leading Edge Pulse Translation.

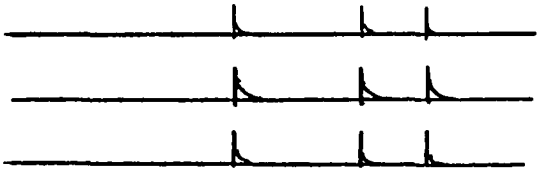


Figure 3.5.1-3. Leading Edge Pulses Repeated.

that the controller interface would either be built for pulse code, or else would have an external translator interposed, possibly at the first node. Each node would contain a two-way translator to couple the subscriber to the network.

This can be accomplished, albeit expensively, by using repeater circuits which retime pulse durations by means of high-speed clocks. The minimum delay through each repeater is of the same order of magnitude as the rise and fall time uncertainty in the transmitter-receiver pairs. Delays of this order should be tolerable in the kinds of networks described here.

3.5.2 Fiber Optics vs. Electrical Conductors

The presence of repeaters in the nodes of a mesh network yields options for linkage technology not easily available to multiplex bus technology. Fiber optics is a prime example of such an option. The fiber optics versions of 1553 multiplex buses reported so far have strayed quite far from the simplicity and robustness of the electrical version. The main reason is that the optical energy must necessarily be divided equally among the subscribers, which eats heavily into design tolerances when the number of terminals approaches twenty or so.

In mesh networks, each link is actually two half-links in a full-duplex arrangement, i.e. one half-link in each direction. Each half-link can be implemented as a fiber optic channel, with one transmitter, one cable, and one receiver. The cost is having numbers of transmitters and receivers in each node equal to the number of ports per node. The benefit is the absence of dynamic range problems and power division problems, which allows design tolerances to be healthy and broad.

It must always be pointed out, in discussing the potential for fiber optics, that it is not yet practical to achieve complete electrical isolation using fiber optics, since electrical power must still be distributed on hard wires to every node in the system. Experimental circuits have used optical power, but this is not anticipated to impact avionics for transport airplanes in the foreseeable future.

3.5.3 Embedded vs. External Location of Nodes

The nodes in a mesh network perform dual roles, combining the functions performed by remote couplers and remote interfaces in 1553 multiplex systems. They couple together the various segments of link-

age to form a coherent network, and they couple complex subscriber circuits to a simple twisted pair interface. In 1553, the remote couplers are fixed to the airframe while the remote interfaces are embedded in the subscribers. In a mesh network, a choice needs to be made between embedding the node in the subscriber, attaching the node to the airframe, locating the node in a separate line-replaceable unit (LRU) from the subscriber, or fragmenting the node into two or more of the above-mentioned locations.

If the node is embedded in subscriber packages, then the network is injured whenever the subscriber package is removed. This may or may not cause a problem, depending largely on whether airplanes would ever be dispatched with missing boxes. Line maintenance and periodic maintenance procedures could also be affected, especially when several boxes are removed while the system is under test.

Another consequence of embedding nodes is the virtual necessity of using form-fit-function specifications so that any vendor of a subscriber element can incorporate a node of his own design.

Embedding a node in a subscriber increases the subscriber's effective failure rate. If the subscriber's original failure rate is much greater than the node's, then this is of no consequence as far as flight safety or maintenance are concerned. If the node's failure rate is greater than the original subscriber's, then the impact on flight safety and maintenance must be calculated. It has already been indicated that node MTBF's of several tens of thousands of hours are apt to be necessary for active-control aircraft applications. Relatively few subscribers are likely to be more reliable than this. Mechanical servos today have reliabilities of this order, but they are not amenable to multiplexing without embedded electronics added, which could substantially reduce their MTBF's. Servo manufacturers, however, are constantly exploring the possibility of incorporating electronic feedback control into hydraulic servos. It would seem that very high MTBF's would be needed in such devices in order for airlines to find them acceptable, even if the electronics modules were easily removed and replaced from the servos, because of the difficulty of accessing servos during line maintenance, which occurs outdoors in ambient conditions. At any rate, if the servo designers succeed, an embedded node must not compromise the servo electronics. This issue, incidentally, exists equally for 1553 interfaces.

If all or part of a node stays in the airframe when the subscriber is removed, the network injury problem is solved. This could be of value in certain military aircraft, where different avionics complements are dispatched for different missions. Another advantage could accrue if fiber optics links are used, since the optical fiber connectors would not be mated and unmated as often as the subscriber's electrical connectors. They could therefore afford to be made more robust than if they were incorporated in a box's end connector. Again, the assumption is made that nodes are extremely reliable.

Another option, possibly suitable for use in avionics bays is to place several nodes in a single LRU package located among the associated subscriber boxes. This approach could be effective for 1553 bus subscribers in a mesh network. The nodes would appear to the subscribers as if they were the remote couplers of a 1553 multiplex bus.

3.5.4 Identification Options

One of the intrinsic necessities of multiplex systems is to have a means whereby the remote terminals, the nodes, and the subscribers are able to determine which messages are and are not directed at them, and which times they are and are not eligible to transmit on the channel. The method used for this purpose almost always involves the assignment of an identity code to each transmitter, so that it can either hear itself called or can count slots until its turn to talk. Only in the case of ring networks is identity unnecessary, since in that case a specific enable signal arrives at the transmitter whose turn it is to talk.

An identity code must be known in at least two places, i.e. the controller and the transmitter, and perhaps secondary controllers. Codes must moreover be disjoint so that each code designates a unique transmitter. An otherwise healty transmitter can become a babbler if its identity code is incorrect.

The method by which identities are assigned impacts performance, economy, and reliability of the network. If the identification is hardwired into the transmitter, then the transmitter is either committed to a single function for its lifetime, or else the controller must be notified as to which of several possible functions a given transmitter code represents during this flight. If the code is to be assigned by the controller writing into an identity register, some means is needed, like the one that exists in the ring network, to tell a specific

transmitter than an identity code assignment message is intended for it. Another possibility is to wire the panel connectors so that the proper codes are presented to whatever transmitter is plugged in. Still another is to have coded buttons or plugs available to insert into transmitter packages.

Whatever method is used, it would be reasonable to diminish the probability that one mistaken identity would map into another. Some form of redundancy, such as replicated identification code words, or error detecting or correcting codes should be considered.

3.5.5 Environmental Considerations

When a network or any other digital transmission system is used to interconnect fully dispersed systems of sensors and effectors, some of the terminals, nodes, and/or subscribers will be located in places where it is difficult to provide environmental control. Places like engines, wings, and tail surfaces experience extremes of heat, cold, and vibration, unlike the fuselage avionics bays, where a relatively benign environment is found.

The impact of a harsh environment on electronics is a reduced MTBF, which has potential impacts on safety and an absolute impact on maintainability. Thus the nodes and subscribers that are most apt to need replacement are those in the more remote spots that are awkward to reach. It is therefore essential that node design take into account the environmental extremes of airplane locations if networks or other multiplex systems are ever fully to displace dedicated passive linkage.

3.6 Network Design Summary

Rather little experience exists as yet with respect to mesh network design for aircraft. Three generations of experimental networks have been designed at the Draper Laboratory, none of which, however, has exceeded ten nodes in size. The following, therefore, represents a prediction as opposed to distillation of experience in the methodology of network design.

3.6.1 Subscriber and Node Locations

Perhaps the first step in designing a mesh network is to decide where subscribers and nodes will be located, subject, of course, to iterations of the design. Environmental issues surface immediately.

3.6.2 Embedment of Nodes

The location of nodes with respect to subscribers and the possible fragmentation of nodes is a major decision. It will be based on the projected scenarios for operation and maintenance for the airplane, as well as on possible utilization of 1553 or ARINC 429 sensors and effectors with multiplex interfaces already installed.

3.6.3 Link Technology

The option to use fiber optics should be decided upon before node technology is selected. This is an enormous issue, which can really be resolved only by future experimentation and experience. Many problems remain to be solved, including susceptibilities to temperature, vibration, x-rays, and repeated mating and unmating of connectors. If and when all the problems are solved, this will be an attractive medium because of its electrical isolation and high bandwidth capabilities.

3.6.4 Node Design

Node design will eventually reduce to a choice among existing designs. At first, however, a substantial challenge exists to create a small, reliable, inexpensive, and capable device.

3.6.5 Multiple Paths

When the channel bandwidth has been determined, any need for multiple channels will be evaluated.

3.6.6 Network Topology

The number of node ports may have been predetermined by available node designs, or perhaps this number may be independently determined at this stage. At any rate, a regular or semi-regular network geometry is to be chosen based on the shape of the airplane, reliability requirements, and multiple path requirements.

3.6.7 Node Assignments

Subscribers will be assigned according to considerations of separate trees, clustering of mutually redundant subscribers, and perhaps distance from the controller.

3.6.8 Operation Principles and Protocol

Protocol and operation principles may have been predetermined. If not, they may be chosen relatively late in the cycle.

3.6.9 Performance and Reliability Assessment

Models are needed to evaluate the compatibility of the design decisions described above. Failure probability, bandwidth, recovery speed, dispatch probability, and maintenance frequency are the principal assessments to be made.

CHAPTER 4

POWER DISTRIBUTION

In this chapter the role of power transmission in flight-crucial active control systems is discussed. To begin with, power in airplanes is presently generated, distributed, and applied in both electrical and hydraulic forms, which is justified on the basis of their respective efficiencies for control processing and actuation. The distinction is not absolute, since, for example, some actuation is electrical, and some control processing is fluidic.

There is a possibility of achieving efficiencies in electrical actuation in the coming years such that it may become appropriate to eliminate hydraulic systems altogether. Much as this would be welcome from an esthetic viewpoint, it should be recalled that despite its many nuisances, hydraulic engineering has successfully come to grips with a full-time flight-crucial availability requirement in some of the more recent airplanes. This is not to say that active control technology was achieved, but the application of redundant power transmission elements giving continuity of service despite faults and damage constitutes an important step. In going from hydraulic to electric actuation, this achievement would have to be matched in electrical power transmission.

Unlike the case of signal transmission, multiplexing is not a significant issue in power transmission. This applies, of course, to the actual power elements themselves. Multiplexing of power control signals is an important issue. To some degree, there is a weak analogy to multiplex signal transmission whenever power is bussed to several destinations. The important difference is that the loads and the sources can be passive in the power bus case, although for practical reasons some form of power interruptor is required to protect the bus from short circuits or leaks, as the case may be.

There are essentially three basic topological forms to consider for power transmission. These are dedicated feeders, buses, and mesh networks. In the first case, power is allocated ot each "subscriber"

at a central location (e.g. a breaker panel) and routed over dedicated channels to the point of application. The bus form of distribution reduces the number of dedicated lines, while requiring that remote protection be employed. A hybrid approach would route power to several remote distribution terminals ("substations") from which dedicated lines would run to nearby subscribers. The mesh network allows power to be shared over a multiplicity of simultaneous paths. Each node can both receive and transmit power. Protection is afforded by limiting the amount of power a node can transmit, or, alternatively, configuring each node so as to be able to switch off its incident power links on command.

4.1 Hydraulic Power Distribution

The problems of hydraulic systems primarily stem from slow leaks, which are difficult to detect if not seen. At least three problems result from leaks. One is loss of fluid, which is potentially critical. A second is the dissolution of paint and insulation, and the third is the fire hazard of a mist of hydraulic fluid, which can be flammable where the liquid form is not. It would obviously be desirable to detect slow leaks and suppress them with valves of some sort. As yet, however, no such detector has been available, at least at an affordable cost. This seriously limits the degree to which hydraulic transmission systems can be improved over their present state.

Present systems consist of three or four separate hydraulic circuits, which draw from separate fluid reservoirs in order to avoid vulnerability from a single leak. In principle, both dedicated and bus distribution forms are possible. A mesh network, however, mixes power from several sources and does not differentiate fluid reservoirs, which makes its application unlikely.

When a power component fails (which is not a particularly rare event), it would be undesirable from several points of view to lose a third or a fourth of the flight control system. This will be, if anything, more true of active control systems. Cross-strapping of power systems is therefore desirable. For this reason, each separate hydraulic channel is likely to be powered by more than one source. Pumps may be geared directly to turbines (engines and APUs), may be electrically driven, or may even be hydraulically driven. Some hydraulic-driven pumps of this sort are symmetrical, so that either channel can power the other. Care must be taken in cross-strapping to avoid situations where the mechanism that supports power sharing has a

failure mode which adversely impacts all of the channels it serves at one time.

4.2 Electric Power Distribution

Unlike hydraulics, electric power distribution has not yet had to confront the problem of full-time availability as a flight-crucial requirement. Autoland requires continuous availability for a brief period of time, and there are other times when loss of electric power would be awkward, but none of these are full-time situtations. Today's airplanes are configured for a five-minute flight period following power loss.

In addition to being intermittent, today's electric power is "dirty", with substantial voltage excursions for brief periods and smaller average excursions. Open circuits are common, and short circuits, although relatively rare, do occur. The nominal reaction to a short circuited power link is that a circuit breaker will open, usually the one that is intended to open, and more rarely one of the breakers hierarchically superior to it. The voltages throughout the system momentarily diminish or vanish, and in some cases require manual intervention to be restored.

A flight-crucial active control system depends on electric power that is effectively free from any interruption. Contemporary practice is to specify every independent "subscriber" to incorporate power conditioning equipment designed to co-exist with a standard power quality (MIL-STD-704). This is effective to some degree, but not wholly so. It is also expensive, and it results in the extensive generation of heat, which tends to increase component failure rates.

Two fundamental problems exist. One is to maintain and distribute a raw supply of power, and the other is to defend it against faults and damage. The first problem is largely solved by having redundant generators, APUs, ram-air turbines, and emergency batteries. It would be desirable to have more energy storage than there is at present, but batteries are hazardous and present a maintenance nuisance. For the foreseeable future, electric power will be derived much as it is now, with the exceptions that mechanical constant-speed drives will probably disappear, and that electronic switching will come into use both for circuit protection and DC-AC conversions in both directions, accepting variable-frequency power from alternators and furnishing either DC or fixed frequency AC at fixed amplitudes to subscriber loads.

The second problem, i.e. defense against faults and damage, requires high-speed reactions to malfunctions. This is not so much a problem for open-circuit faults, because power can be taken from independent sources and merged through passive devices. Short circuits, however, present problems similar to the problems presented by hard-over actuators, because the effects of these faults can propagate throughout the system unless the system possesses the ability to neutralize them. It is moreover not feasible to vote electric power analogous to the way that actuators can be force-voted. Consequently, some means is required to disconnect power from a subscriber. This means must not be a part of the subscriber itself, in case a damage event should simultaneously short circuit the subscriber's power and incapacitate the disconnection element. Moreover, the entire power system, from its source to its loads, needs protection, and transmission links are vulnerable as well as the subscribers.

The design problem for power disconnect capability is further compounded by the fact that the system is as much imperiled by erroneous disconnections as by short circuits. This threat ranges from a malfunctioning power control subsystem computer, to a number of circuit breakers accidently opened by a momentary surge caused by a lightning strike, to a damaged breaker panel, to a faulty resource management algorithm. Circuit breakers do not appear to solve the problem effectively. Individual batteries are not practicable. Relays located in regional distribution subsystems may be adequate, as long as each subscriber receives power from at least two such distribution subsystems, and some means exists for protecting the power system from damage to the subsystems. In some sense the problem moves from the subscribers to the distribution subsystems.

4.3 *The Substation Approach*

A number of contemporary developmental systems have experimented in the use of computer control for electrical power distribution [6,7]. These systems have each used a number of remote subsystems that act as electrical substations, receiving power from redundant main buses, and switching subscriber loads by digital commands from a central controller. Critical subscriber loads can be powered from more than one substation, so that if one substation should fail passive, then each affected load can be supplied from another substation. With the control switches located in the substations, the system is protected against short circuits that take place in subscriber circuits, or in

the power leads between subscribers and substations. If control switches were to be located solely within subscriber circuits, the power leads would not be protected. However, if the substation were to be damaged so as to short circuit numerous power leads, the local switches could protect subscribers against a consequent loss of power. Similarly, if main buses are shorted at a substation, system survival would depend on protective switches in the vicinity of the primary power sources.

Thus the substation approach needs to be augmented by remote control switches local to subscribers and power sources if it is to provide coverage for all conceivable failure modes. The management of such a system by computer control would be a moderately complex task, comparable to that of managing a mesh data network.

The reliability of a substation-type system can be made quite high, despite the fact that it is quite a bit more complex than a conventional breaker system. The redundancy provided by multiple power sources for critical subscribers is the principal reason. The failure of individual remote switches can be moderately probable without seriously affecting the system success probability. For example, in a system of 100 subscribers, each of which receives power from two substations, the power switches need only have MTBF's on the order of 10,000 hours for the system to have a failure rate below 10^{-9} per hour.

4.4 Current Limiting

If most of the switches in a substation-type system were to be replaced by current limiting devices, a great deal of complexity could be saved with respect to the software needed to manage the system. It would be necessary to incorporate only those switches needed to provide normal power on and off to each subscriber. The attractive feature of current limiters is their autonomous operation. They are, however, unusual, complex, and/or expensive devices.

If a subscriber receives power from two or more sources through current limiters, then if the subscriber becomes short circuited, it will draw a limited amount of current, which will not overload the remainder of the power distribution system. If one of the limiters should be faulty and not limit properly, then one of the main buses could become unusable. This kind of fault is latent in a system, and a test would have to be devised to discover it. Otherwise, a full set of current limiters might all possess the same kind of fault, in which

case all buses might be brought down.

Not surprisingly, current limiting is much easier to accomplish for low currents than for high currents. A good example is an electronic regulator circuit where the control variable is current rather than voltage. A switching regulator can conveniently be made to operate this way. It can have a so-called foldback characteristic, if desired, where the current supply is cut off if the demand exceeds a set level, and stays off until the demand is reset to zero. Alternatively, it can be designed to saturate at a set level. The efficiency of switching regulators is likely to be on the order of eighty per cent. This should be quite acceptable for modest size loads. Considering the fact that these regulators can condition the power's quality well over that of MIL-STD-704, it should be possible to save more than twenty per cent by designing the subscriber power supplies to a more benign input specification.

A second type of current limiter is based on A.C. saturable reactor principles, and is shown schematically in Figure 4.4-1. The series impedance windings at the left of the diagram are connected in series with the load. They normally contribute a negligible impedance to the circuit, because normally legs 2 and 3 are magnetically saturated by flux from the D.C. bias winding. The shorted turn acts as a low-pass filter on the flux in leg 1 to maintain this bias at a constant level. The load current produces alternating fields in legs 2 and 3. An excessive current will produce fields large enough to alternately cancel the bias fields in legs 2 and 3. When this happens, the respective legs become unsaturated, and the impedance of the windings becomes high, placing an upper limit on the load current, assuming constant source voltage. Figure 4.4-2 shows a typical voltage-current characteristic for a series connection of a resistive load and a current limiter with variable source voltage, V. I is the load current.

The saturable reactor current limiter can be large and heavy. It is made from transformer material, i.e. iron alloys and copper wire. For a given power level, its size and weight are comparable to the size and weight of a transformer for the same power level. As a rule of thumb, 400 Hz transformers will weigh the order of one Kilogram, and will occupy a volume the order of 15 cm^3 (10 in^3) for every 100 watts of power transmitted. Higher frequencies would reduce these penalties.

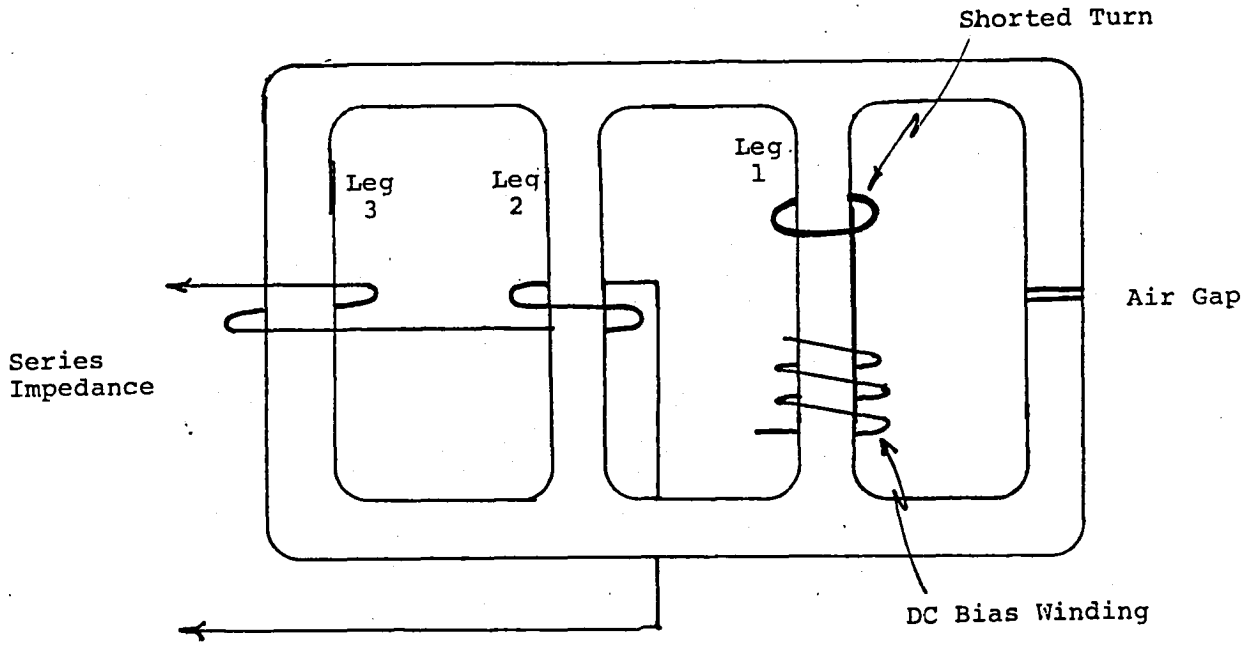


Figure 4.4-1. Saturable-Reactor Current Limiter.

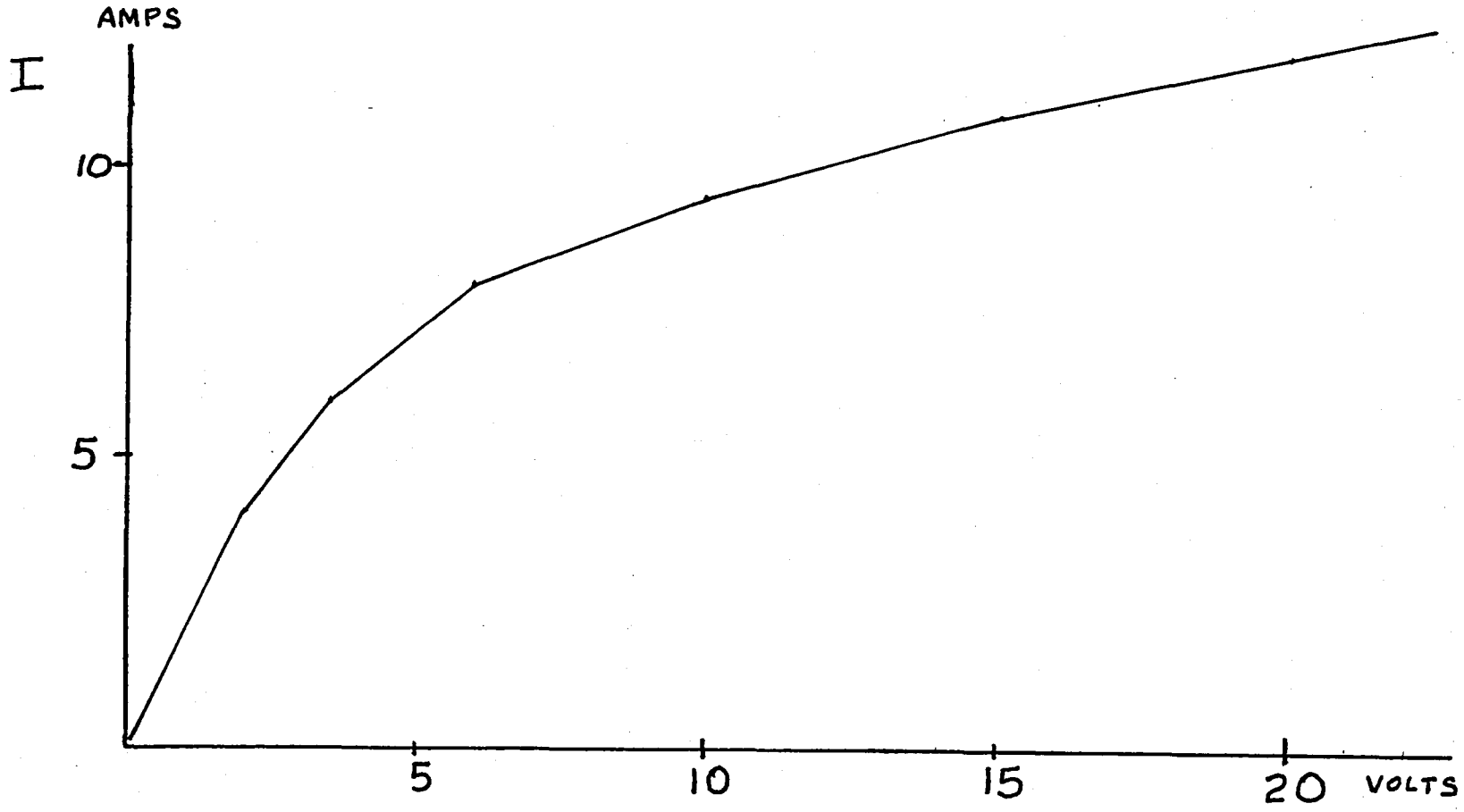


Figure 4.4-2. Typical Saturation Characteristic for a S-R Current Limiter.

4.5 Mesh Networks for Power

Networks are used on a continental scale for distributing electrical power. It is natural, therefore, to seek a power analogy to the mesh networks for data discussed in Chapter 3. Such an analogy does exist, although there are some important differences between data and power networking. The principal transmission concern in power networks is quantity, whereas in data networks it is quality. Thus data is distributed over trees of connected links in order to avoid problems that would result from multiplex arrivals with different delays. Power transmission does not suffer from multiple arrivals, but rather is helped if every available link can share the burden.

Prime power injected at numerous dispersed points in a regular network is allowed to flow through every link from node to node to meet load demands. A simplified example is shown in Figure 4.5-1. Assume that prime power is injected at equal potential at each of the six shaded nodes, and is consumed at an equal rate at every node. The arrows indicate direction of current flow. Current flows out of the shaded nodes on all three links. In twelve nodes current flows in on all three links, and in the other eighteen it flows in on one link and out on two.

Thus far, nothing has been said as to how the links are joined inside a node. If they are simply connected together, then a single short circuit will short the entire system. An open circuit, however, would be tolerated by a redistribution of currents. To tolerate short circuits and other overloads, the nodes are made capable of interrupting or limiting the current on each link separately. In this way, each node defends itself against an overly greedy neighbor.

If switches were used to interrupt link current, they would be controlled from within the node, based on sensed current flow, as shown in Figure 4.5-2. The controller's job is to prevent excessive outbound current. Current limiters can be used instead of switches, as shown in Figure 4.5-3. No controller is necessary in this case, although a switch in the subscriber's power port will be necessary. The figure assumes bidirectional limiters, such as the saturable reactor type, which requires alternating current. In this case, the arrows in Figure 4.5-1 show power transfer rather than current.

Figure 4.5-4 shows one-way D.C. current limiters connected to limit outbound current, but not inbound. This would apply to switching regulator limiters. Unlike data networks, which use double links,

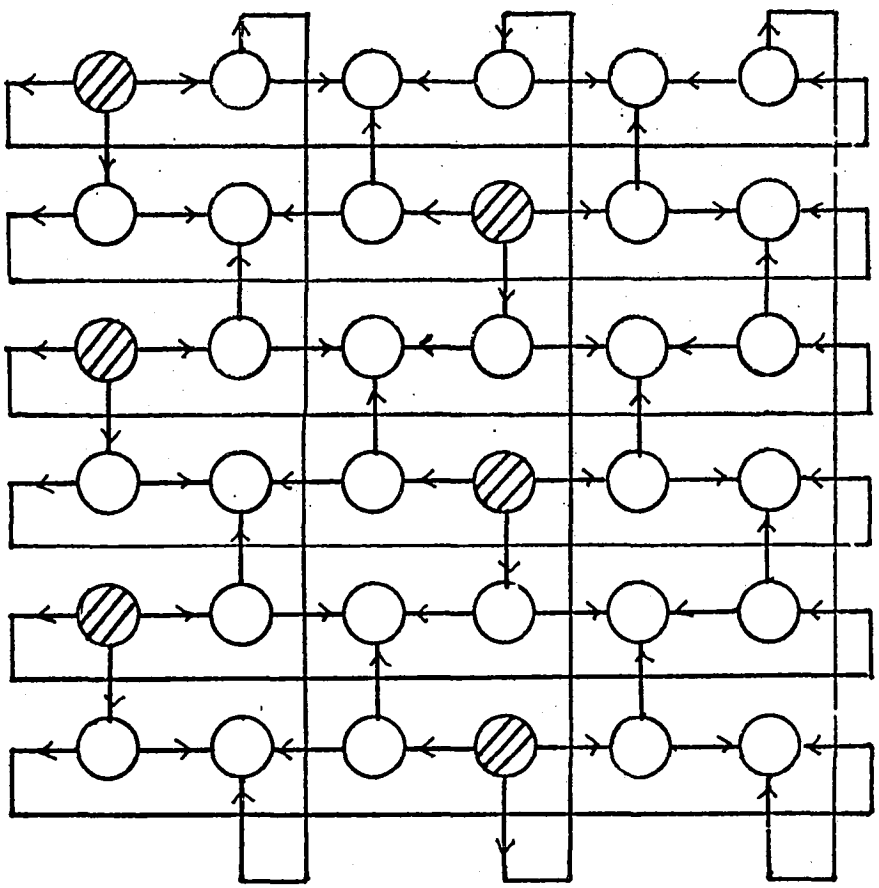


Figure 4.5-1. Current Distribution in a Power Network.

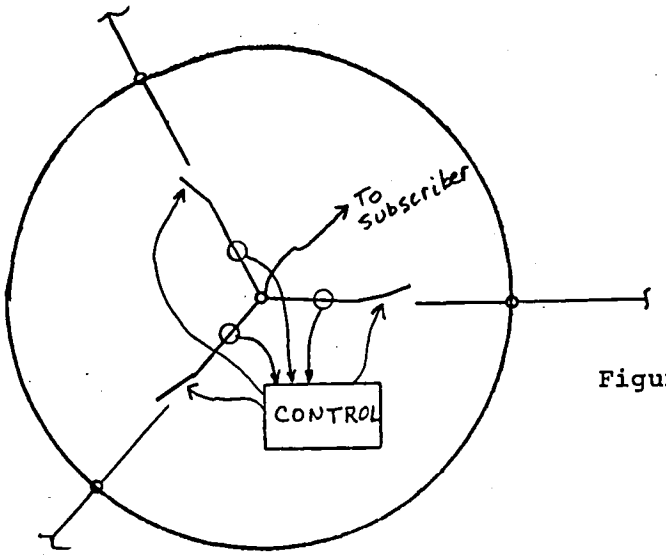


Figure 4.5-2. Switched Links.

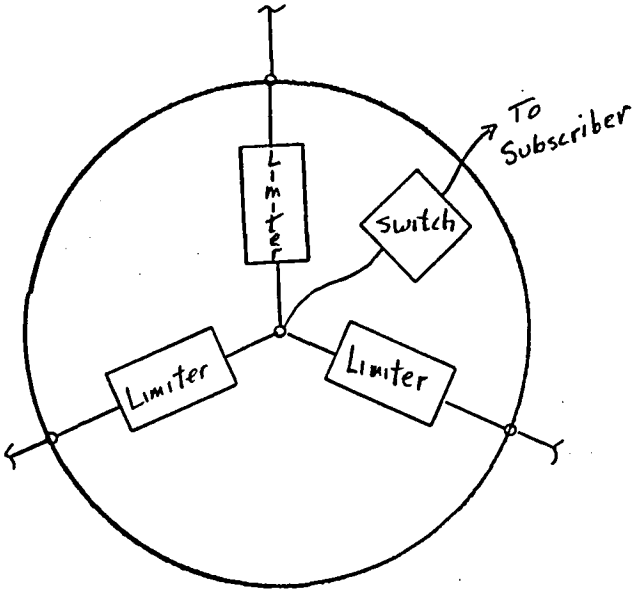


Figure 4.5-3. Current-Limited Links.

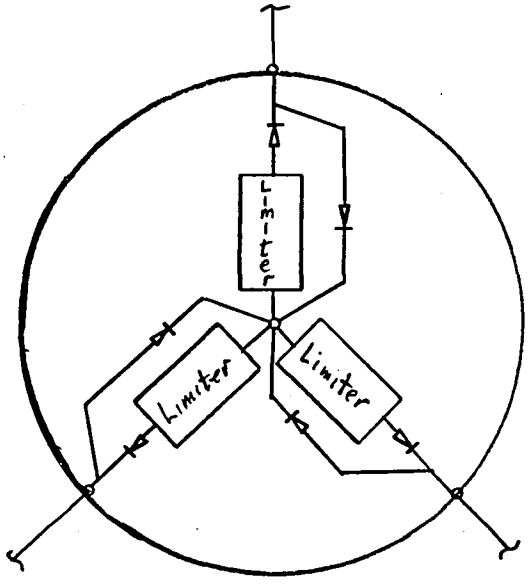


Figure 4.5-4. One-Way Current Limiter Links.

power networks would use single links.

The performance of current-limited networks in normal operation as well as in case of short circuits is imperfect. There is voltage loss from node to node, and power loss in the current limiters with possible voltage drops from node to node. In normal operation, the power loss is small in magnetic limiters, being only copper losses, while it is substantial in electronic limiters due to junction drops. In case of a short circuit, however, the electronic limiter is better at containing additional power loss than the magnetic limiter.

As a numerical example, consider the 36-node power network shown in Figure 4.5-1 using electronic limiters that are 80 per cent efficient. Neighbor nodes of the power nodes have a 25 per cent overhead on power they use, and their neighbors have a 56.25 per cent overhead. Weighting these overheads by the numbers of nodes of each class yields an average overhead of 31 per cent. Next, one can calculate the power overhead using magnetic limiters the characteristics of which are shown in Figure 4.4-2. The results are sensitive to loading parameters, so several cases were calculated as shown in Table 4.5-1.

To elaborate on the same example, calculations were made for faulty network cases. Table 4.5-2 shows overheads for a case where prime power is withdrawn from one of the shaded nodes, and for four different short circuit cases. The results from Table 4.5-1 are repeated for comparison. The extreme power overheads for short circuits are greatest where the normal overheads are least.

In the electronic limiter case, the overhead current for a shorted node is three times the limit current increased by the normal power overhead at the nodes supplying them. The limit current can be set somewhat in excess of normal currents. In a typical case from the preceding magnetic limiter exercise (case 1), the nominal power delivered from each power node is 50 watts, since it effectively powers five other nodes. Each of three links from a power node carries 50/3 watts at 30 volts, or 5/9 ampere. Suppose the limit is set at 2 amperes. Then an extra six amperes, approximately, must be supplied if a node shorts. If the nodes supplying the extra current have power overhead percentages of 56.25, then the total overhead power is 6 amperes times 1.5625, or 9.375 amperes, times 30 volts, equaling 281 watts. The total nominal power is 350 watts, so the overhead percentage in this case is 80 per cent. This is about one forth of the overhead for the magnetic limiter, and could be made lower still by reducing

TABLE 4.5-1

POWER OVERHEADS WITH MAGNETIC LIMITER

CASE	PRIME POWER VOLTAGE, VOLTS	POWER DELIVERED TO EACH NODE, WATTS	OVERHEAD
1	30	10	0.85%
2	20	10	1.95%
3	10	10	9.28%
4	30	5	0.42%
5	20	5	0.96%
6	10	5	4.12%

TABLE 4.5-2

PERCENT OVERHEADS FOR VARIOUS MAGNETIC LIMITER CASES

CASE	NORMAL OVERHEAD	OPEN POWER NODE	SHORT POWER NODE	SHORT NEIGHBOR NODE	SHORT SECOND NEIGHBOR	TWO SHORT NODES, DISTANCE TWO APART
1	0.85	1.24	314	330	325	540
2	1.95	2.90	184	192	190	-
3	9.28	15.71	-	85	84	-
4	0.42	0.61	624	658	647	1,077
5	0.96	1.40	358	379	374	620
6	4.12	6.32	140	152	148	243

the value of the current limit.

In networks using magnetic limiters, the power loss results in voltage drops from one node to the next. In the preceding examples, the nominal voltage drops are on the order of a volt or less. When one node is shorted, some other nodes drop as low as half of the prime voltage. With two nodes shorted, one node goes below one tenth of the supply voltage. If conventional subscribers are to be used, the nodes should contain voltage regulators.

Electronic limiters can be designed so as to incorporate voltage regulation up to their current limit, thus avoiding large voltage differences from node to node.

4.6 Power Distribution Summary

Much of this chapter has been devoted to discussions of relatively new and untried power distribution techniques. Such techniques are sought because of the intrinsic vulnerability of contemporary techniques to momentary or substantial interruption. Aircraft designers have a large job on their hands to design an electrical system for an autoland-equipped airplane. The outlook for active control is that the job will be a great deal more difficult.

Substation approaches will be helpful in this situation. A great deal of initial research and development has been applied to such systems, and the results should be available for incorporation into next-generation airplanes.

Current limiting technology is barely in its infancy, and even when developed will probably not be universally applicable. The advent of electrical actuators will mean that some very heavy loads will have to be handled, which may be beyond the scope of current limiting devices. They are moreover apt to cause severe line transients.

Far-future aircraft systems may be able to benefit from systems where moderate critical loads can be handled by substations or mesh networks employing current limiter technology, while heavy loads are handled by specialized dedicated power links with the appropriate levels of redundancy. Much research remains to be done in this area, however.

CHAPTER 5

TECHNOLOGY ISSUES

Although component technology evolves rapidly in many respects, system technology tends to evolve slowly, largely because systems bring together many technologies which do not necessarily evolve in the same direction. The advantages of multiplexing have been well-known for decades, yet multiplexing is just beginning to be used in aircraft systems. The reasons for not using it earlier include traditional practices as well as technology issues. This chapter discusses several important technology issues as they relate to contemporary and advanced communication systems. Topics covered are:

- Integrated circuits
- Packaging
- Transmission media
- Lightning susceptibility
- Software
- Terminal design

The chapter concludes with a concept of a 1553-compatible network node.

5.1 Integrated Circuits

Except for discrete interfaces, digital channels require substantial amounts of circuitry. Even a simplex serial channel requires serial-parallel conversion at either end, timing circuits, and sequential control circuitry. More complicated channels, such as 1553 multiplex buses, require this and more. A typical 1553 interface requires the order of a hundred medium-scale integrated (MSI) devices and possibly some large-scale (LSI) bit slice devices. This may be a tolerable cost if the number of interfaces is few, but communications for active-control airplanes will require many interfaces. A cost incentive therefore exists to realize complex interfaces in very-large-scale integrated circuits (VLSI) form. In commercial data processing, a serial communication protocol called SDLC is used in very large

quantities, which has justified the development of a VLSI circuit for the interface. So far, however, aircraft data interface volume has not been great enough to justify development of VLSI circuits comparable to the SDLC circuit. It is reasonable to expect, though, that such circuits will be available in a few years time.

Given an adequate market for VLSI interface circuits, some questions emerge concerning environmental tolerance and reliability. As to the first question, it has been generally true that LSI circuits can be made to withstand military environments. There is no particular reason to expect that VLSI can not also be made to do so. Also, by the 1990's, it is possible that new semiconductor materials, notably gallium arsenide (GaAs) will be available in quantity for VLSI circuits. The temperature range for GaAs devices is substantially greater than that for silicon devices.

The second question, concerning reliability, has been the subject of some debate in the past. The development of VLSI reduces the interconnection level, but it has been conjectured that this would be offset by the fragility of the device itself. MIL HDBK 217C dated May 1980 gives the most recent data for calculating the reliability of LSI devices. The handbook also contains reliability information on interconnections. It is an interest exercise to use this reference to look at the effects of going to higher complexity microcircuits on reliability. The equation for computing the reliability of LSI devices is:

$$\lambda_P = \pi_Q \left[C_1 \pi_T \pi_V + (C_2 + C_3)\ \pi_E\right] \pi_L$$

where:

λ_P is the device failure rate in $F/10^6$ hours

π_Q is the quality factor

π_T is the temperature acceleration factor, based on technology

π_V is the voltage derating stress factor

π_E is the application environment factor

C_1 and C_2 are device complexity failure rates based upon gate count

C_3 is the package complexity failure rate

π_L is the device learning factor

It is assumed for purposes of comparison that there is an option of doing a particular job with 50x100-gate circuits or 5x1000-gate

circuits or 1x5000-gate circuit.

Table 5.1-1 gives the values for the constants to be used in calculating the reliability of the various sized circuits.

In selecting the constants in Table 5.1-1 several assumptions have been made. Quality level B parts and an air transport environment have been assumed. It has also been assumed that the parts come from an established line and process. A maximum junction temperature of 105°C for CMOS technology was selected as realistic. The number of leads per package was set arbitrarily at 22 for 100 gates, 40 for 1000 gates and 64 for 5000 gates. The calculated failure rate for each of the three devices is shown at the bottom of Table 5.1-1.

The failure rate for the semiconductor parts of the 3 versions of realizing the function would be:

50x100 gate devices	50x.240 = 12.000 $f/10^6h$
5x1000 gate devices	5x.437 = 2.185
1x5000 gate devices	1x1.140 = 1.140

This shows over a tenfold reduction in failure rate simply by going to more complex LSI devices. If we chose to compare the 5000 gate LSI with a small scale integration SSI and medium scale integration MSI realization, the gain in reliability would be even more dramatic. This is not the total story, because to make a fair comparison between the reliability of different degrees of integration we must also calculate in the reliability of interconnecting the 50-100-gate packages and the 5x1000 - gate packages.

If we assume that a multilayer board is used to interconnect the packages in the two less-integrated examples and that the number of holes in the board is 1.5 times the total pins on the packages interconnected, we find that we have an additional failure rate for the 50 package case of 8.25 $f/10^6$ hours and for the 5 package case 1.5 $f/10^6$ hours.

In addition we must count the failure rates of the solder joints between the package leads and the multilayer printed circuit card. If we assume wave soldering this is .319 $f/10^6$ hours for the 50 package version and .058 $f/10^6$ hours for the 5 package version.

TABLE 5.1-1

RELIABILITY EQUATION PARAMETERS

	100 GATES	1000 GATES	5000 GATES
π_Q	1	1	1
C_1	.011	.021	.053
π_T	21	21	21
π_V	1	1	1
C_2	.0007	.001	.0017
C_3	.008	.015	.025
π_E	4	4	4
π_L	1	1	1
λ_P	.240	.437	1.140

The total failure rates for the three realizations of the function are:

50x100 gate devices	12.000	
50x1650 thru hole MLB	8.250	
50x1100 solder joints	.319	
		20.569 f/10^6 hr.
5x1000 gate devices	2.185	
5x300 thru hole MLB	1.500	
5x200 solder joints	.058	
		3.743 f/10^6 hr.
1x5000 gate device	1.140	
		1.140 f/10^6 hr.

The reliability benefits of going to higher levels of integration are obvious.

5.2 Packaging of Dispersed Electronics

There are substantial reasons for centralizing electronics in aircraft and there are equally impressive reasons for decentralizing electronics. The electronics packaging engineer looks upon centralized electronics as a means of easing his burden. A central compartment can easily be pressurized, cooling can be provided by a central system, the location can be selected for low ambient vibration, interconnections between subsystems are short, and power conditioning can be done more efficiently. Centralization in turn can reduce equipment cost and maintainance cost and can improve reliability. It can also ease thermal, vibration, and moisture-related design problems. The Navy has funded efforts for the past three years to develop a Standard Aviation Module and the Navy is moving more toward centralization in its Standard Electronics Module Program where modules are plugged into large environmentally-controlled cabinets.

System designers, on the other hand, have determined that the degree of reliability necessary for future flight-crucial systems dictates that the electronics not be centralized. The electronics must be distributed so that a damage event which does not render the airframe unflyable, also does not cause the flight-crucial control systems to fail. There is also a strong impetus to decentralize electronics which can reduce the weight of wire in the airplane. It therefore seems likely that there will be an increase in electronics decentralization over the next few decades despite the packaging engineers' abhorrence of this trend. It is likely that much of this distribution can be done within the confines of multiple electronics bays where many of the advantages of centralization can be maintained. However, there are items such as sensors, effectors, and engine controllers, from which it will become less and less desirable to bring all the raw data to the bays. Also there is a trend toward replacing hydraulic and mechanical controllers, sensors, and effectors, which are by nature distributed, with electrical devices and electronics. Fault-tolerant data communication systems will also add to the distributed electronics because of the desirability of placing nodes with, or near, the system's subscribers.

In view of the increasing pressures to decentralize the electronics, it seems prudent to examine ways in which the penalties of distributed electronics can be minimized. The most difficult environmental factor in localized electronics is temperature.

Electronics which does not see appreciable aerodynamic heating, nor significant power dissipation in adjacent equipment, can regularly see temperatures to -55°C and lower. On the other hand, heating from engines, skin friction, and the sun can create temperatures at which most electronics components cannot reliably function or cannot function at all. In general, MOS devices are more temperature-sensitive than TTL and are often only specified for 0 to 70°C operation. Fiber-optics devices are also more limited in temperature capability than the military (-55°C to 125°C) temperature range. In addition, failure rates increase rapidly at higher temperatures, and failures are also increased by highly variable temperatures which create mechanical stresses and fatigue in electronic components and assemblies. Unfortunately there is no single solution to this problem. One must select component technologies which have the best temperature capabilities. This selection must take into consideration the widely different power requirements, and thus self heating, of different semiconductor technologies. The packaging engineer must minimize thermal excursions and high temperatures. This will undoubtedly mean some sort of active thermal control in the form of cooling and/or heating, and may also require thermal isolation for the electronics. Engine controllers may be cooled by fuel if the fuel is cool enough itself. Thermoelectric coolers can also be used as heaters by reversing the current flow, and they would be an excellent solution to the environment modification problem if they did not suffer from reduced reliability themselves at higher temperatures. Other possibilities include the use of conditioned air or the use of air ducted in locally from the outside. One might also consider liquid flow systems including the hydraulic system with appropriate heat exchangers or refrigerators as a means of cooling.

Vibration is an easier environment to modify than temperature. Standard approaches should work well. The solution is to mount the electronics package so that the vibration of the surface on which it is mounted is not transmitted through to it. This works well for high frequencies, but loses effectiveness below 100 Hz. Low frequency vibrations are not a problem if the electronics package is designed so there are no resonances below 100 Hz.

Electronics replacement can be facilitated by providing the same plug-in features found in rack-mounted equipment. Essentially, the act of placing the equipment and engaging the holddown also engages the connector or connectors. To achieve this requires all signal and power wiring and all fiber optics to be terminated to a connector or

connectors which are in turn mounted to the equipment mounting bracket. Although this places an additional burden on the wiring installation, it will not only ease equipment replacement, it will reduce harness-wire and fiber-optic breakage. The shock mounts can support the bracket, and thus the electronics, or they can be placed between the electronics and bracket, in which case the connectors must be allowed to float relative to the bracket.

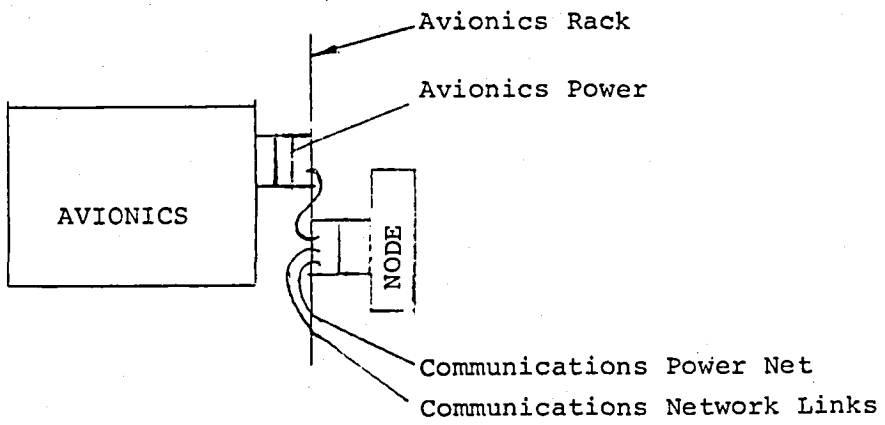
Equipment placed outside the pressurized envelope will require an environmental seal and must utilize connectors which are environmentally sealed.

In summary, distributed electronics should be in an environmentally sealed package and be mounted on vibration isolators in such a way that all electrical and optical connections are made in the mounting operation. In addition, the package should be provided with a means of thermal control which will keep components in a temperature range which will not unduly accelerate their failure.

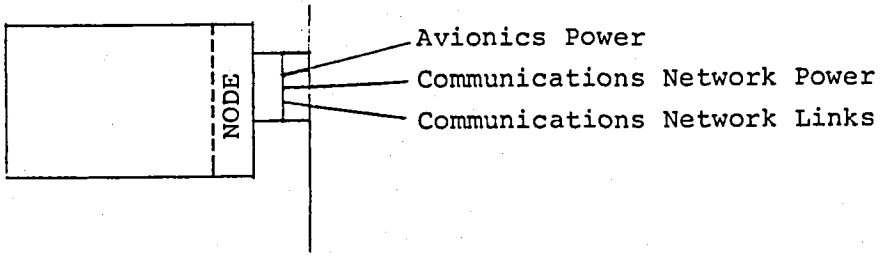
Communication network node packaging can be broken into two general problems. Nodes which are associated with electronics in environmentally controlled bays will benefit from the thermal control and mechanical isolation that are provided for the electronics. Nodes which are associated with remote sensors or effectors will be provided with their own packaging and must survive in the harsher thermal and mechanical environments of these locations.

In the first case, there is a possible problem with embedding a node directly in an electronics box. If the box is removed for any reason, the node is also removed from the network, which puts neighboring nodes in greater jeopardy of being isolated in the same way that a failed node would. In military aircraft which change equipment for different missions, this would be a problem. The solution is to make the nodes a part of the equipment rack as shown in Figure 5.2-1a rather than a part of the boxes which plug into the rack. Commercial passenger aircraft might upgrade equipment during the life of the airframe, but these changes take place only occasionally. It is therefore assumed that the node will be an integral part of the electronics with which it is associated as shown in Figure 5.2-1b.

When a node is associated with a remote sensor or effector, a different tack must be taken, particularly if the MTBF of the item with which it is associated is on the order of, or longer than, the MTBF of the node. Ease of maintainance requires that the node be separately



a. Rack Mounted Nodes



b. Nodes Integrated in Avionics

Figure 5.2-1. Rack Mounted Equipment.

removable. This can be accomplished by having the node attach and plug into the sensor or effector as shown in Figure 5.2-2b. If this makes the node inaccessible, or if it causes the node environment to be too harsh, the node could be mounted to the airframe near the item which it services, as shown in Figure 5.2-2c.

In order to consider what the node packaging might consist of, it is necessary to have some idea of what the node will consist. For purposes of discussion it is assumed that the node will contain between five and twenty silicon circuits, and that the total lead count for these circuits will be between 100 to 400 leads. In addition to the silicon circuits, there may be a small number of components associated with decoupling, line termination, lightning protection, and auxiliary power. It is estimated that there may be five to twenty of these components with a total lead count of ten to forty.

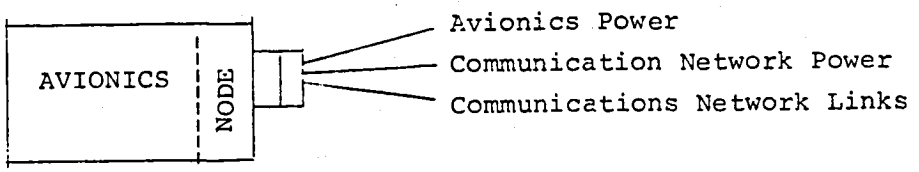
A network node must be associated with at least three links, and it is doubtful that it would be connected to more than four. Each link contains a sending and receiving line, and therefore a node would have to have connectors for six to eight twisted pairs or six to eight fiber optic lines. The node must be powered, probably via a power network dedicated to the communication system. This network will likely use its own return lines (twisted pair) and therefore there will be six to eight connector pins devoted to node power connections. There will also be two or more connector pins devoted to sensor or effector inputs to the node. The node power is estimated not to exceed five watts and probably will be much less.

Figure 5.2-3 shows schematically the contents of a node. The sense which this figure portrays is that a substantial number of connections are required for a small volume of electronics.

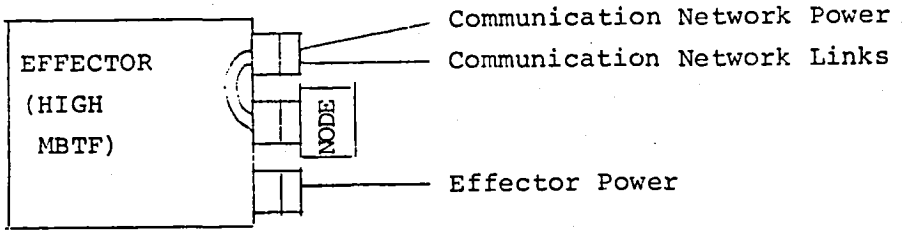
Packaging trends are toward the use of chip carriers and hybrid circuits. Table 5.2-1 shows the advantages, spacewise, of these approaches over older packaging techniques.

Chip carriers are receiving impetus from many fronts. There are JEDEC standards for several varieties of carriers. Texas Instruments has started to sell standard products in chip carriers, and although the price is now 8 times that of a comparable DIP-product price, TI projects that the price will be comparable by 1985.

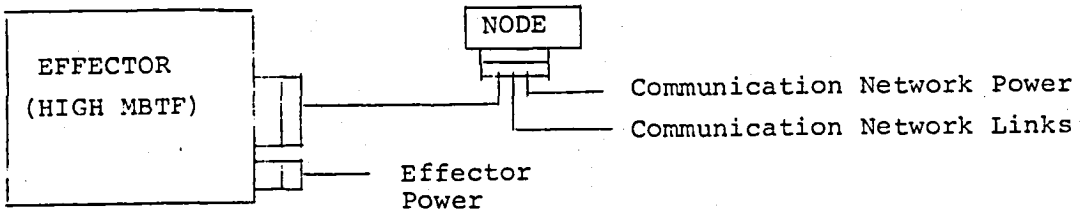
Chip carriers are typically interconnected using planar wiring structures made from epoxy glass, polyimide, alumina, and so forth. The truly leadless chip carrier must have a substrate which closely



a. Nodes Integrated in Avionics.



b. Node Mounted on High MBTF Effector.



c. Node Mounted Near High MTBF Effector.

Figure 5.2-2. Remote Equipment.

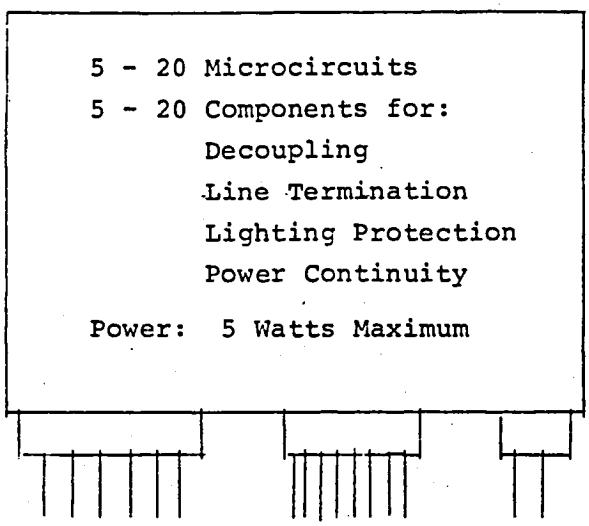


Figure 5.2-3. Node Contents.

matches the thermal coefficient of expansion of the wiring structure if they are to operate over a large temperature range. If they do not match, the solder joints between the carrier and wiring structure fatigues and cracks. The hermetic chip carriers are made with bodies of alumina, and this implies an alumina wiring board.

TABLE 5.2-1
PACKAGE DENSITITES

Page Type (16 Leads)	Number of Packages That Can Be Interconnected By 6 CM^2 or 1 IN^2 Of Multilayer Board
Dual In-Line Package (DIP)	2-3
Flatpacks	4-5
Chip Carrier	9-11

There are at least two efforts underway which may change this picture. One involves a leaded chip carrier which will provide stress relief and thereby allow alumina chip carriers to be reliably mounted to other than alumina wiring boards. The other involves the development of a chip carrier ceramic material with a higher expansion coefficient than alumina to more closely match epoxy glass. Even if these developments are successful, it still may be desirable to use an alumina wiring board to minimize the thermal drops between the semiconductor junctions and the node case.

Hybrid circuit technology is often thought of as an alternative where space is at a great premium. However, it is also frequently used to produce standard modules which cannot be made on a single silicon chip. Hybrids have even become popular for automotive applications when long life under severe environment is required. By 1990 it is likely that it will be practical to tool for production quantities of node hybrid circuits which will be price-competitive with nodes made using chip carriers.

If a node is made using chip carrier technology, the electronics assembly will be from 10 to 40 CM^2, or 1.5 to 6 IN^2 by about 0.5 CM, or 0.2 IN thick. If it is made using hybrid circuit technology it would fit in a flatpack from 3 to 10 CM^2, or .5 to 1.5 IN^2, by 0.25 CM or

0.1 IN thick. In either case the size of the electronics is small. If it is integrated into an electronics assembly it will fit into a single module, or perhaps onto a portion of a module. If it is made as a separate unit, it will be a simple small box, with a single connector, and will either be held in place by fasteners or a quick release locking mechanism.

It is anticipated that such a small package can be made rugged enough so as not to require mechanical isolation. Furthermore it is small enough so that heat sinking can reduce the electronics temperature rise to a minimum, and thereby permit reliable operation without active cooling.

Active cooling in the form of thermoelectric devices may be considered, but is apt to be rejected for several reasons. The devices are expensive. Their reliability is a strong function of heat sink temperature, decreasing with increasing temperature. Lastly, it is felt that all nodes can be placed so that their ambient temperatures will fall within the range of -55°C to 71°C and that the semiconductor junctions can be maintained below 105°C, which will provide for highly reliable operation.

5.3 Transmission Media

There are many types of transmission media which may be applicable to designing communication links. These include the following:

Electrically conductive types:

- Single wire with common ground
- Twisted pair
- Twisted shielded pair
- Twisted double shielded pair
- Shielded balanced pair (4 conductors)
- Planar parallel
- Microstrip
- Stripline
- Coaxial
- Triaxial
- Twin-axial

Free space - for radio through optical frequencies

Radio frequency wave guides

Optical wave guides

In selecting the appropriate transmission medium for the network links the following characteristics are important:

Signal Transfer Properties

The communication links must be capable of transmitting digital signals over distances of a few hundred feet at up to a ten-megabit rate.

EMI Properties

The communication links must provide adequate rejection for all spurious signals. These include noise from electrical generating equipment, radios, radars, relays, and lightning.

Environmental Properties

The links must provide an adequate service life under diverse environments which may be as benign as a controlled-temperature equipment bay, or as difficult as an engine controller.

Reliability

The reliability of the link as a whole (including the link driver and receiver) must be adequately high. In many cases the drivers, receivers and connector pins will have a shorter MTBF than the transmission medium, and therefore will strongly influence any decision to use the medium.

Cost

Cost includes the initial cost of the medium, its installation and maintainence, and the cost of all necessary interface electronics and connectors.

Size and Weight

Deltas in the size and weight are directly translatable into a delta cost of operation for a given aircraft.

Some of the transmission choices can be quickly disposed of. Since line-of-sight optical transmissions are impractical in the complex aircraft structure this is certainly not a practical system. Radio transmissions inside or outside the aircraft would also be

highly perturbed by the aircraft structure, and are not practical for this reason and several others.* The use of RF wave guides is impractical from weight and cost considerations. Thus we are left with the electrically conductive transmission lines and with optical wave guides.

5.3.1 Conductive Media

A study by SCI Electronics, Inc., of transmission systems for the space shuttle, covered many of the pertinent points in selecting from among the electrically conductive types of transmission media.

The use of a single wire with a common ground and unshielded twisted pair would be entirely unsatisfactory from an EMI point of view, and will not be considered further. The parallel conductor techniques such as planar parallel, microstrip, and stripline offer particular advantages when many lines are required from one point to another, which is not the case in the communications network. If such conductors were considered seriously for an aircraft environment, the stripline would provide the most EMI protection, but then it is also the most expensive and difficult to terminate of these parallel conductor techniques.

The SCI report gives information on various tests of twisted shielded pair, twin axial, co-axial, triaxial, and shielded balanced pair. In their characterization of electromagnetic interference, three types were considered: Low-frequency steady-state field generated by AC power alternators, high-frequency steady-state fields generated by communication equipment or radar, and impulsive noise caused by relays, spark ignitors, and the like. No comments were made on the effects of lightning, which is covered separately in this report. Because of the large number of relays in the space shuttle, an unsuppressed relay was selected as a noise source in making a comparison of RG-110 coaxial, TRC-75-2 (triaxial), twisted shielded pair #22 AWG, and TWX-124-2 (twin-axial). This listing is in order of increasing immunity to EMI effects from a relay with the addition of a 2 MHz filter. The noise on the twisted pair was reduced four-fold. The recommendations from this series of tests were:

- A balanced transmission line and receivers be used

*Radio transmission is considered a viable medium for entertainment multiplex.

- That the cable be grounded as near receiver ground as possible
- Low-pass filtering should be employed at the receiver.

EMI rejection at low frequencies was shown to be best with twin axial, next best with twisted shielded pair and worst with coaxial.

The conclusion drawn from this extensive test program is that the use of "rather inexpensive and light weight twisted shielded pair #22 appears adequate for EMI considerations ..." For the fault-tolerant communication network it may be necessary to use extra shielding or to use the more expensive and heavier twin axial cable consisting of two twisted pair conductors and two dielectric spacers with an overall shield.

Environmental considerations for a conductive transmission medium fall mainly into three areas:

1. The ability of the materials to physically survive the temperature range to which they will be subjected.
2. The change in transmission characteristics with temperature.
3. The ability of the medium to withstand the mechanical stresses which cause abrasion and cold flow.

The high temperature limit for service of copper and silver plated copper is 120°C, while tin plated copper is good to 150°C. The mechanical and physical properties of insulating materials are shown in Table 5.3-1. The chemical properties of insulating materials are shown in Table 5.3-2, and the electrical properties of insulating materials are shown in Table 5.3-3. Several of these materials are reasonable candidates for insulation in the communication network link conductors. The SCI Systems report indicates that changes in attenuation arise mostly from changes in the conductor resistivity. Changes in phase shift and impedance at high frequencies are proportional to changes in the square root of the dielectric constant. For example, the dielectric constant of Kapton changes from 3.5 at 25°C to 3.0 at 200°C which would cause a 9% change in phase shift and impedance. It is not anticipated that changes of this order will cause problems.

The selection of an insulating medium for communication link conductors becomes a matter of determining the minimum cost material

TABLE 5.3-1

MECHANCIAL AND PHYSICAL PROPERTIES OF INSULATING MATERIALS

INSULATION	COMMON DESIGNATION	TENSILE STRENGTH, PSI	ELONGATION, %	SPECIFIC-GRAVITY	ABRASION RESIST.	CUT-THROUGH RESIST.	TEMPERATURE RESISTANCE (MECHANICAL)
POLYVINYL CHLORIDE	PVC	2,400	260	1.2-1.5	POOR	POOR	FAIR
POLYTETRAFLUOROETHYLENE	TFE	3,000	150	2.15	FAIR	FAIR	EXCELLENT
FLUORINATED ETHYLENE	FEP	3,000	150	2.15	POOR	POOR	EXCELLENT
MONOCHLOROTRIFLUOROETHYLENE	KEL-F°+	5,000	120	2.13	GOOD	GOOD	GOOD
POLYVINYLIDINE FLUORIDE	KYNAR*	7,100	300	1.76	GOOD	GOOD	FAIR
POLYIMIDE FILM	Kapton**	18,000	707	1.42	EXCEL.	EXCEL.	GOOD
POLYSULFONE	-	10,000	50-100	1.24	GOOD	GOOD	GOOD
POLYIMIDE-COATED TFE	TFE/ML**	3,000	150	2.2	GOOD	GOOD	GOOD
POLYIMIDE-COATED FEP	FEP/ML**	3,000	150	2.2	GOOD	GOOD	GOOD

*Trademark, Pennsalt Chemicals Corp.
**Trademark, E.I. du Pont de Nemours & CO.
+Trademark, 3M CO.

TABLE 5.3-2

CHEMICAL PROPERTIES OF INSULATION MATERIALS

INSULATION	COMMON DESIGNATION	PROPERTY				
		FLUID RESISTANCE	FLAMMABILITY	RADIATION RESISTANCE, RADS GAMMA EXPOSURE	TEMPERATURE RESISTANCE	COMMENTS
POLYVINYL CHLORIDE	PVC	GOOD	SLOW TO self-extinguishing	10^6 - 10^-	-55-105	
POLYTETRAFLUOROETHYLENE	TFE	EXCEL.	NON-FLAMMABLE	10^6	-80-260	
FLUORINATED ETHYLENE PROPYLENE	FEP	EXCEL.	NON-FLAMMABLE	10^6	-80-200	
MONOCHLOROTRIFLUOROETHYLENE	Kel-F	GOOD	NON-FLAMMABLE	10^6	-80-200	FLUIDS TEND TO PERMEATE @ HI TEMP.
POLYAMIDE FILM	KAPTON	EXCEL.	NON-FLAMMABLE	10^9	-80-260	
POLYSULFONE	POLYSULFONE	FAIR	SELF-EXTINGUISHING	--	-65-150	SOLUBLE IN CHLORINATED HYDROCARBON
POLYIMIDE-COATED TFE	TFE/ML	EXCEL.	NON-FLAMMABLE	10^6	-80-260	
POLYIMIDE-COATED FEP	FEP/ML	EXCEL.	NON-FLAMMABLE	10^6	-80-260	

TABLE 5.3-3

ELECTRICAL PROPERTIES OF INSULATION MATERIALS

INSULATION	COMMON DESIGNATION	PROPERTY			
		DIELECTRIC STRENGTH, VOLTS/MIL	DIELECTRIC CONSTANT, 10^3 Hz	LOSS FACTOR, 10^3 Hz	VOLUME RESISTIVITY, OHM-CM
POLYVINYL CHLORIDE	PVC	400	5-7	0.02	2×10^{14}
POLYTETRA-FLUORO-ETHYLENE	TFE	480	2.1	0.0003	10^{18}
FLUORINATED ETHYLENE PROPYLENE	FEP	500	2.1	0.0003	10^{18}
MONOCHLORO-TRIFLUORO-ETHYLENE	Kel-F	431	2.45	0.025	2.5×10^{16}
POLYVINYLIDINE FLUORIDE	Kynar	1,380 (8 mils)	7.7	0.02	2×10^{14}
POLYIMIDE FILM	Kapton	5,400 (2 mils)	3.5	0.003	10^{18}
POLYSULFONE	-	425	3.13	0.0011	5×10^{16}
POLYIMIDE-COATED TFE	TFE/ML	480	2.2	0.0003	10^{18}
POLYIMIDE-COATED FEP	FEP/ML	480	2.2	0.0003	10^{18}

which is adequate for the job. The development of new insulations, and variations and combinations of old ones continues, and it is a certainty that a listing of wire insulations in the year 1990 or 2000 will incorporate new members worthy of consideration. The fact remains that there are adequate insulations available now to do the job.

5.3.2 Optical Media

Fiber optics is an emerging technology which offers significant advantages over metal conductors, especially when applied to telecommunication systems.

If one considers the application of fiber optics to the fault-tolerant fully embedded avionic system for the 1990 to 2000 time frame, the advantages are not as clearly defined as for telecommunications. This results from the nature of the application and the constraints imposed by the environment. New components could greatly modify this picture by 1990, however.

Fiber optics has gained acceptance because of its ability to transmit information along a glass fiber over long distances with little attenuation. Cable losses of less than 1 dB/km have been reported with new records for long distance transmission being announced virtually daily. The low attenuation figure means that information can be transmitted using far fewer repeaters than are presently required by copper cables. In communication, the trend has been toward very fine single fibers to minimize dispersion.

Wide bandwidth is another important advantage of fiber optics. Bandwidth up to one gigahertz is possible with 50 to 3000 megahertz being a more common range. Injection laser emitters and avalanche photo diode detectors are usually employed to achieve these high bandwidths. Laser emitters provide a high optical power output, a fast response time (typically less than 1 nsec) and a spectral width less than 1 nanometer. Their high optical power output allows transmission of a signal through a long fiber. A narrow spectral width reduces pulse spreading.

Light-emitting diodes (LED's) which are also commonly used as emitters, have lower optical power output, a slower response time (typically 20 to 30 nsec) and a spectral width of about 50 nm. LEDs can operate with a maximum bandwidth of about 20 megahertz.

Another important advantage of fiber optics is its immunity to electromagnetic interference (EMI), electromagnetic pulses (EMP), and

lightning. The fiber optics cable is immune to these radiation sources and does not radiate to other fibers or to copper wire in the area. Therefore there will be no crosstalk between adjacent fibers or cables. These same characteristics also make it difficult to tap the line, and therefore fiber optics provides better security than wire cables.

Telecommunication fiber optics with its small fibers (in the order of 100 μm dia.) and high bandwidth also provides low cost and weight. A small fiber optics cable can replace a large copper wire cable. The lower cable weight and size makes it easier to handle. It also costs less based on its data handling capability.

A fault-tolerant embedded avionic system will have different requirements for the application of fiber optics from those of the communication industry. These requirement changes are imposed by the environmental constraints and differences in the application. In remote sections of the aircraft, temperatures can extend from -55 to 100°C and can cycle between the extremes during normal operation. Atmospheric pressure and humidity vary over significant ranges. Mechanical stress can be imposed on the cables by shock and vibration when located near the engines, auxiliary power unit, landing gear, etc., as well as during installation. A fully embedded avionics communication system has design constraints imposed by branching and routing the cable through the structure, requirements for many connectors, and constraints imposed by maintenance procedures.

For a long telecommunication line, attenuation is of extreme importance because it affects the number of repeaters that will be required to span a given distance. In the aircraft environment, the longest data transmission lines will be typically 30 meters with a maximum of 100 meters. With a line of 100 meters, line attenuation is not particularly important. However, coupling efficiencies to the detectors, emitters and multiple connectors are still important. Instead of a very small single mode fiber that would be ideal for land-based communication, an avionic application may be better served by a multifiber cable with a large numerical aperture. The multifiber produces redundancy, so that one broken fiber has little effect on transmission, and the large numerical aperture increases the coupling efficiency.

The problem of fiber termination on board an aircraft must be solved. The fiber is usually terminated by mounting its end in a fixture and polishing it to a fine surface finish. There are also

techniques to scribe and break single fibers with special tools. The connector attenuation is strongly affected by the technique used. In addition, when connectors are disconnected, their ends must be protected from dirt and scratches. Since dirt, scratches and misalignment can increase connector attenuation, point-to-point attenuation checks may be required to assure that a fiber optics link has retained sufficient margin to operate over temperature.

It is important that the reliability of the communication links in an avionic system be high both for safety and for minimizing maintenance costs. The avionic thermal environment has a great effect on fiber optic emitters and detectors. Even though lasers have high optical power output, very small beam width dispersion and very high response time, they are extremely temperature-dependent and have short life times. Development is going on to increase life times and to decrease temperature dependence.

Light emitting diodes, LED's, are the next best thing, but they still have a limited temperature range of about 0 to 70°C which is not adequate for all applications in all sections of the aircraft.

LED's also exhibit a few basic properties that could be a problem to a high reliability system. One is that the optical output of an LED decreases with increasing temperature. As one drives the LED harder it is possible to have little optical power increase due to the increase of temperature caused by increased dissipation. Second, high temperatures decrease the life of the LED. Third, LED light output decreases with time. For a fault-tolerant avionic system this means that a pre-flight margin check on emitter light output may be required to determine whether a link will operate over temperature or if the link has technically failed.

Avalanche photodiode detectors are almost as bad as lasers, as regards temperature effects. At the present time the PIN diode matches the LED in performance and has a 0 to 70°C temperature range. The LED/PIN combination seems to be the best match. With these devices a bandwidth up to about 20 megahertz is possible. New developments will likely improve this temperature range.

Fiber optic cables are not completely immune from the aircraft environment. Some reports indicate that fiber attenuation does change under temperature and that stress corrosion, which is the propagation of micro-cracks in glass, can result in failures.

A recent development in fiber optics is the transmission of power optically. This technique is presently being proposed for control systems in hazardous environments and is also being tested for lower power applications. Power efficiencies of 33.4% using GaAs diodes receiving 820 nm wavelength have been measured by Sandia Laboratories. This techniques could help solve the EMP and lightning problem in aircraft by providing electrical isolation in the power lines.

Fiber optics can be designed to survive lightning and EMP. In addition, fiber optics provides data lines that are electrically decoupled from other subsystems. With a properly designed power distribution network, i.e. one that also provides D.C. electrical isolation, the subsystem can be completely electrically isolated from other systems. Without a current path through the subsystem box, the lightning current can be constrained to the outer surface of the aircraft. This single feature could be the impetus that propels the use of fiber optics in aircraft in the 1990 to 2000 time frame. A necessary prerequisite will be optical component improvements which will make their use in the aircraft environment practical.

5.4 Lightning Effects and Countermeasures

Material in this section concerning lightning effects was derived primarily from Reference 9. It is instructive to quote some of the preface of this reference as an introduction to the subject.

"The impetus for writing this ... springs from two sources — the increased use of nonmetallic materials in the structure of aircraft and the constant control and navigation functions. Nonmetallic structures are inherently more likely to be damaged by a lightning strike than are metallic structures. Nonmetallic structures also provide less shielding against the intense electromagnetic fields of lightning than do metallic structures. These fields have demonstrated an ability to damage or cause upset of electronic equipment."

"The persons who can best use aircraft protection from lightning are the aircraft designers and operators, but generally they are not among those who produced this information. Moreover they are often unaware of its existence, and they seldom have the background to distill from it the important facts that should be applied to achieve safer designs."

"Even though much has now been learned about lightning effects on aircraft and how to design protection, there are still some lightning

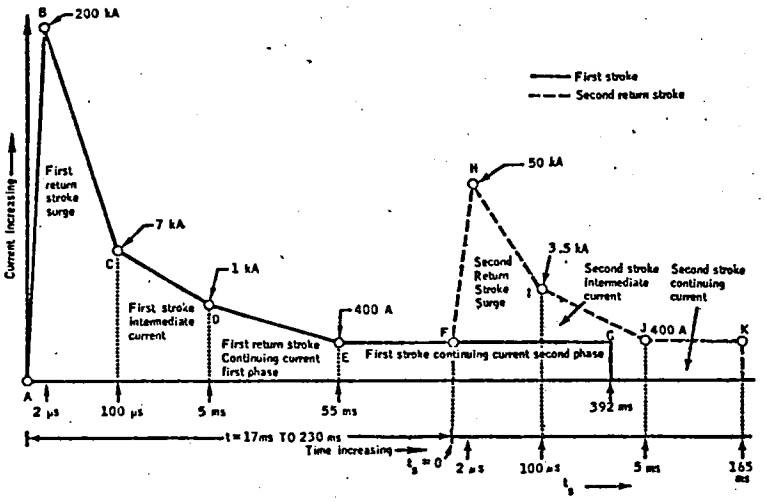
effects which are not fully understood. Examples of these are (1) the mechanism by which lightning currents diffuse into interior structural members and conducting parts together with the extent to which this happens and (2) the effects of electromagnetic radiation from the lightning arc upon aircraft electrical and electronic systems."

Lightning strikes commercial transport aircraft on the average of once every 3000 hours. The damage caused ranges from none to serious. In addition to structural damage, lightning can be disruptive to the power system and to electronics either by causing an intermittent malfunction or by causing permanent damage.

It is possible for lightning to couple directly into the plane's electrical system by striking through a non-conductive structure or by striking navigation lights. More often indirect effects are caused by voltages induced on wiring through electromagnetic field changes inside the plane and resistive voltage drops in the structure. A record of 214 strikes between 1971 and 1974 reported by a group of U.S. Airlines shows that malfunctions occurred in the electronics 20% of the time and half of these malfunctions were permanent in nature.

A communications network which reaches the total avionics complement is obviously a candidate for similar malfunctions, unless adequate measures are taken in the design of the network. A lightning stroke can last a significant part of a second, so it is important to minimize, if not eliminate, transient malfunctions. Indeed, this may prove to be more important than permanent damage to a small portion of the network. A diagrammatic representation of a lightning model is shown in Figure 5.4-1. This shows that the highest rate of change of current occurs during the first two microseconds of the primary and return strokes. This raises the prospect that transient malfunctions, which occur over only a short interval of the most severe strokes, may be handled by the system without adverse effects to the aircraft.

Transients are considered to be produced by three types of coupling to the interior of the airframe. First is the IR drop associated with the huge currents passing along the airframe. Second are magnetic fields drawn through apertures and through the skin, which is not a complete magnetic shield, and third are electric fields which enter entirely through apertures because the conductive skin provides excellent shielding against electric fields. The degree to which these effects occur varies greatly with location in the aircraft. The use of composites, which is expected to increase, will aggravate this



Diagrammatic representation of lightning model.
(Note that the diagram is not to scale.)

Figure 5.4-1. (From Ref. 9).

coupling problem.

Considerable shielding advantage can be gained by preferential routing of electrical conductors. In general, they should be laid as close as possible to continuously conducting skin or structure. They should avoid areas of small skin radii such as the leading edge of a wing. Routing in conduits is only beneficial if the conduits are continuously conducting, well grounded to structure and have all access holes conductively covered.

Because the airframe does not provide adequate shielding for lightning effects, it is necessary that the transmission system in a network contain supplemental shielding which will provide adequate protection. The primary (outside) cable shield should be highly conductive, be grounded at least at both ends, and have a minimum of holes and discontinuities. A solid shield is best. Tape wound shields are not recommended. A woven shield has enough small holes to allow significant coupling. However, it has been found that double layer woven shield can approach a solid shield in quality.

It is very important that the outer shield be continuous through the connector. The conductive connector shell should electrically contact the shield around its total circumference. It may be desirable to use an inner shield to provide even more protection. The intuitive reaction of many designers would be to ground this shield only at one end for best control of low frequency, low level interference. Some experiments have shown that in many cases it is better to ground both ends of the inner shield too for best protection from lightning.

Until more definitive experimental work is done for an actual application, the conservative approach would be to use double-shielded transmission line. In some high interference locations, the outer shield should be a double layer of braid. Extreme care should be taken in the design and execution of the shield terminations to minimize leakage apertures. Whether the inner conductor should be grounded at one or both ends will depend upon the length of the cable, location in the ground plane, other noise sources, and the details of the terminating circuitry for the conductor being protected.

Magnetically induced voltages may appear between wires of a two wire circuit, or between either wire and the airframe. The former voltages are referred to as line-to-line voltages and the latter as common-mode voltages.

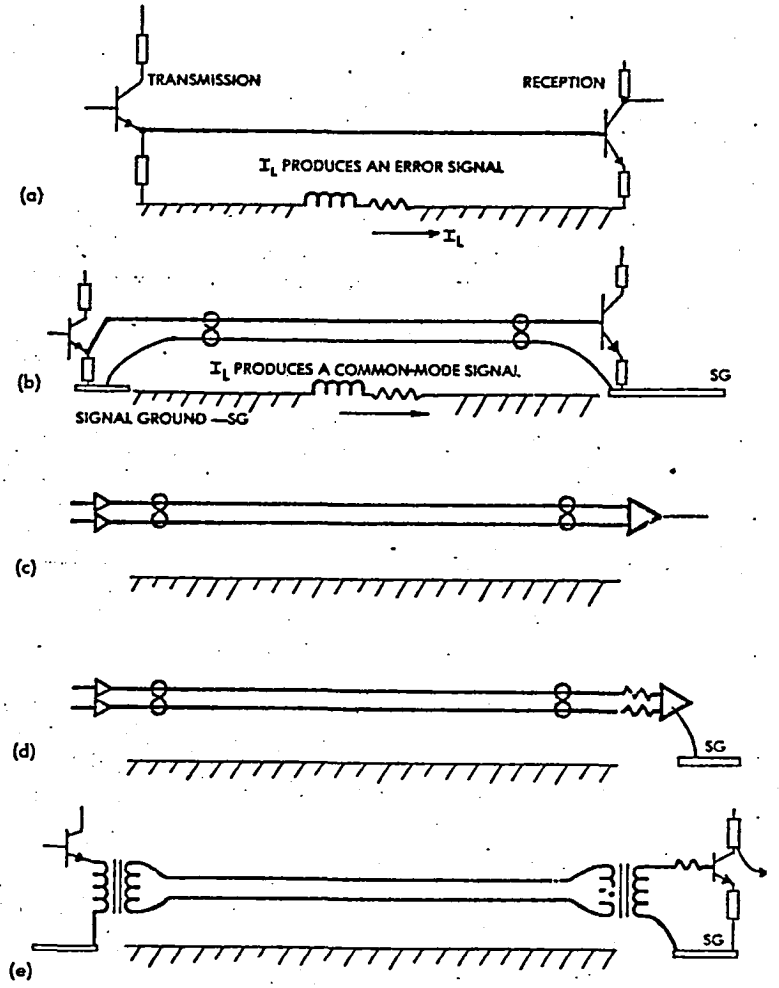
The noise produced by lightning has a broad frequency spectrum. The energy collected by a receptor increases with increasing pass band. Unfortunately, digital circuits are by nature broad band, and therefore other techniques than limiting bandwidth must be used to minimize noise effects.

In designing a network signal link several things can be done to minimize transient and permanent damage. First, the structure should not be used as a ground path as shown in Figure 5.4-2a. Second the use of twisted-pair transmission line, Figure 5.4-2b, does not preclude the appearance of common mode voltages unless differential transmission and reception devices are utilized as shown in Figure 5.4-2c. The use of resistors in series with semiconductor junctions, Figure 5.4-2d, will greatly increase the chances of transistor survival during transients. Transmission through balanced transmisson lines and transformers coupled with transistor input protection, Figure 5.4-2e, provides even more protection.

Certainly the power lines which power the nodes must also receive attention to minimize the noise they carry into the receiving and transmitting electronics of the node. The use of shielded twisted pair for power will offer the same advantages as for signal.

An assessment of the probability of lightning causing damaging transients in an actual application using the foregoing design principles must be made to determine the desirability of using transient protection devices. There are two basic types of transient protective devices capable of protecting against lightning produced transients. The first type is represented by Zener diodes and varistors, which by virtue of their non-linear current to voltage relation will shunt currents produced by overvoltages. Spark gaps, the second type, switch to a highly conductive state upon the application of an overvoltage and will not shut off until the voltage falls to a low level, which may require the removal of normal line voltage, depending on its level, to occur.

Direct lightning damage to wiring and avionics has occurred in numerous instances. Externally mounted hardware which offers a point of entry for lightning includes navigation lights, antennas, windshield heaters, and pitot tube heaters. One incident to a general aviation aircraft will serve to indicate the degree of damage which can be caused by such a strike. Lightning struck a navigation light which was mounted on a fiber-glass wing tip and which was not adequately grounded



Considerations regarding circuit design.

Figure 5.4-2. (From Ref. 9).

to conductive structure. The lightning vaporized the housing ground wire and shattered the lamp globe and bulb which allowed it to enter the power wires. The resulting damage included all the navigation light switches and all the lights, several instrument lights, two fuel tank quantity indicators, and a VHF communication set. Also 75% of the circuit breakers popped and only half of them could be reset.

Adequate grounding measures and inclusion of conductors in otherwise non-conductive structure can greatly lessen the change of this sort of damage. However, the consequences of lightning directly entering the power wiring of an active-control aircraft are potentially so serious that measures must be taken to limit the possible damage caused by such an event. This might be done by isolating the power for the communication system from power lines which are vulnerable to direct effects and/or by adding over-voltage protective devices to the communication network power lines.

5.5 Software

In all forms other than the dedicated link form of data communication, the medium is multiplexed. This means that it is shared among functional and structural components of the system. System controllers are almost always digital computers, and subscribers and nodes often have embedded computation elements. The sharing of the medium is therefore largely software-dependent. If the software fails, the entire medium, and hence the airplane system, can fail. Good software engineering practice is therefore necessary.

The software relating to node control is special, in that it is actually a form of firmware. The normal task to be accomplished is relatively simple and testable. The code is brief, and it cycles continuously in the node. It is hence reasonable to expect that the node control software will be made correct for normal operation. The problem is with respect to error recovery. Even though most failure modes can be covered rather easily, it will always be difficult or impossible to prove that node software is correct in response to errors, since the variety of sequential errors due to probable faults will most likely be too large to enumerate.

In order to meet the safety requirements for active control, the coverage of faults must be high enough so that the product of a fault's probability times its non-coverage probability is small enough to be negligible. High coverage can be provided by arming the system with

the capability of intervening in cases where one or more nodes misbehave. In mesh networks, nodes can be isolated from misbehaving neighbors. In addition, gateman circuits can be designed to reset and restart the computer in a node. In multiplex buses, watchdog timers and subscriber power control may be used for software fault defense as well as hardware fault defense.

5.6 Terminal Redundancy Design

The word "terminal" is used here to denote any node or subscriber that interacts with the data communication medium. In a system with redundant sensors, redundant effectors, and redundant controllers, the nature of the communications among these elements largely determines the system's ability to tolerate faulty elements. Systems that are organized into separate strings have the simplest communication structures, but have the drawback that a single element failure can render a quarter, a third, or half of the total equipment useless.

Systems that are more integrated than string systems require communications that allow more general associations among elements. Sensors will be able to be interrogated by redundant controllers, effectors will be able to be commanded by redundant controllers, and any controller may be called upon to read the sensors and command the effectors. The impact on communications is to require many or all terminals to be multiported. This is not as simple as joining wires, because a short circuit at one terminal could kill the signal for all of its destinations. It is standard practice, therefore, to use a separate buffer for each port.

Multiport sensors are easily arranged for analog and broadcast digital transmission. For command-response protocols, however, the complete interface must be replicated in order to allow independent access. Multiport effectors are more difficult. The terminal must decide which port, or what aggregation of signals, to follow; and provision has to be made for reconfiguration and testing. Mesh networks simplify sensor and effector interfaces by using multiport nodes.

Multiport controllers have no difficulty as far as design is concerned, but the number of interface circuits can be greatly inflated if the communications are all dedicated on a per-signal basis. The situation is somewhat better when broadcast buses or dedicated per box links are used, but the number of ports will still be in the hundreds. With multiplex buses, the controllers interface with each redundant

copy of the bus, which make them equivalent to sensors and effectors in this regard. With mesh networks, the controller can have an arbitrary number of ports.

5.7 A 1553-Compatible Network Node

The 1553 multiplex bus standard has no provision for network forms, and a mesh network would not be able to meet the standard. It is possible, however, to design a network that interfaces with standard 1553 remote terminal equipped subscribers. The concept of a node for such a network is shown in Figure 5.7-1. As explained in Chapter 3, the signals between nodes would be in a pulse code form derived from the Manchester biphase code. The reason for using pulse code is to allow instantaneous repetition of the signals at each node. If Manchester code is amplified and repeated many times, the signal is apt to be distorted out of tolerance.

At the upper left side of the figure are three full-duplex link interfaces, where the R's are receivers and the T's transmitters. Received signals are repeated and retimed in boxes labeled RP. Just to the right of the RP boxes is the link switching matrix, where switches labeled A through F selectively connect repeated signals to transmitters on other links. These switches are operated by the node control microprocessor at the extreme right of the figure. To the right of the link switching matrix are conversion circuits between pulse code and Manchester signal forms. Manchester signals from the receivers are selectively gated to the 1553 subscriber interface, and signals from the 1553 subscriber are selectively gated to converters that lead to the transmitters. Selections are again made by the node control microprocessor via outputs G through M. In order to prevent feedback of Manchester signals, the pairs of pulse code to Manchester and Manchester to pulse code converters are mutually inhibitory.

The ungated received signals, after conversion to Manchester are converted to parallel format at the right of the figure, where they are accessible via the parallel bus to the node controller. The node controller scans all received signals looking for gateman commands. Status replies are sent via the parallel to Manchester converter at the lower right corner of the figure.

Node power is shown to be supplied by a mesh network with current limiters at the power interfaces. The limiters limit outbound current, not inbound. Depending on the nature of the subscriber and the available technology, the node may supply subscriber power as well.

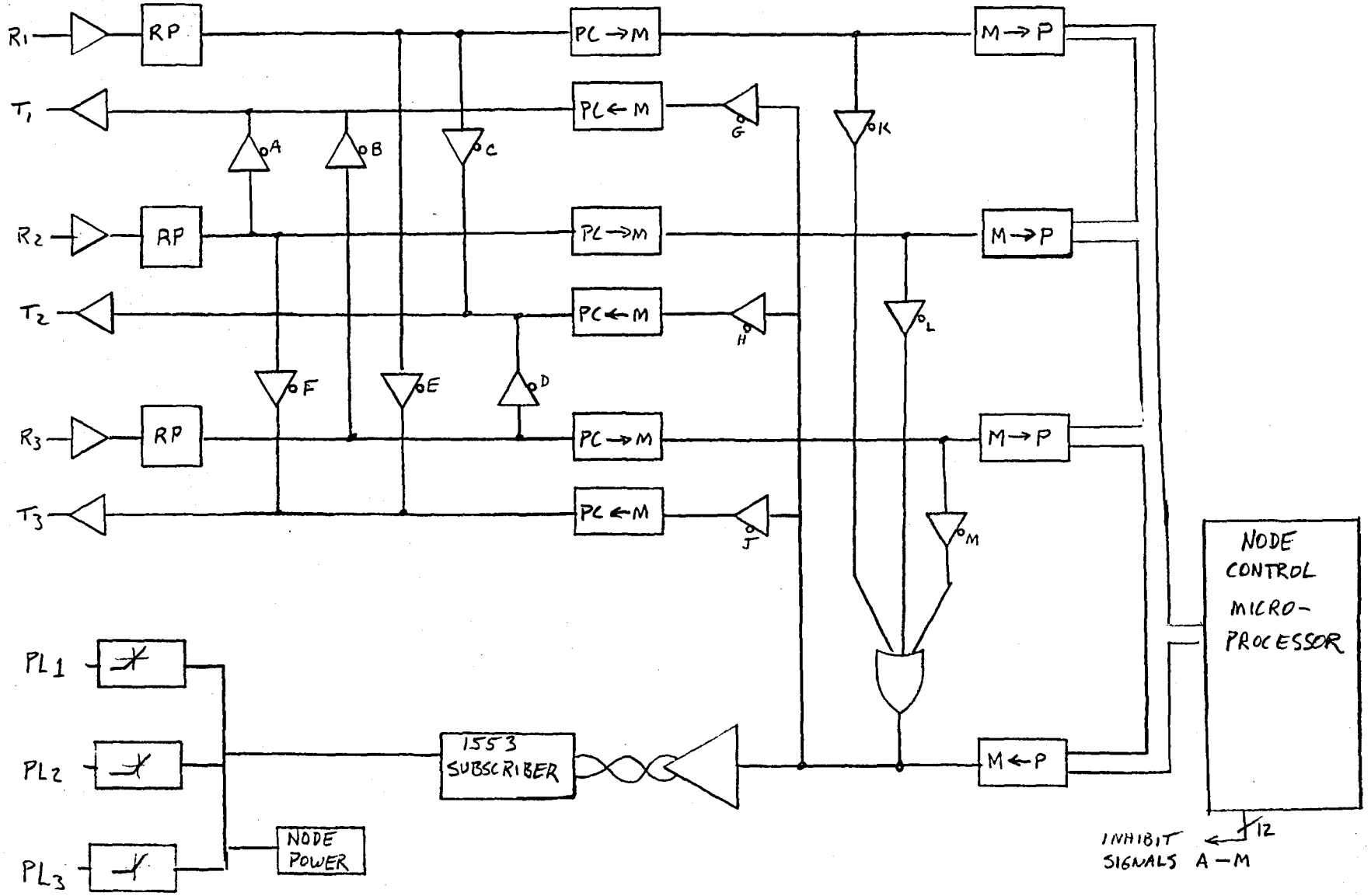


Figure 5.7-1. 1553 Network Node Concept.

CHAPTER 6

ANALYSIS AND MODELING DISCUSSIONS

In attempting to make quantitative judgements about communication and/or power systems, one is often able to use broad generalizations and approximations to arrive at useful "ballpark" estimates. In some instances, however, approximations are not obviously trustworthy, and in other instances approximations are not practicable at all. This chapter treats various approaches developed and/or used in this project to support the generation of trade-off data as well as qualitative conclusions.

Subjects treated here are the following:

- Network reliability
- Network connectivity
- Network dispatch probability
- Bus reliability and dispatch probability
- Remote power control reliability
- Power networking with current limiters
- Reliability analysis tool

6.1 Network Reliability

A network can fail in several ways. There are correlatable failures such as those caused by power failure, lightning, and damage events. There are random failures which occur as a natural consequence of the finite lifetime of the components of the system. It is the random failures and their effect on a network communication or power system which will be investigated here. There are two objectives to this study. One is to determine how the reliability of the network will vary with design variations. The other is to develop methods of analysis for networks which will provide a basis for predicting their reliability.

There are several ways of looking at the reliability of a network. If we consider the subscriber's point of view, the network has

failed if either the subscriber's node has failed or if this node has been cut off from the rest of the network by other failures. From the overall system view, a subscriber failure, a subscriber node failure, or an inability to communicate with an operational subscriber and its node, all amount to the same thing; the subscriber is lost to the system. It is this latter view which is the more germane and will be used in this analysis.

The node belonging to a subscriber has the effect of virtually reducing the reliability of the subscriber, because we are now interested in the failure rate of the node subscriber pair rather than the subscriber itself. There are two obvious ways of improving this situation. One is to improve the reliability of the nodes. A second method, which may be justified for every important subscriber, is to assign it two or more redundant nodes. There is little doubt that a system containing a fault-tolerant computer will require such a multinode interface to the computer.

It is appropriate at this point to consider what would be reasonable realibility numbers for a data communication node based on its electronic complexity. Figure 6.1-1 schematically shows a node with three links attached. Referral to MIL HDBK 217C along with some estimates of complexity suggest that the following lifetimes for the node might reasonably be expected.

Subscriber Interface	MBTF 80,000 Hr
Routing & Protocol Electronics	MBTF 40,000 Hr
Link	MBTF 120,000 Hr
Link, Routing & Protocol Combined	MBTF 30,000 Hr

The reasons for listing the lifetime for the combination of routing and protocol electronics plus one link will become apparent as we get deeper into the probabilities of losing communications to a node through no fault of the node itself.

Figure 6.1-2a shows a small network made of 3-port nodes. The circled node, which we will call an innocent node because it has not failed, can be isolated by failures of the three links attached to it or by failure of the protocol and routing electronics of the three next nearest nodes. Figure 3b shows a network with the same number of nodes which has been connected up in such an unfortunate manner that the failure of two links can cut the network in half. Obviously this type of network is to be avoided. Not quite so obvious is the fact

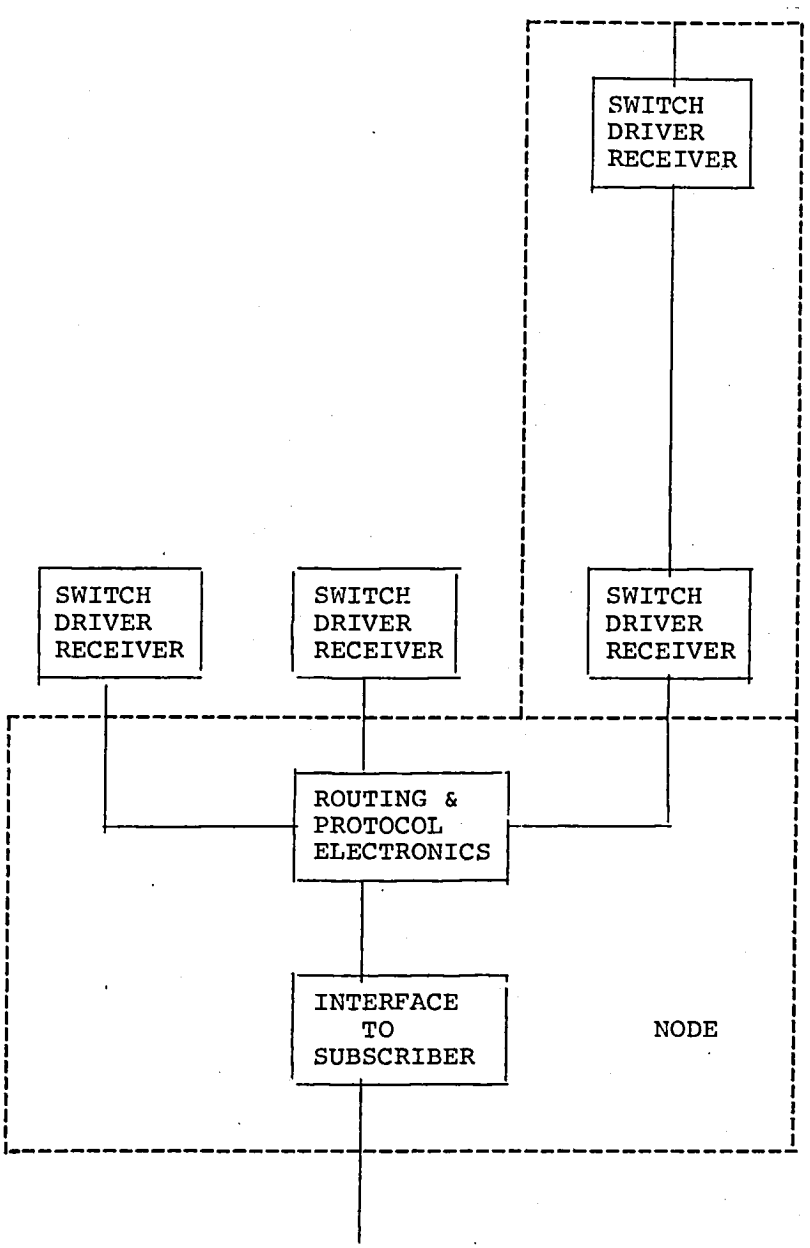
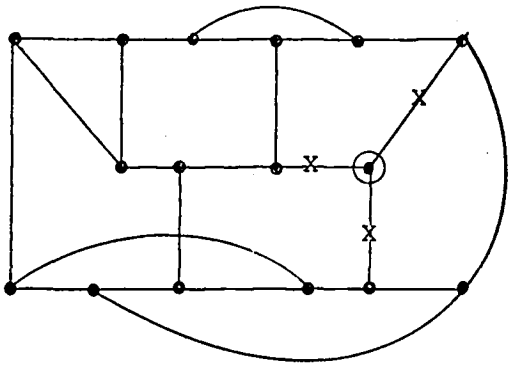
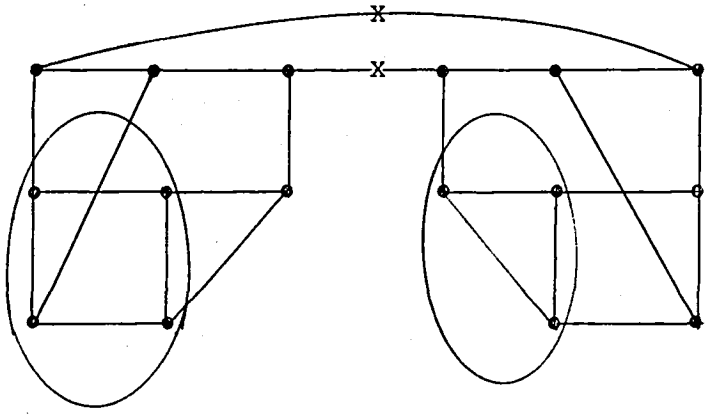


Figure 6.1-1. Schematic of Data Communication Node and Link.



A



B

Figure 6.1-2

Node Isolation In Irregular Networks

that both networks a and b have been cut into two pieces. The seriousness of the cut can be measured in terms of the number of nodes isolated. In a, it is one, and in b, eight. Further examination of a shows that three nodes can be isolated by three link breaks and four nodes can be isolated by four link breaks. If this network were hooked up in a regular fashion as in Figure 6.1-3 it would require four link breaks minimum to isolate two nodes, five link breaks minimum to isolate three nodes and six link breaks minimum to isolate four nodes, which obviously is a more reliable situation. We can conclude that regular networks minimize the chances for node isolation.

Figure 6.1-3b shows a network boundary and the fact that the irregularity caused by the boundary again creates a situation where three nodes can be isolated by three link failures. If a network could be designed without boundaries this problem would also disappear. Such a network that is easily to visualize is a cube with nodes at the corners and links along the edges. Also 20 3-port nodes can be placed at the corners of a dodecahedron with the same effect. We have termed such networks regular and closed.

In general, networks of either 3-port or 4-port nodes can be made regular and closed by interconnecting the nodes as shown in Figure 6.1-4. Networks connected in this way can be visualized as being evenly spread on the surface of a torus. This type of network has the desirable property that the relationship of one node to all its neighbors is the same as the relationship of any other node to all its neighbors. Such a network is much simpler to analyze, and the rest of this work will concern itself only with regular closed networks.

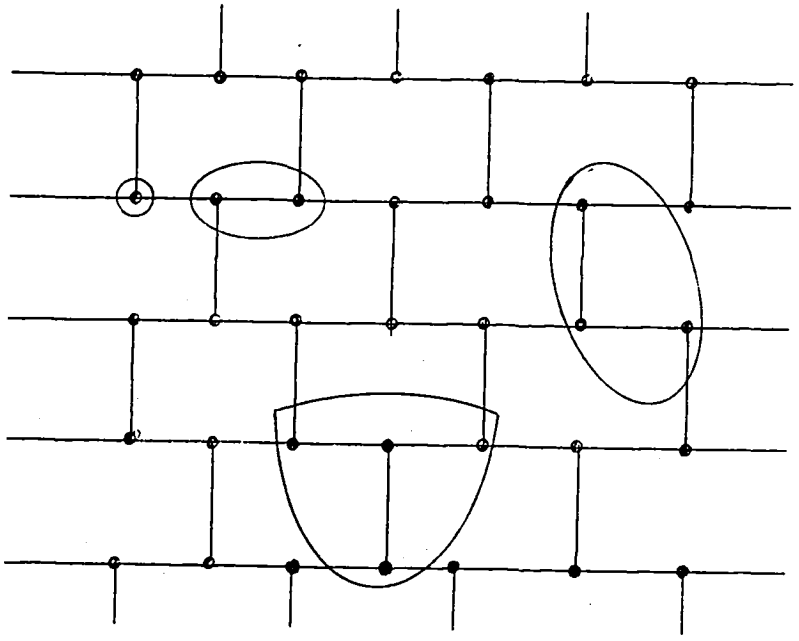
Probability of Isolating Innocent Nodes

If we have a network of 3-port nodes which is closed and regular and reasonably large, it is possible to derive the first terms of an expression for the probability of isolating an innocent node. If the probability of failure of a link or node is suitably small the higher order terms of this expression become negligible.

The probability that exactly x failures will occur out of a group of ℓ devices is approximately equal to

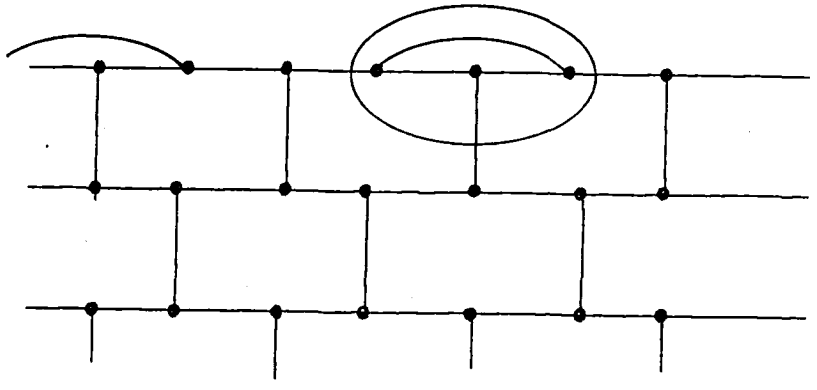
$$\frac{\ell!}{x!(\ell-x)!} (\lambda t)^x (1-\lambda t)^{\ell-x}*$$

*Here we have assumed that λt is .01 or less so that the reliability, $e^{-\lambda t}$, can be approximated by $1 - \lambda t$.



A

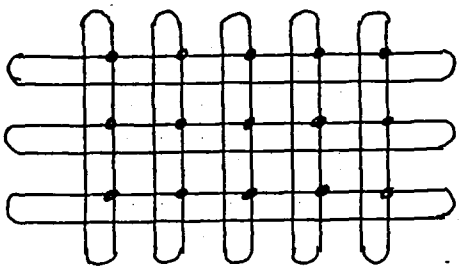
Unbounded Network



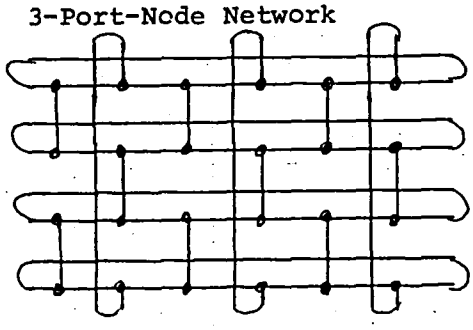
B

Bounded Network

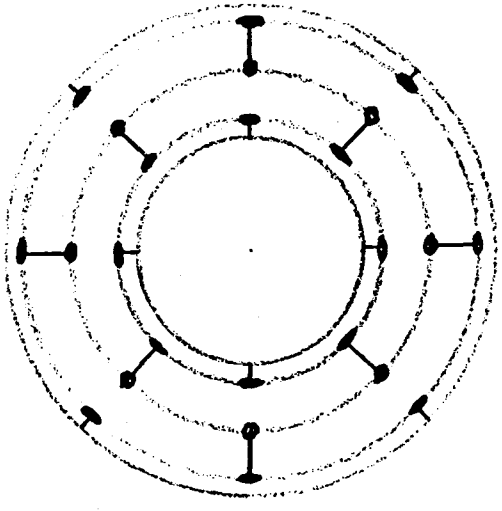
Figure 6.1-3. Large Regular Network.



3-Port-Node Network



4-Port-Node Network



3-Port-Node Network
Drawn on Torus
(Plan View)

Figure 6.1-4
Regular Closed Networks

where λ is the failure rate and t is the time interval of interest. If the number of failures of nodes or links in our network is 3 or more, an unfailed node may be isolated from the rest of the network. Figure 6.1-5 shows how three link breaks can isolate such a node. For a network of n nodes there are n ways in which a node can be isolated by three link breaks. The probability of having exactly three link failures in a network of n nodes (ℓ links where $\ell = 3/2\ n$) is:

$$\frac{\ell!}{3!(\ell-3)!}\ (\lambda t)^3\ (1-\lambda t)^{\ell-3}$$

Since we know that there are n sets of exactly three failures which can isolate an innocent node, then the fraction of all sets of three failures which can isolate an innocent node is

$$\frac{n}{\frac{\ell!}{3!(\ell-3)!}}$$

If we multiply the probability for having exactly three failures times the fraction of three failure sets which will isolate an innocent node, we get the probability that three failures will isolate an innocent node.

$$P_{I_3} = n\ (\lambda t)^3\ (1-\lambda t)^{(\ell-3)}$$

It is instructive to look at this problem a little differently. If we focus on a particular innocent node, then the chance of having a random link failure help to isolate the node is $3/\ell$. A second failure must strike one of the two remaining good links to the node under question which has a $2/\ell$-1 chance of happening. The third failure must strike the remaining good link which will happen with $1/\ell$-2 probability. Then the chance of having exactly three random failures isolate the innocent node under question is

$$\frac{3}{\ell} \cdot \frac{2}{\ell-1} \cdot \frac{1}{\ell-2} = \frac{3!(\ell-3)!}{\ell!}$$

The chance of having three failures isolate any of the n innocent nodes is

$$n\ \frac{3!(\ell-3)!}{\ell!}$$

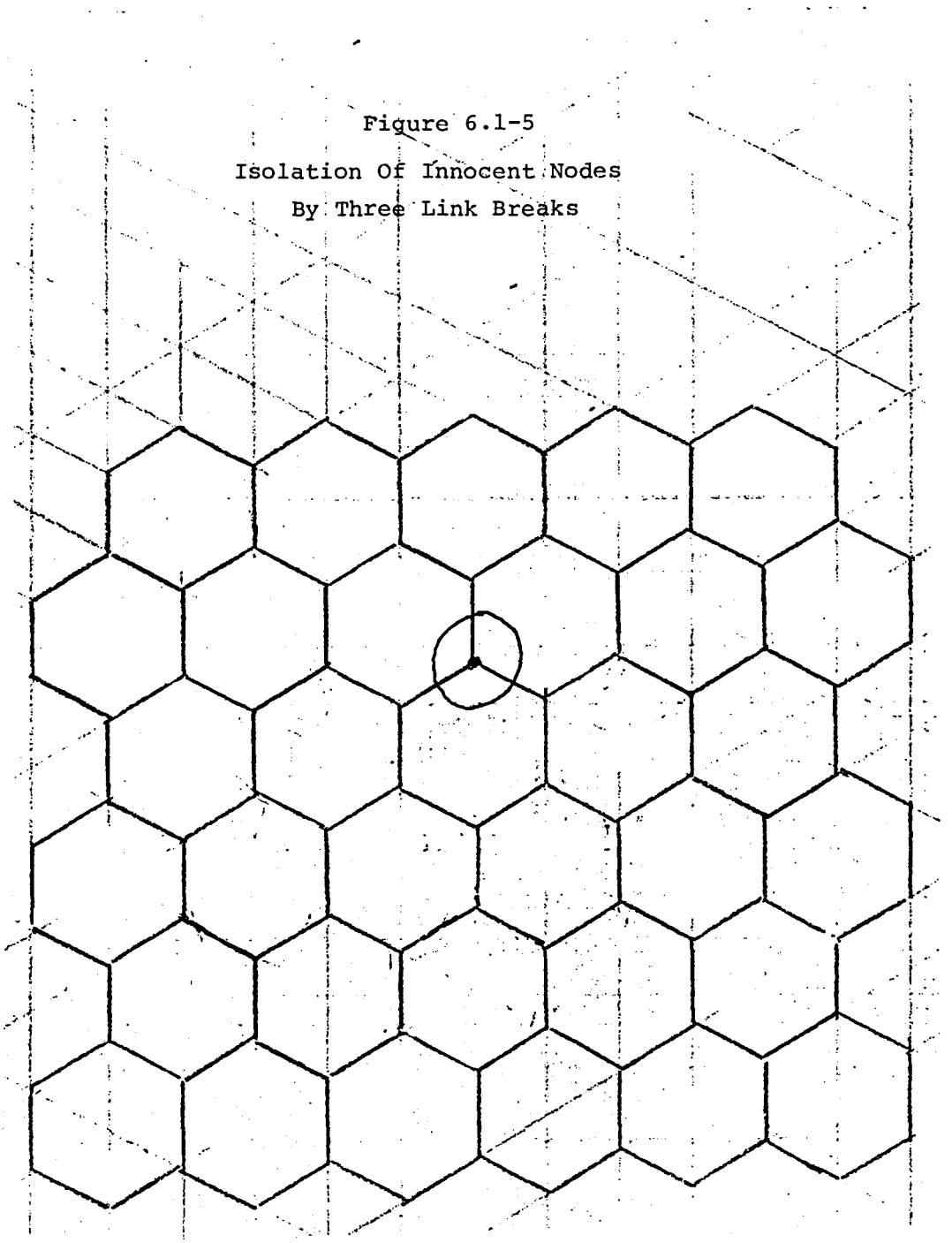


Figure 6.1-5

Isolation Of Innocent Nodes

By Three Link Breaks

The chance of having exactly three failures is

$$\frac{\ell!}{3!(\ell-3)!}\ (\lambda t)^3\ (1-\lambda t)^{(\ell-3)}$$

The product of these last two expressions is the chance of having exactly three failures isolate an innocent node

$$P_{I_3} = n\ (\lambda t)^3\ (1-\lambda t)^{(\ell-3)}$$

which agrees with the previous analysis.

Now if there are four failures, two nodes can be isolated as shown in Figure 6.1-6. Also, there is the possibility of isolating a single node with three breaks as in the previous analysis and having the fourth break appear harmlessly elsewhere in the network. The probability of having one or more nodes isolated with exactly four link breaks is

$$P_{I_4} = \left(\frac{\ell}{\frac{\ell!}{4!(\ell-4)!}}\right) \frac{\ell!}{4!(\ell-4)!}\ (\lambda t)^4\ (1-\lambda t)^{\ell-4}$$

$$+ n\ \left(\frac{\ell}{\frac{\ell!}{4!(\ell-4)!}}\right) \frac{\ell!}{4!(\ell-4)!}\ (\lambda t)^4\ (1-\lambda t)^{\ell-4}$$

The rationale for the first term in the equation is as follows. There are ℓ ways of isolating an innocent node using all four link breaks. To see this, note that each such isolation event is uniquely representable by a closed curve containing one entire link.

Simplifying the above expression, we have

$$P_{I_4} = n\left(\ell - \frac{3}{2}\right)\ (\lambda t)^4\ (1-\lambda t)^{\ell-4}$$

Therefore with five link breaks there are several possibilities. Three of the breaks can isolate one node. Four breaks can isolate two nodes. Five breaks can isolate three nodes or they can isolate two nodes as shown in Figure 6.1-7. (Figure 6.1-8 shows the six-break case for interest.)

The number of ways the three-break isolation can occur is found by focusing on a single node isolated by three link breaks. The two other link breaks occur harmlessly elsewhere and they can occur in

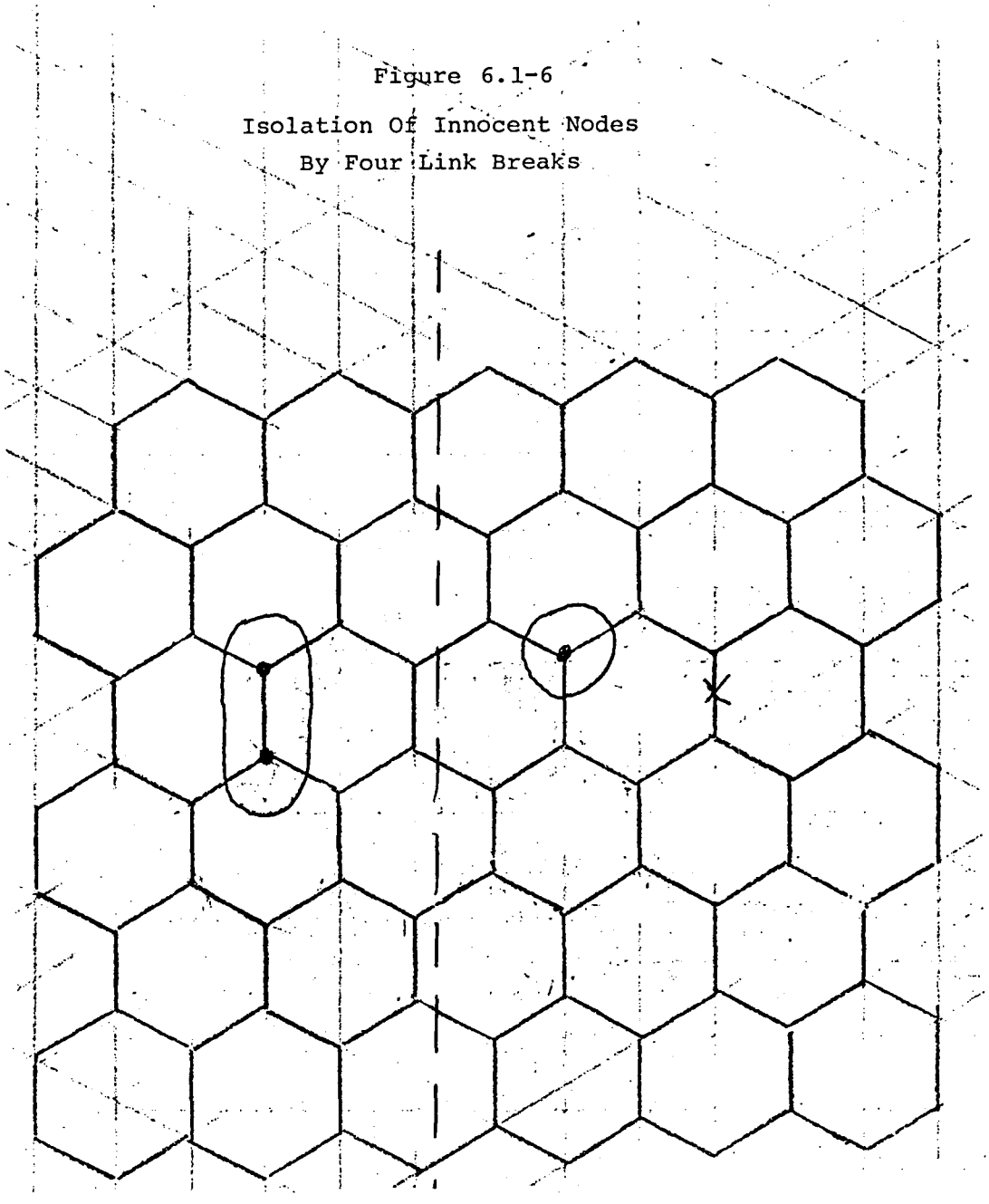


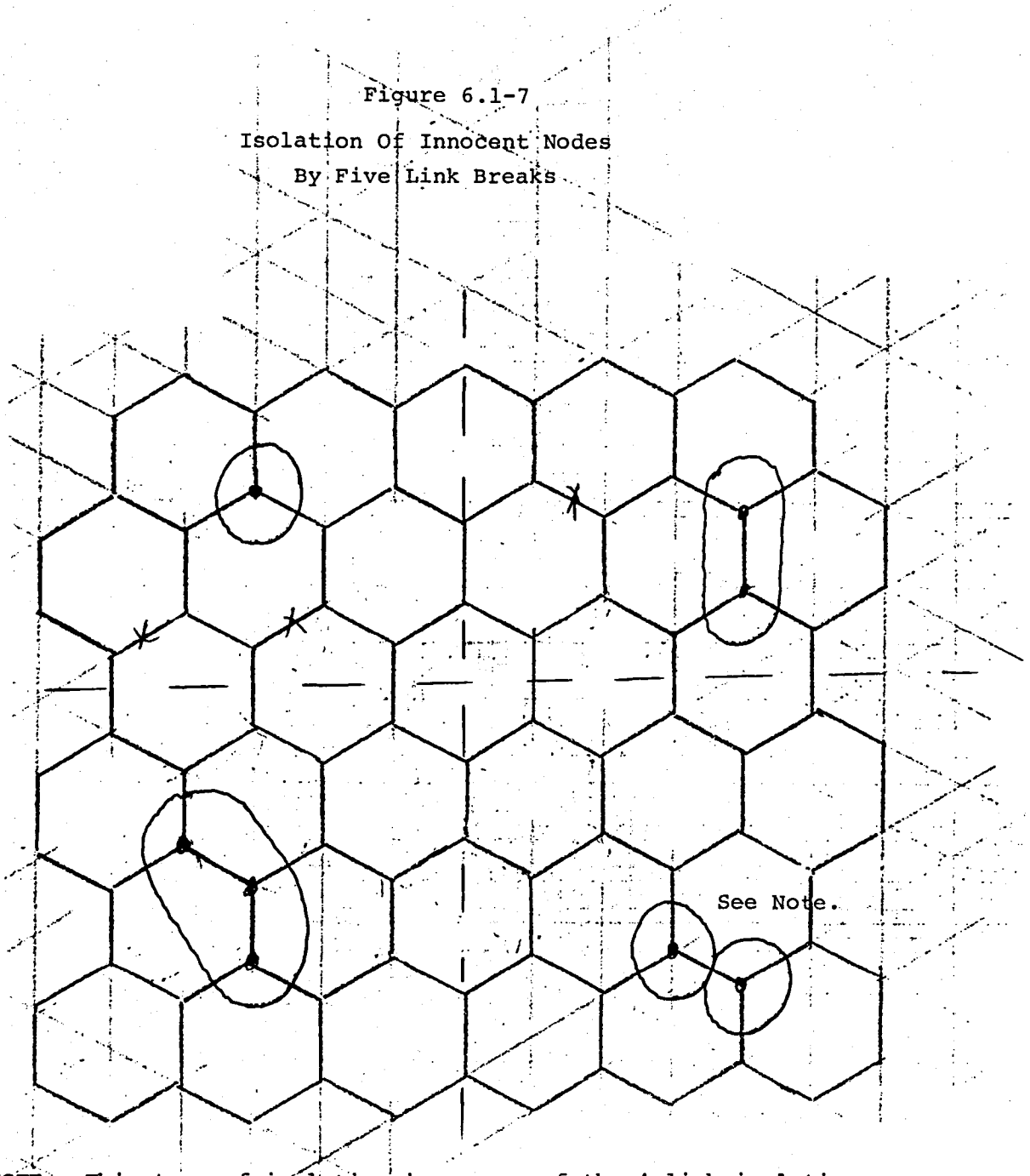
Figure 6.1-6

Isolation Of Innocent Nodes

By Four Link Breaks

Figure 6.1-7

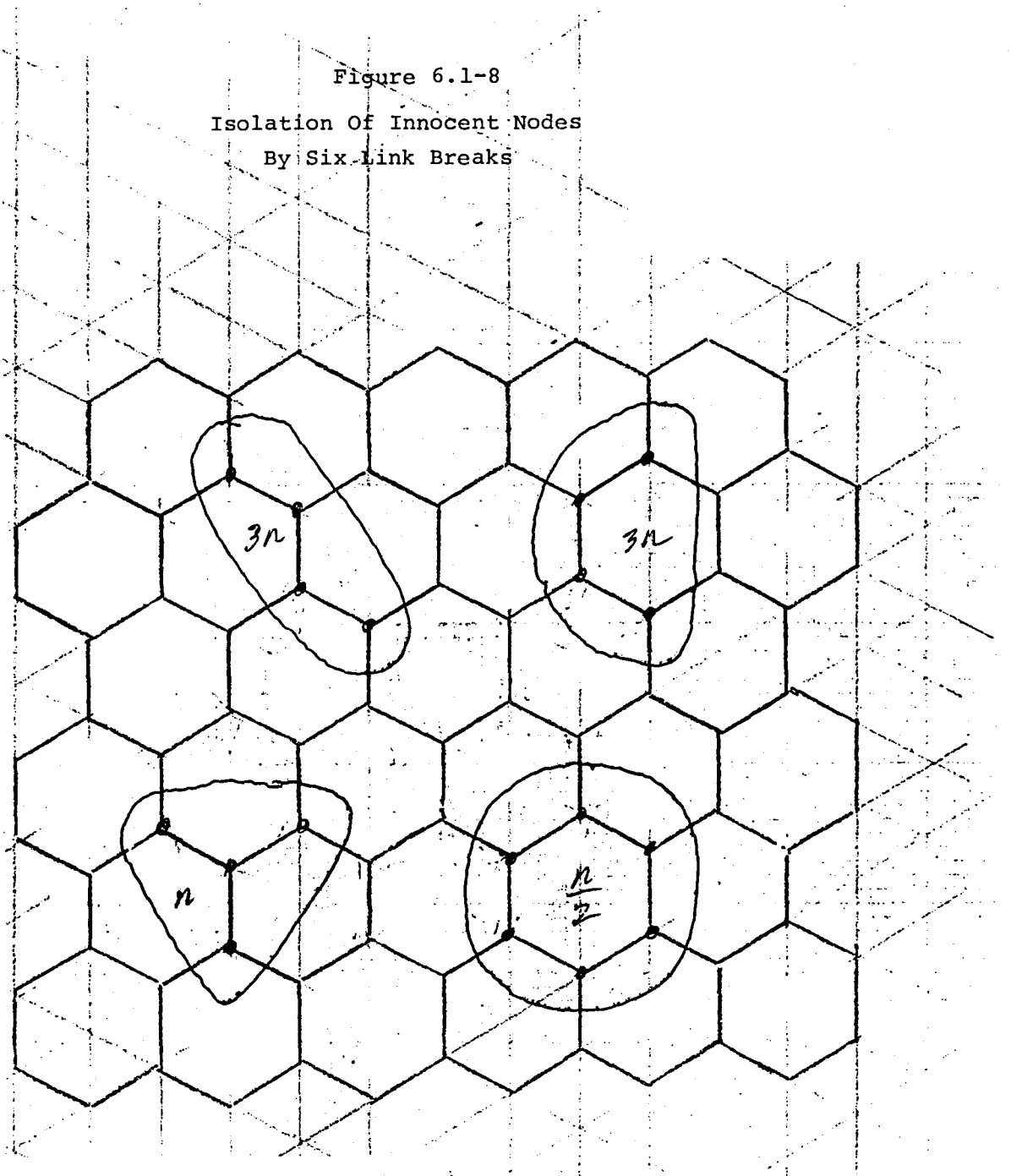
Isolation Of Innocent Nodes
By Five Link Breaks



NOTE: This type of isolation is a case of the 4 link isolation plus a random break. It is also a case of a 3 link isolation plus two random breaks. Therefore it should not be counted separately as a different kind of break pattern. In fact it should either be counted among the 3-link isolation patterns or among the 4-link isolation patterns, but not both. There are ℓ such patterns and if subtracted from the 4-link isolation case would give a corrected number of 4-link isolation patterns of $\ell(\ell-4)-\ell$. This effect is small for networks of 20 or more nodes, and not including it errs on the conservative side. It therefore will be neglected in this analysis.

Figure 6.1-8

Isolation Of Innocent Nodes
By Six Link Breaks



Plus all the possible combinations shown for 3, 4, and 5 breaks.

$$\frac{(\ell-3)!}{2!(\ell-5)!}$$

ways among the remaining $\ell-3$ links. The total number of such patterns which can cause isolation is then the number of nodes, n, times the number of combinations of two link breaks out of the remaining good links.

$$\frac{n}{2}\ (\ell-3)(\ell-4).$$

The number of ways the four break isolation can occur is found by multiplying the number of links ℓ, or $\frac{3}{2}$ n, times the number of ways a single break can occur out of the remaining $(\ell-4)$ links,

$$\frac{3}{2}\ n\ (\ell-4).$$

The number of ways the five break isolation can occur is a little more difficult to visualize. One way is to realize that there are nine patterns that can isolate a given node along with two others. This suggests that there are a total of 3n combinations for this type of isolation. It is perhaps easier to look at Figure 6.1-7 and note that the cut pattern can be stepped one node at a time in three possible directions which gives rise to the 3n term.

At this point it should be noted that although we have been assuming link failures, the same equations would result if we had assumed node failures for the three-and-four-failure cases. In both cases the failure rate would be that for the node routing electronics plus one whole link including the driver & receiver electronics at both ends. For the five failure case which isolates two adjacent nodes and also breaks the link between the nodes, (see Figure 6.1-7), the interchangeability of node and link failures begins to break down.

The probability of isolating one or more innocent nodes is

$$P_{I_5} = n\ [\frac{(\ell-3)(\ell-4)}{2} + \frac{3}{2}\ (\ell-4) + 3]\ (\lambda t)^5\ (1-\lambda t)^{\ell-5}$$

$$= n\ [\frac{\ell^2}{2} - 2\ell + 3]\ (\lambda t)^5\ (1 - \lambda t)^{\ell-5}$$

The total probability for isolation by 0 through 5 failures is:

$$P_{I_{0-5}} = n\ [1+(\ell- \frac{3}{2})\lambda t\ (1-\lambda t)^{-1} + (\frac{\ell^2}{2} - 2\ell + 3)\ (\lambda t)^2\ (1-\lambda t)^{-2}]$$

$$(\lambda t)^3\ (1-\lambda t)^{\ell-5}$$

For $\ell > 100$ and $\lambda t > 10^{-3}$, this can be simplified to

$$P_{I_{0-5}} = [1 + \ell\lambda t + \frac{1}{2} (\ell\lambda t)^2] \; n \; (\lambda t)^3 \; (1-\lambda t)^{\ell-3}$$

It is apparent that if $\ell\lambda t$ is substantially less than 1, that the terms of this series will become small rapidly, and that the series has a limit because there are fewer than ℓ terms. It would therefore appear that rules of thumb can be developed to relate network size, node failure rates, flight hours between repairs and requirements for not isolating an innocent node. For example we might say that

$$\ell\lambda t < \frac{1}{10}$$

and

$$P_I < 2 \times 10^{-7}$$

This would suggest that λt be kept less than 10^{-3} for ℓ of about 100, which represents reasonable numbers for a system which commences operation with no failures.

For a regular closed network made up of 4-port nodes as shown in Figure 6.1-9, the minimum number of failures which can isolate an innocent node is four. The probability of this occurring is

$$P_{I_4} = n \; (\lambda t)^4 \; (1-\lambda t)^{(\ell-4)}$$

For five failures - Figure 6.1-10

$$P_{I_5} = n \; (\ell-4) \; (\lambda t)^5 \; (1-\lambda t)^{(\ell-5)}$$

For six failures - Figure 6.1-11

$$P_{I_6} = [2n + \frac{n}{2} (\ell-4)(\ell-5)] \; (\lambda t)^6 \; (1-\lambda t)^{\ell-6}$$

Therefore the probability of up to six failures isolating one or two nodes is

$$P_{I_{0-6}} = [1 + (\ell-4)\lambda t(1-\lambda t)^{-1} + (\frac{\ell^2}{2} - \frac{9\ell}{2} + 12)(\lambda t)^2(1-\lambda t)^{-2}] \; n \; (\lambda t)^4 \; (1-\lambda t)^{\ell-4}$$

which simplifies to

$$P_{I_{0-6}} = [1 + \ell\lambda t + \frac{1}{2} (\ell\lambda t)^2] \; n \; (\lambda t)^4 \; (1-\lambda t)^{\ell-4}$$

Figure 6.1-9

Isolation Of An Innocent Node By
Four Link Breaks

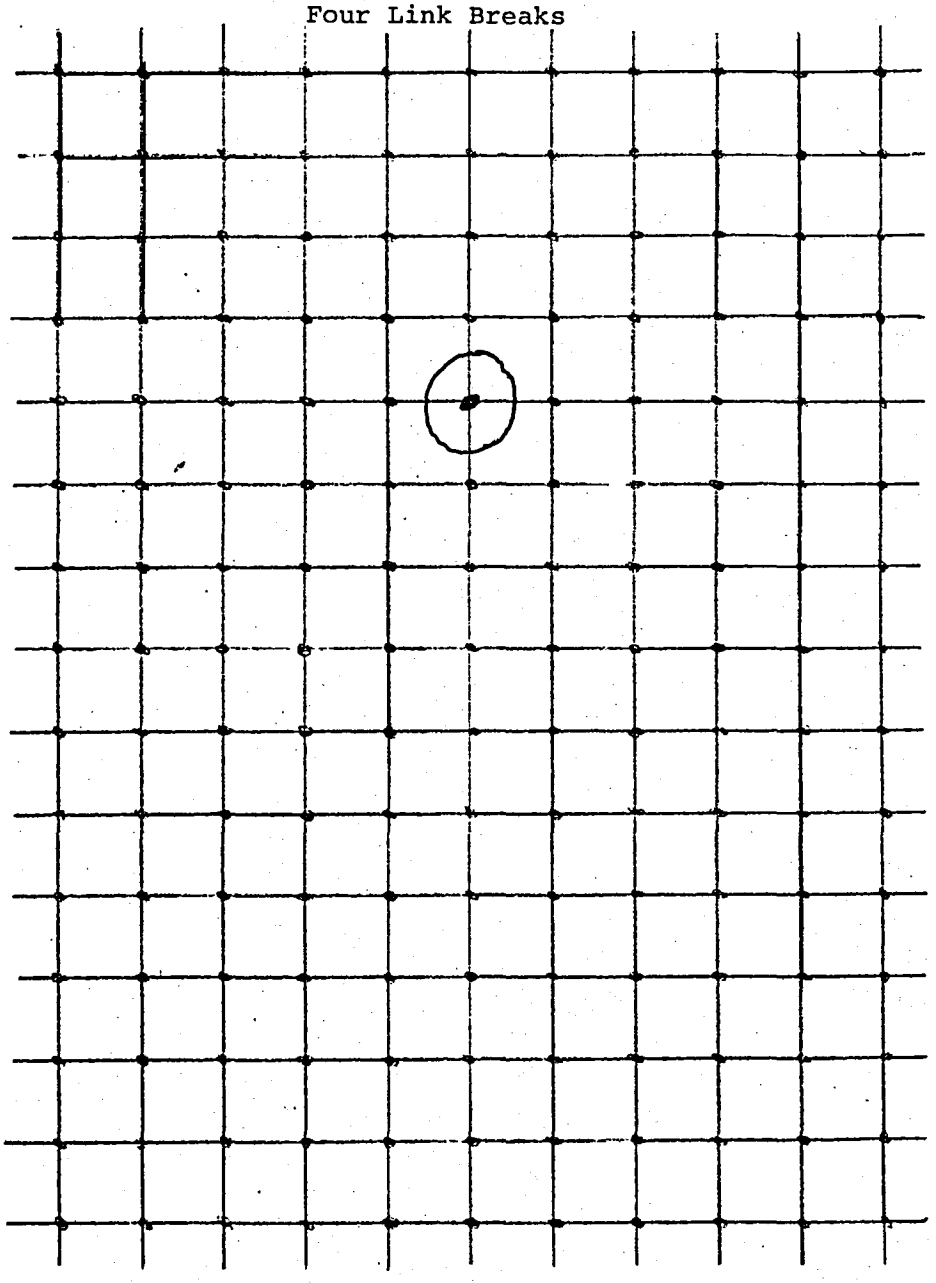


Figure 6.1-10

Isolation Of An Innocent Node By
Five Link Breaks

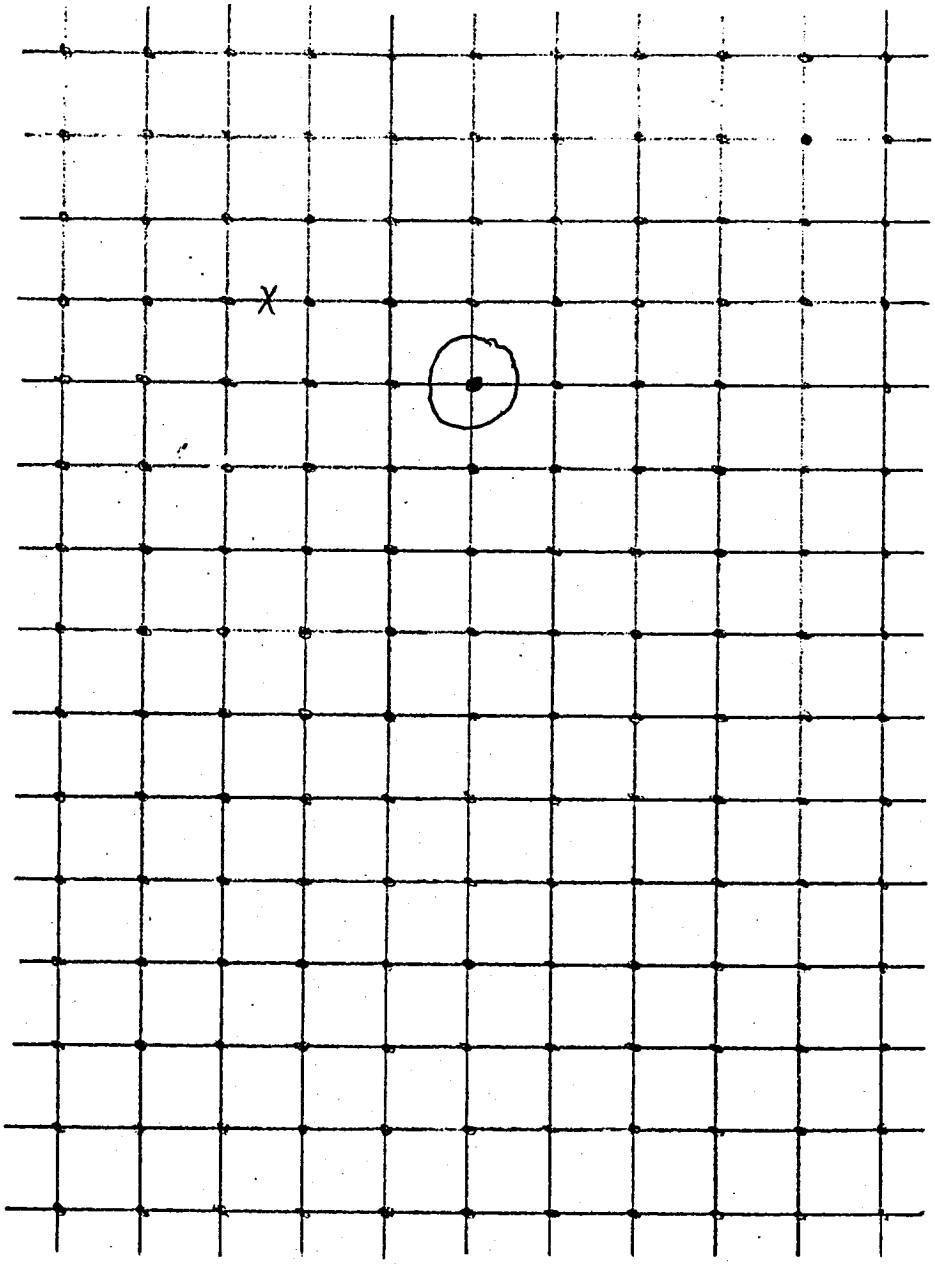
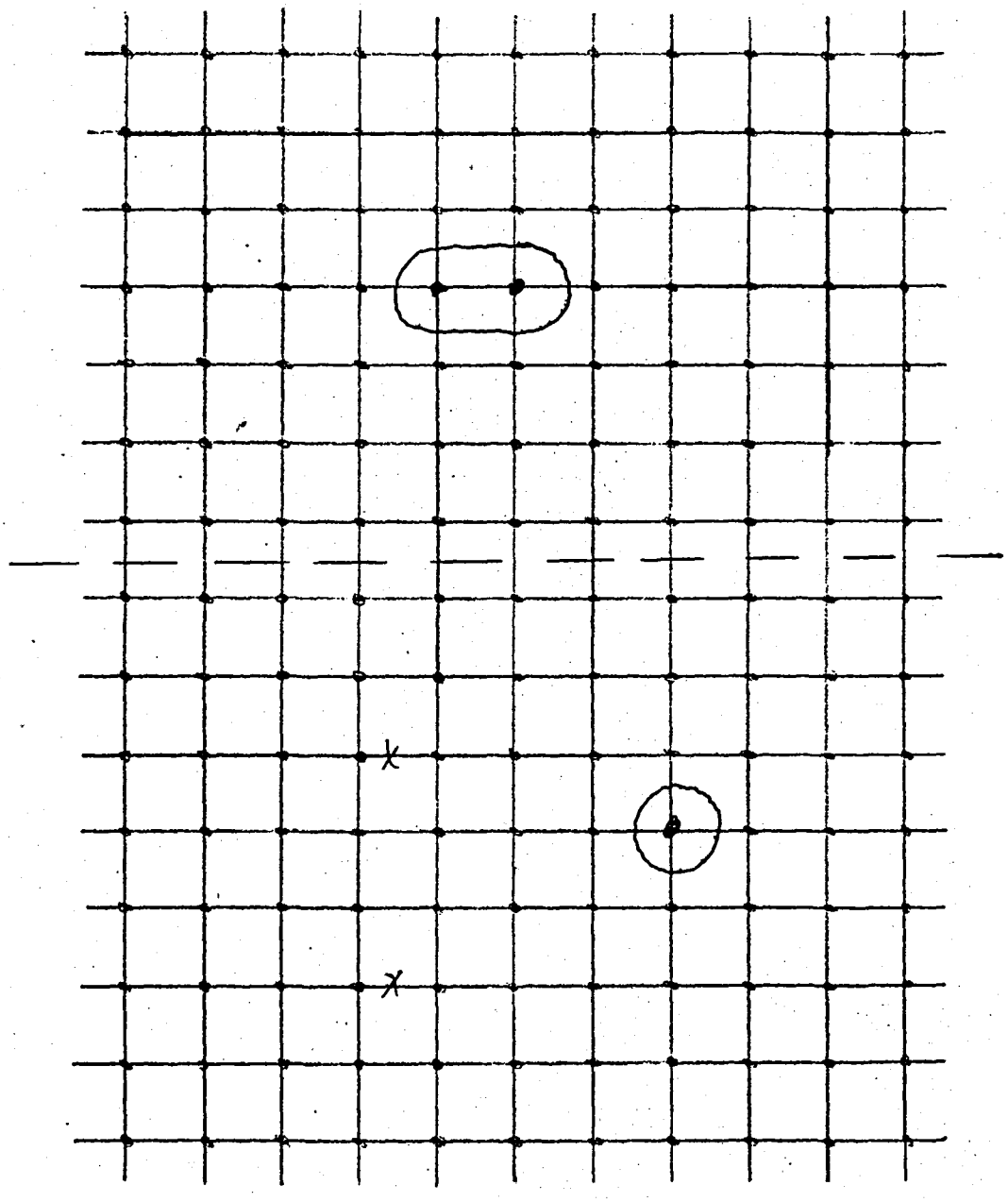


Figure 6.1-11

Isolation Of An Innocent Node By
Six Link Breaks



If this is compared with the equation for 3-port nodes it is apparent that the probability of isolating an innocent node is lower by a factor of almost λt when 4-port nodes are used instead of 3-port nodes.

A practical question at this juncture is, what is the probability of isolating an innocent node if one or more nodes have already failed. For the 3-port node case, if there is a single link failure then the chance of two more link failures joining with the single link failure to isolate a node is

$$P_{I_{1+2}} = 2\ (\lambda t)^2$$

The chance of isolating a node in a 4-port node network after one link failure with three more link failures is

$$P_{I_{1+3}} = 2\ (\lambda t)^3$$

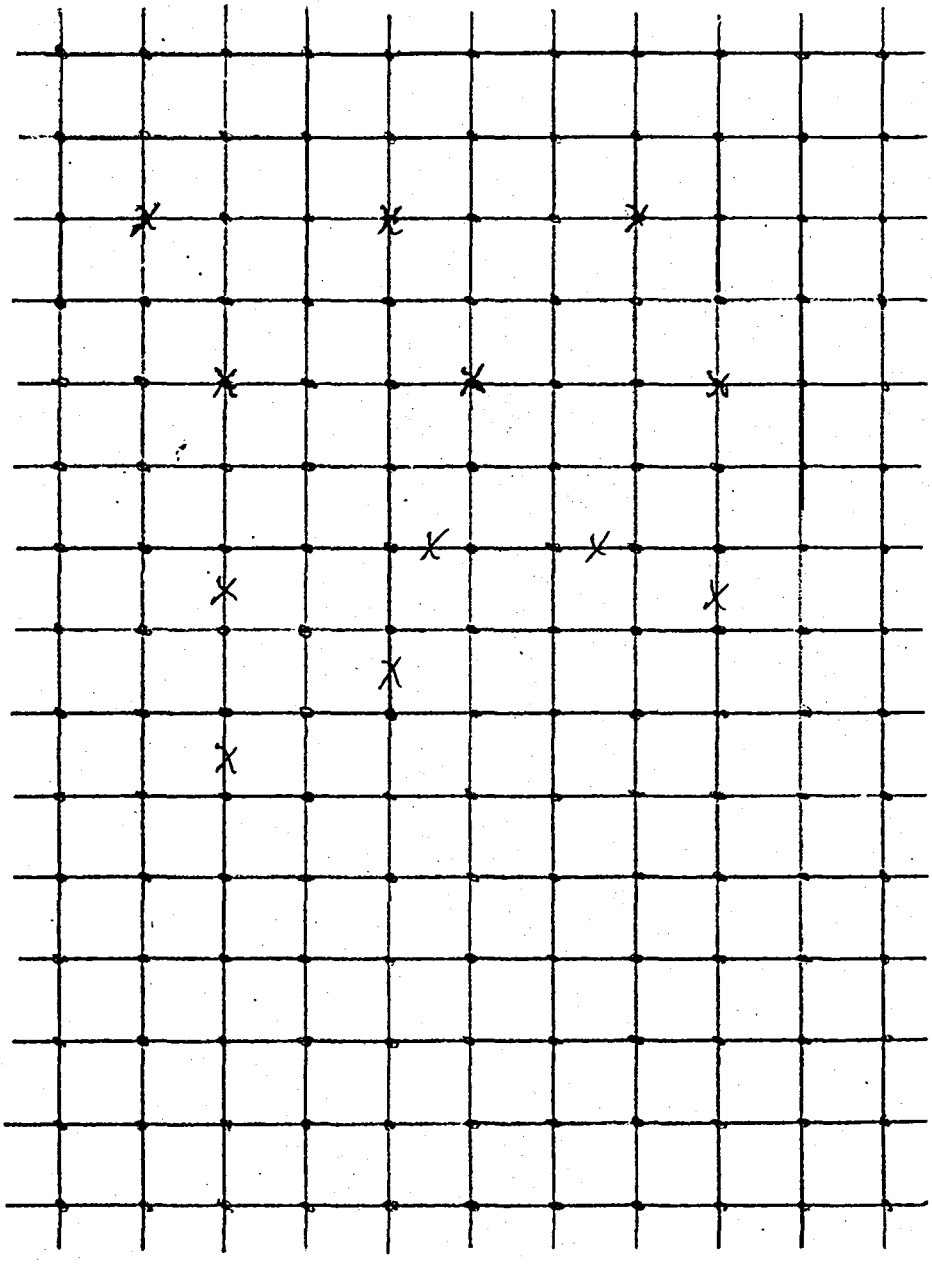
It is obvious that the chance of isolating an innocent node in a network which contains a failure is significantly lower in a network made of 4-port nodes.

To argue this question further we must ask what the effects of two or more failures are. Obviously if these failures are too close to one another in the network the chance of isolating an innocent node increases greatly. For example if two links to an innocent node are broken, only one more needs to be broken in a 3-port node network to isolate an innocent node. The chances for isolation become much less if a rule is adopted which limits the closeness of failures to remain unrepaired in a flightworthy system.

Figure 6.1-12 shows a 4-port node network. If we do not permit node failures to be closer to another node failure than the third node away, then at least three more failures are required to isolate an innocent node. Likewise two failed links must have at least one good link between them (except of course for the correlated link failures which occur when a node fails). The same rules are good for a 3-port node network except that now only two more failures may isolate an innocent node. This is limited by the number of ports on the node and cannot be improved by increasing the spacing between failures which may be tolerated in a flightworthy system.

Figure 6.1-12

Minimum Spacing For Node And
Link Failures



Additional Network Reliability Equations

Here we will develop some approximate expressions for comparing the reliability of regular closed networks made from 3-port and 4-port nodes. The equations are only approximate for the following reasons.

1. Reliability, $e^{-\lambda t}$, is approximated by $1 - \lambda t$. This is a very good approximation for $\lambda t < .01$.

2. The chance of isolating an innocent node or nodes is based on the approximate number of link break patterns which will cause isolation relative to the exact number of link break patterns of x broken links out of ℓ that are possible. The number of link break patterns which will cause isolation is approximate because of double counting of patterns in some instances. See Figure 6.1-7. Also it is possible for two isolation patterns to occur in a single network. Offsetting these approximations is the fact that the number of broken links in one isolation pattern has been limited to six in a 3-port node network and eight in a 4-port node network.

3. The assumption that link and node failures are equivalent breaks down for large numbers of node failures.

It is possible to bound the errors to some degree by different methods of calculation. It is also possible to look at double counting and overlapping patterns directly for the simpler cases and assess their effect. A Monte Carlo test can also be made by computer to provide further verification.

There are some interesting questions that can be asked about network failures which will be addressed here

1. What is the probability that a particular innocent node will be isolated, given x link breaks in a network of ℓ links and n nodes?

2. What is the probability that one or more innocent nodes will be isolated given x link breaks in the network?

3. What is the likely number of link breaks which will cause one or more innocent nodes to be isolated in a network?

4. What is the probability that x link breaks will occur in a network and isolate one or more innocent nodes?

We will call the probability that x link breaks will isolate a particular innocent node P_{I_x}. Tables 6.1-1 and 6.1-2 list the ratios of the number of possible isolating patterns of x breaks around the object node to the total number of patterns possible with x breaks out of ℓ links.

In Table 6.1-1, the coefficients 1,3,9 and 31 are the numbers of link break isolation patterns possible around a particular node for 3, 4, 5, and 6 link breaks. The coefficients 1, 4, and 18 were found the same way for Table 6.1-2. Figure 6.1-13 shows the possible 8-link break patterns around a given node. The chance of isolating a given node is

$$P_{I_x} = \frac{x!}{(x-3)!}\ \frac{(\ell-3)}{\ell!} + 3\ \frac{\ell!}{(x-4)!}\ \frac{(\ell-4)!}{\ell!} + 9\ \frac{\ell x!}{(x-5)!}\ \frac{(\ell-5)!}{\ell!}$$

$$+ 31\ \frac{x!}{(x-6!)}\ \frac{(\ell-6)!}{\ell!}$$

3-port nodes

$$P_{I_x} = \frac{x!}{(x-4)!}\ \frac{(\ell-4)!}{\ell!} + 4\ \frac{x!}{(x-6)!}\ \frac{(\ell-6)!}{\ell!} + 18\ \frac{x!}{(x-8)!}\ \frac{(\ell-8)!}{\ell!}$$

4-port nodes

The chance of isolating some node in the network is just n times the above expressions. It is noted that these expressions are good for small x but for x larger than $\ell/5$ for the 3-port node equation and $\ell/4$ for the 4-port node equation, the errors become significant. Another way of looking at the problem is to consider the probability of not isolating a particular node

$$1 - P_{I_x}$$

The probability that no innocent nodes in the network are isolated equals

$$P_{ONI} = (1 - P_{I_x})^n$$

This expression has the advantage that it provides reasonable answers for larger values of x. The probability that one node is isolated is

$$P_{1NI} = \frac{n!}{(n-1)!}\ P_{I_x}\ (1-P_{I_x})^{n-1}$$

and the probability that y nodes are isolated is

$$P_{YNI} = \frac{n!}{(n-y)!}\ P_{I_x}{}^{y}\ (1-P_{I_x})^{n-y}$$

TABLE 6.1-1

PROBABILITY OF GETTING SPECIFIC ISOLATION PATTERNS AROUND A PARTICULAR NODE IN A 3-PORT NODE NETWORK

NUMBER OF BROKEN LINKS	3-LINK BREAK ISOLATION PATTERN	4-LINK BREAK ISOLATION PATTERN	5-LINK BREAK ISOLATION PATTERN	6 LINK BREAK ISOLATION PATTERN
3	$\frac{1}{\frac{\ell!}{3!(\ell-3)!}}$			
4	$\frac{(\ell-3)}{\frac{\ell!}{4!(\ell-4)!}}$	$\frac{3}{\frac{\ell!}{4!(\ell-4)!}}$		
5	$\frac{\frac{(\ell-3)(\ell-4)}{2!}}{\frac{\ell!}{5!(\ell-5)!}}$	$\frac{3(\ell-4)}{\frac{\ell!}{5!(\ell-5)!}}$	$\frac{9}{\frac{\ell!}{5!(\ell-5)!}}$	
6	etc.	$\frac{\frac{3(\ell-4)(\ell-5)}{2!}}{\frac{\ell!}{6!(\ell-6)!}}$	$\frac{9(\ell-5)}{\frac{\ell!}{6!(\ell-6)!}}$	$\frac{31}{\frac{\ell!}{6!(\ell-6)!}}$
7		etc.	$\frac{\frac{9(\ell-5)(\ell-6)}{2!}}{\frac{\ell!}{7!(\ell-7)!}}$	$\frac{31(\ell-6)}{\frac{\ell!}{7!(\ell-7)!}}$
8			etc.	$\frac{\frac{31(\ell-6)(\ell-7)}{2!}}{\frac{\ell!}{8!(\ell-8)!}}$
x	$\frac{x!}{(x-3)!}\frac{(\ell-3)!}{\ell!}$	$\frac{3\,x!}{(x-4)!}\frac{(\ell-4)!}{\ell!}$	$\frac{9\,x!}{(x-5)!}\frac{(\ell-5)!}{\ell!}$	$\frac{31\,x!}{(x-6)!}\frac{(\ell-6)!}{\ell!}$

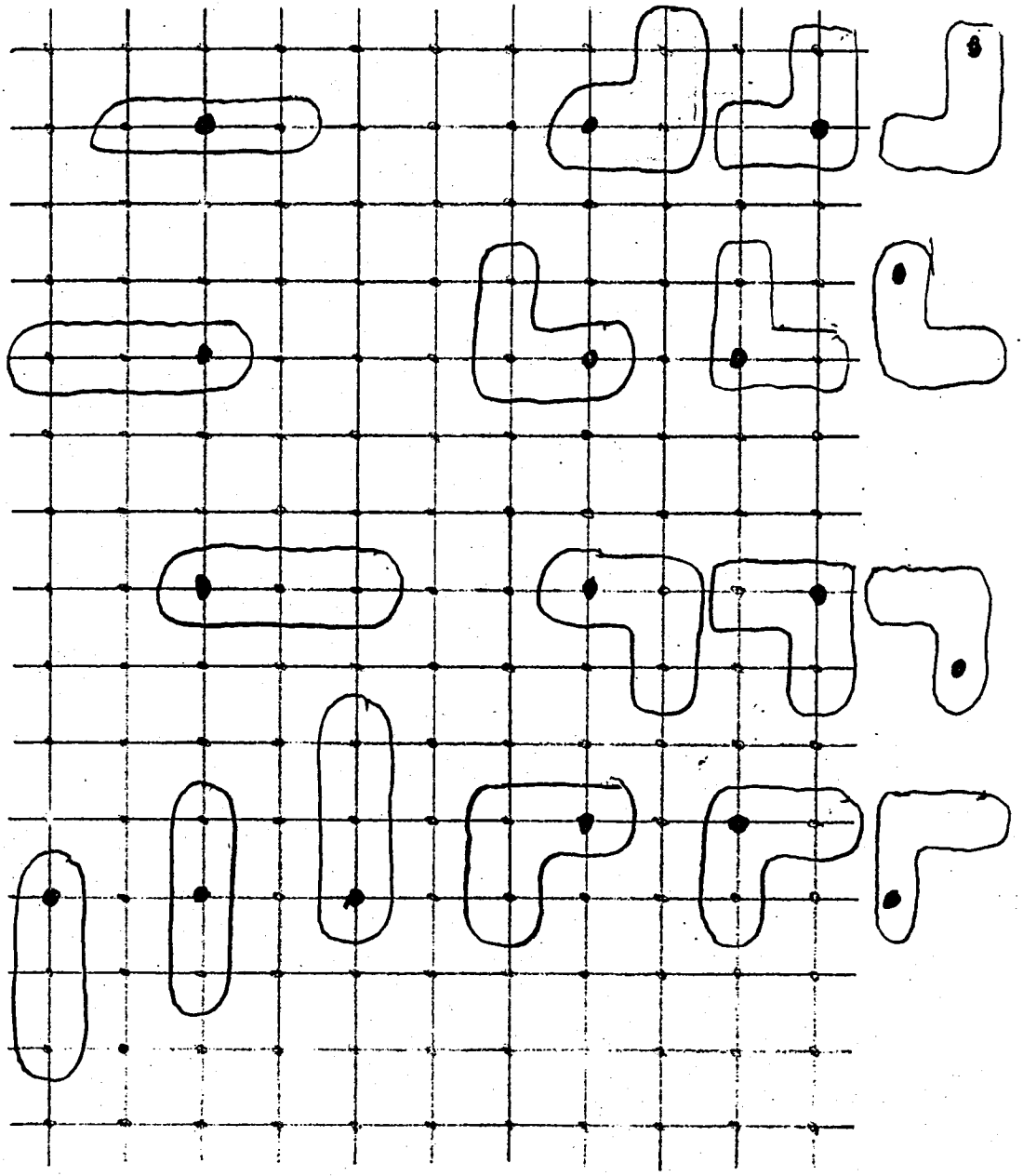
TABLE 6.1-2

PROBABILITY OF GETTING SPECIFIC ISOLATION PATTERNS AROUND A PARTICULAR NODE IN A 4-PORT NODE NETWORK

NUMBER OF BROKEN LINKS	4-LINK BREAK ISOLATION PATTERN	6-LINK BREAK ISOLATION PATTERN	8-LINK BREAK ISOLATION PATTERN
3	$\frac{1}{\frac{\ell!}{4!\ (\ell-4)!}}$		
4	$\frac{(\ell-4)}{\frac{\ell!}{5!\ (\ell-5)!}}$		
6	$\frac{\frac{(\ell-4)\ (\ell-5)}{2!}}{\frac{\ell!}{6!\ (\ell-6)!}}$	$\frac{4}{\frac{\ell!}{6!\ (\ell-6)!}}$	
7	etc.	$\frac{4\ (\ell-6)}{\frac{\ell!}{7!\ (\ell-7)!}}$	
8		$\frac{\frac{4\ (\ell-6)\ (\ell-7)}{2!}}{\frac{\ell!}{8!\ (\ell-8)!}}$	$\frac{18}{\frac{\ell!}{8!\ (\ell-8)!}}$
x	$\frac{x!}{(x-4)!}\ \frac{(\ell-4)!}{\ell!}$	$\frac{4\ x!}{(x-6)!}\ \frac{(\ell-6)!}{\ell!}$	$\frac{18\ x!}{(x-8)!}\ \frac{(\ell-8)!}{\ell!}$

Figure 6.1-13

All Purpose 8-Link Break Patterns
Which Will Isolate a Given Node



The probable number of link breaks to isolate a node in a network can be found by setting P_{ONI} equal to 0.5 and solving for the number of link breaks.

Some feeling for the likely number of link breaks to cause isolation can be gained by realizing that a network can remain in one piece under ideal circumstances with one fewer link than there are nodes. The next link break will cause isolation. Also we know that no isolation will occur in a 3-port node network with only two link breaks. Therefore the likely number of link breaks to cause isolation will be more than 2 and less than $\frac{n}{2} + 2$ for a 3-port node network and more than 3 and less than n + 2 for a 4-port node network.

A third way of looking at the problem of the likely number of link breaks that will isolate an innocent node is to calculate the probabilities that various types of break batterns will not occur in the probability that there are no isolating patterns. For example the probability of having a particular 3-link isolation pattern appear in a 3-port node network is

$$\frac{x!}{(x-3)!} \quad \frac{(\ell-3)!}{\ell!}$$

There are n possible 3-link isolation patterns in the network so the chance of not having any patterns of this type is

$$(1 - \frac{x!}{(x-3)!} \quad \frac{(\ell-3)!}{\ell!})^n$$

The chance of having a 3-link isolation pattern is

$$1 - (1 - \frac{x!}{(x-3)!} \quad \frac{(\ell-3)!}{\ell!})^n$$

Similar terms can be written for 4, 5 and 6 break isolation patterns. If these terms were independent they could be multiplied times each other to give the probability that none of these isolation patterns appear. Actually if one of the isolation patterns does appear, it is considerably less likely that a second pattern will appear in the same network. However, we will add these terms, to be conservative, and to tend to cancel the error introduced by truncating the size of the break pattern considered. The expressions for the isolation of one or more nodes are then

$$P_{I_{\substack{1\text{-}6 \text{ nodes} \\ 3 \text{ ports}}}} = 4 - [1 - \frac{\ell x!}{(x-3)!} \quad \frac{(\ell-3)!}{\ell!}]^n - [1 - \frac{x!}{(x-4)!} \quad \frac{(-4)!}{!}] \quad \frac{3n}{2}$$

$$- \left[1 - \frac{x!}{(x-5)!}\,\frac{(\ell-5)!}{\ell!}\right]^{3n} - \left[1 - \frac{x!}{(x-6)!}\,\frac{(\ell-6)!}{\ell!}\right]^{\frac{15n}{2}}$$

$$P_{\substack{I \\ 1\text{-}3 \text{ nodes} \\ 4 \text{ ports}}} = 3 - \left[1 - \frac{x!}{(x-4)!}\,\frac{(\ell-4)!}{\ell!}\right]^{n} - \left[1 - \frac{x!}{(x-6)!}\,\frac{(\ell-6)!}{\ell!}\right]^{2n}$$

$$- \left[1 - \frac{x!}{(x-8)!}\,\frac{(\ell-8)!}{\ell!}\right]^{6n}$$

Figures 6.1-14 and 6.1-15 show plots of probabilities for innocent node isolation as a function of link failures for regular closed 36 node networks made with 3-port and 4-port nodes. In addition to plots of the three equations derived for isolation of innocent nodes, simulation results are shown, and a straight line plot is made between the known end point for no isolation and for certain isolation.

It is quite obvious from the plots for innocent node isolation that the fourth port on a node does a great deal for improving the reliability of the network from the point of view of innocent node isolation.

The probability that x link breaks will occur and isolate a particular innocent node is the probability that x link breaks will cause isolation times the probability that x link breaks will occur

$$P_{I_x} \cdot P_x = P_{I_x}\,\frac{\ell!}{x!(\ell-x)!}\,(\lambda t)^x\,(1-\lambda t)^{\ell-x}$$

Any of the isolation probability expressions can be used for P_{I_x} in the above equation as long as x is not too large.

Figures 6.1-14 and 6.1-15 include results of a Monte Carlo simulation based on the flowchart shown in Figure 6.1-16. The flowchart assumes that failures can be assigned to subscribers, nodes, and links. For Figures 6.1-14 and 6.1-15, however, only link failures were allowed. To illustrate the significance of link failures vs. node failures, Figure 6.1-17 shows a comparison of probability densities for a three-port network of 36 nodes. The node curve rises more steeply, because each node failure eliminates three links. However, since each node failure eliminates a potential innocent victim, the node curve approaches unity gradually, whereas the link curve stops suddenly when less than one link for each node remains.

Figure 6.1-14. Probability for Innocent Node Isolation for a 36 3-Port Node Network.

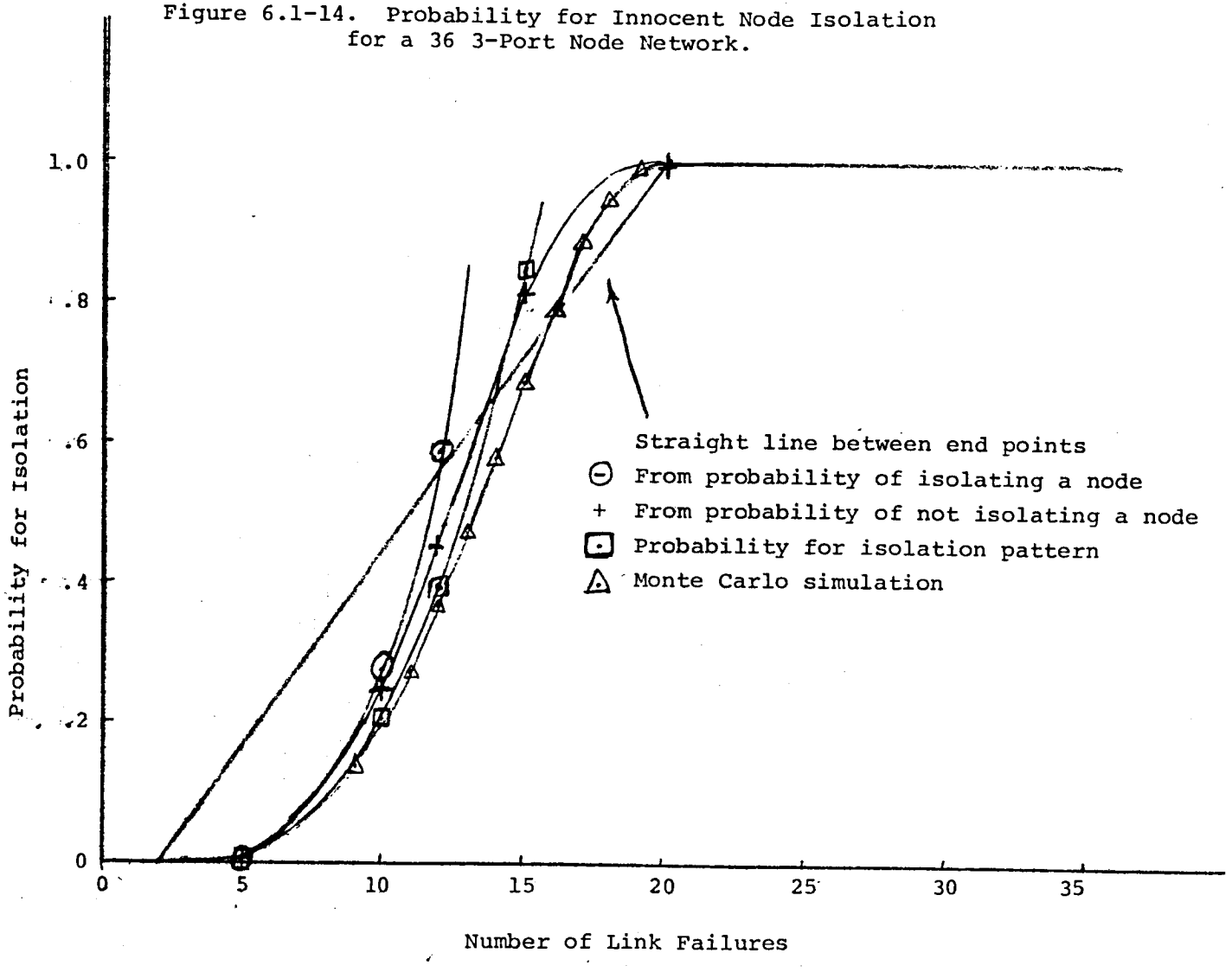
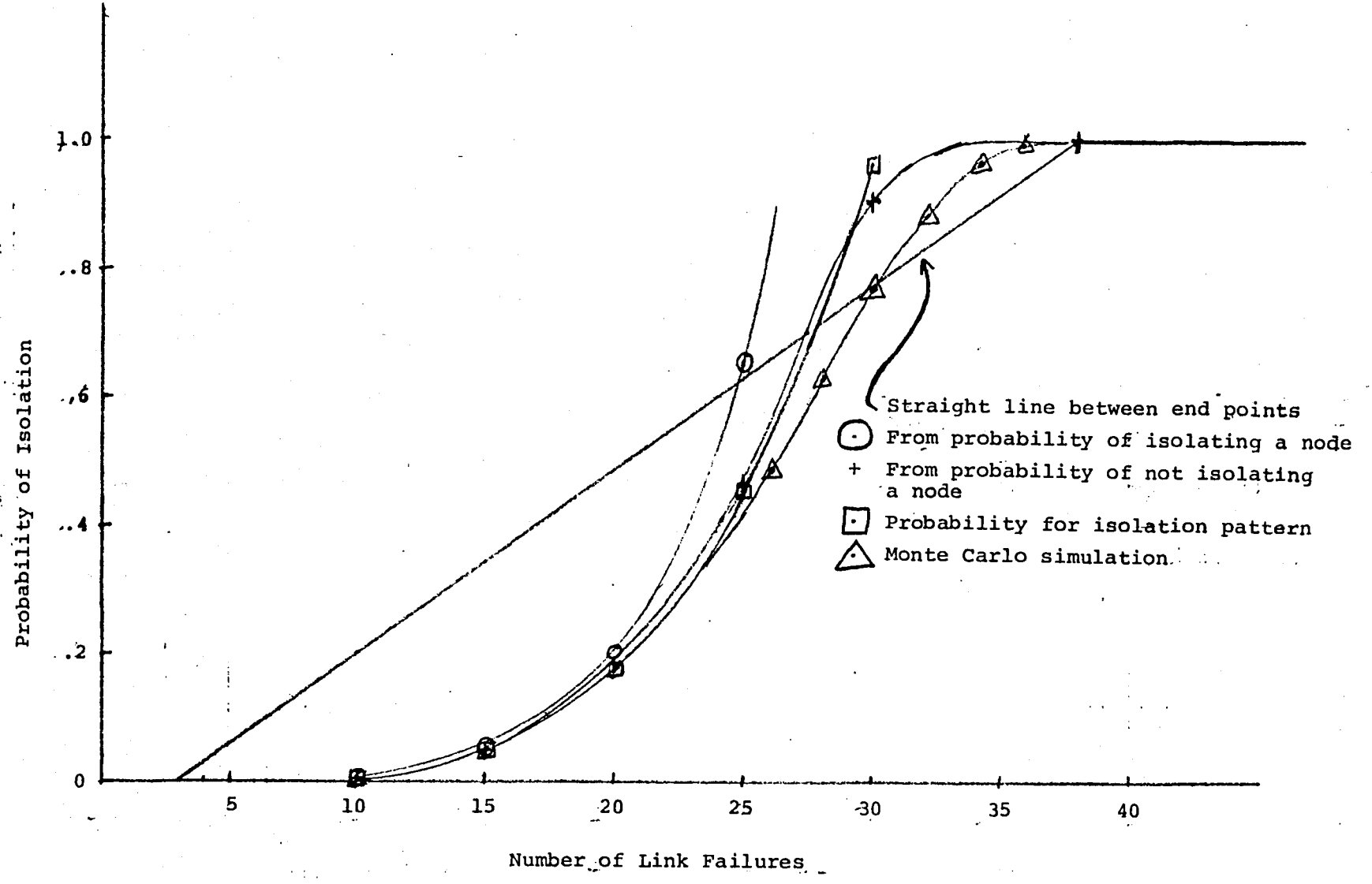


Figure 6.1-15. Probability for Innocent Node Isolation for a 36 4-Port Node Network.



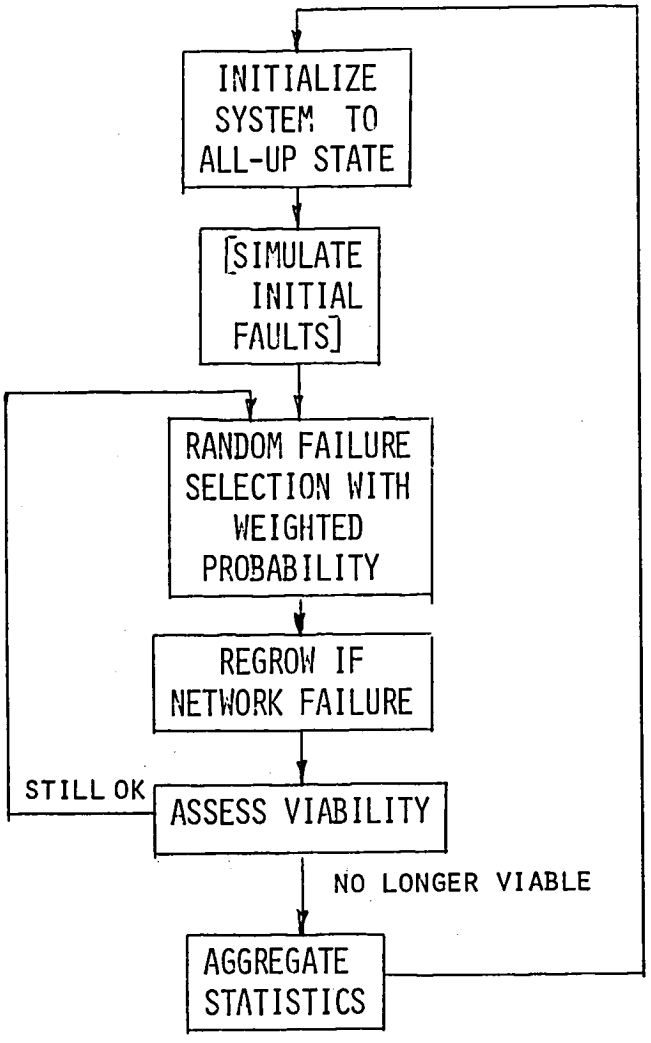


Figure 6.1-16. Failure Simulation Algorithm.

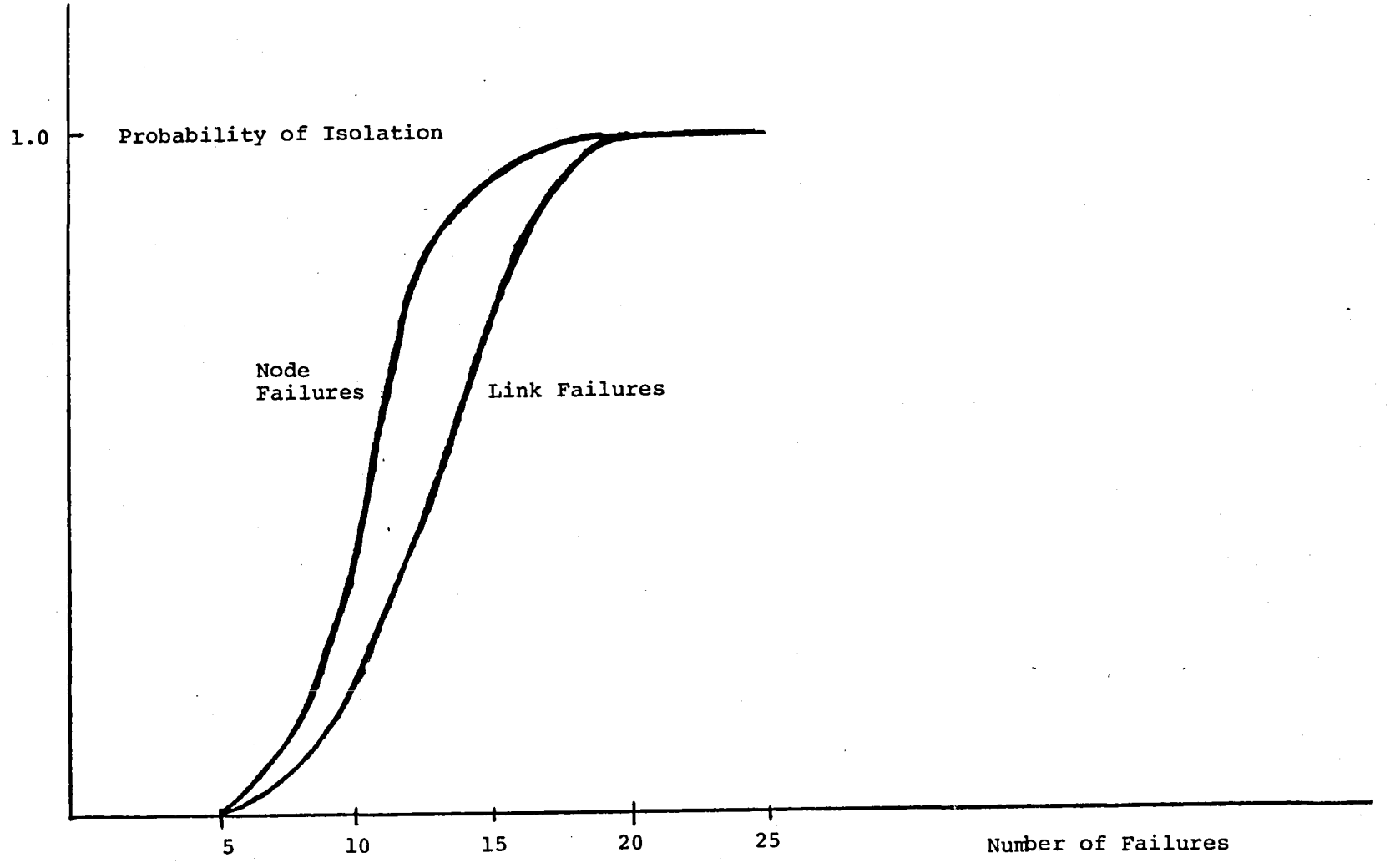


Figure 6.1-17. Link Vs. Node Failures.

6.2 Network Dispatch Probability

To approximate the probability that network faults will prevent dispatch, we must first identify the dispatch criteria. For a simple criterion, such as "dispatch is allowed unless x or more faults exist," the non-dispatch probability is simply the probability that x or more faults exist. If only nodes could fail, for example, and the node failure rates were all equal to λ_n, then the leading term in the probability expression for nondispatch would be

$$P(x) = \frac{N!\ (\lambda_n t)^x\ (1-\lambda_n t)^{N-x}}{x!\ (N-x)!} \qquad (6.2\text{-}1)$$

where t is the elapsed time of concern, such as an operational day, and N is the number of nodes. If $\lambda_n t$ is small, this term is a sufficiently close approximation to the desired value. Moreover, the factor

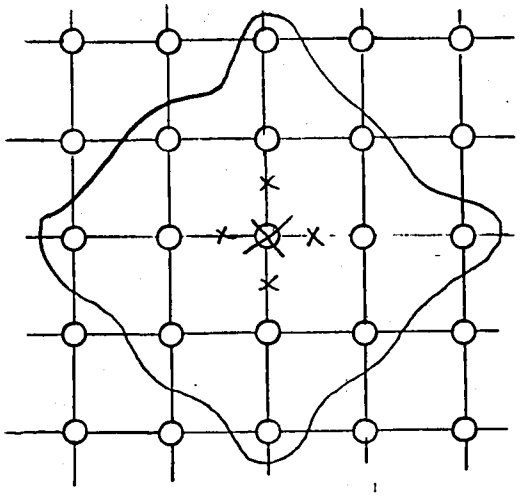
$$(1 - \lambda_n t)^{N-x}$$

may be set equal to unity. Thus if one fault is permitted (x = 2), and a hundred node network operates twenty hours (N = 100, t = 20) and the failure rate per node is 10^{-4} per hour, then the non-dispatch probability is approximately

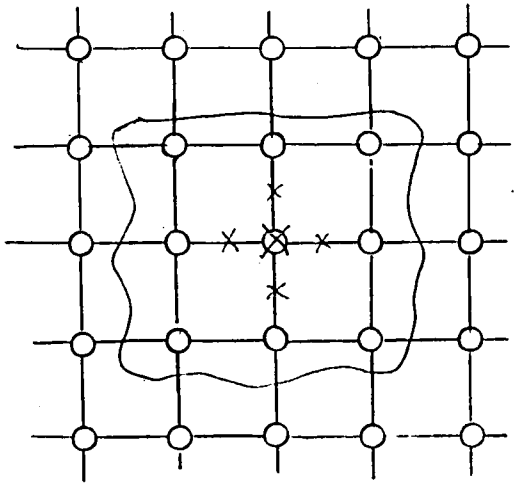
$$P(2) = \frac{100!\ (20 \times 10^{-4})^2}{2!\ 98!} \approx 0.02$$

Next, consider a more elaborate, more permissive criterion, where any number of faults may exist so long as they are mutually distant. (It must also be true that these faults do not violate the minimum equipment criterion for the subscribers, but this is not the present concern.) Let a buffer zone be defined surrounding each faulty node or link. The dispatch criterion may be stated as "no fault lies in the buffer zone of another fault." The size of the buffer zone may be chosen arbitrarily. Figure 6.2-1 shows three examples.

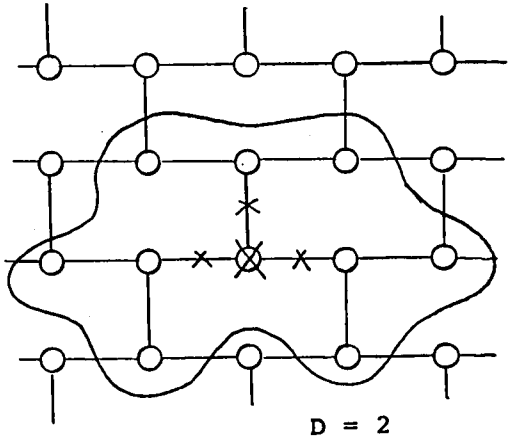
The probability of having exactly x node faults is given by eq. (6.2-1). When x equals zero or one, the system is dispatchable. When x = 2, there is a chance that the faults are too close. This is given approximately by B ÷ N-1, where B is the number of nodes in the buffer zone, and N is again the total number of nodes. N is assumed to be large, and link failures improbable for this example. If there are two faults, then, the conditional probability that the system is



D = 3
B = 12



D = 2
B = 8



D = 2
B = 9

Buffer zones around failed nodes
D = Minimum linkage for dispatch
B = Number of nodes in zone

Figure 6.2-1. Buffer Zones.

dispatchable is $1 - \frac{B}{N-1}$. If a third fault arrives, it must clear two buffer zones, and the conditional probability for dispatch becomes $(1 - \frac{B}{N-1})(1 - \frac{2B}{N-2})$. For succeeding faults, this conditional probability expression is of the form

$$P_C(x) = (1 - \frac{B}{N-1})(1 - \frac{2B}{N-2})(1 - \frac{3B}{N-3}) \ \ldots \ (1 - \frac{(x-1)B}{N-x+1}),$$

truncated and set equal to zero if negative. The dispatch probability expression is as follows.

$$P_D = P(o) + P(1) + P(2)\ (1 - \frac{B}{N-1}) + P(3)\ (1 - \frac{B}{N-1})\ (1 - \frac{2B}{N-2})$$

$$+ \ \ldots \ P(x)\ P_C(x) + \ \ldots$$

Figure 6.2-2 shows representative values of $1 - P_D$ as functions of N and $\lambda_n t$ for the three-port node buffer zone in Figure 6.2-1. If $\lambda_n = 10^{-4}$, $t = 20$, and $N = 100$, the probability of nondispatch, $1 - P_D$, is about 2×10^{-3}.

6.3 Network Connectivity

As aircraft communication and power distribution systems become more complex, and the level of flight criticality of such systems demands higher and higher reliability performance, designers will be faced with the complex problem of determining whether or not a proposed design provides adequate connectivity. In an effort to provide several alternative pathways of communication and power delivery to critical systems, the designer must be assured that all elements in the system which require multiple pathways for extreme reliability are covered in the final design. For systems with very large numbers of elements and communication links, the problem of determining the level of connectivity is very complex. The problem becomes even more severe when a design is characterized as an irregular network.

As has been described in the earlier sections of this report, there are levels of criticality of particular elements in the total system structure. Some elements are flight critical and require the highest reliability performance. Other subsystems are not as critical. Thus, designers are very likely to create irregular networks of elements in which certain parts of the network require more pathways of connection than others in order to ensure performance. The entire system may be made up of regions of regular networks which are interconnected in an irregular way to reach the necessary design objectives.

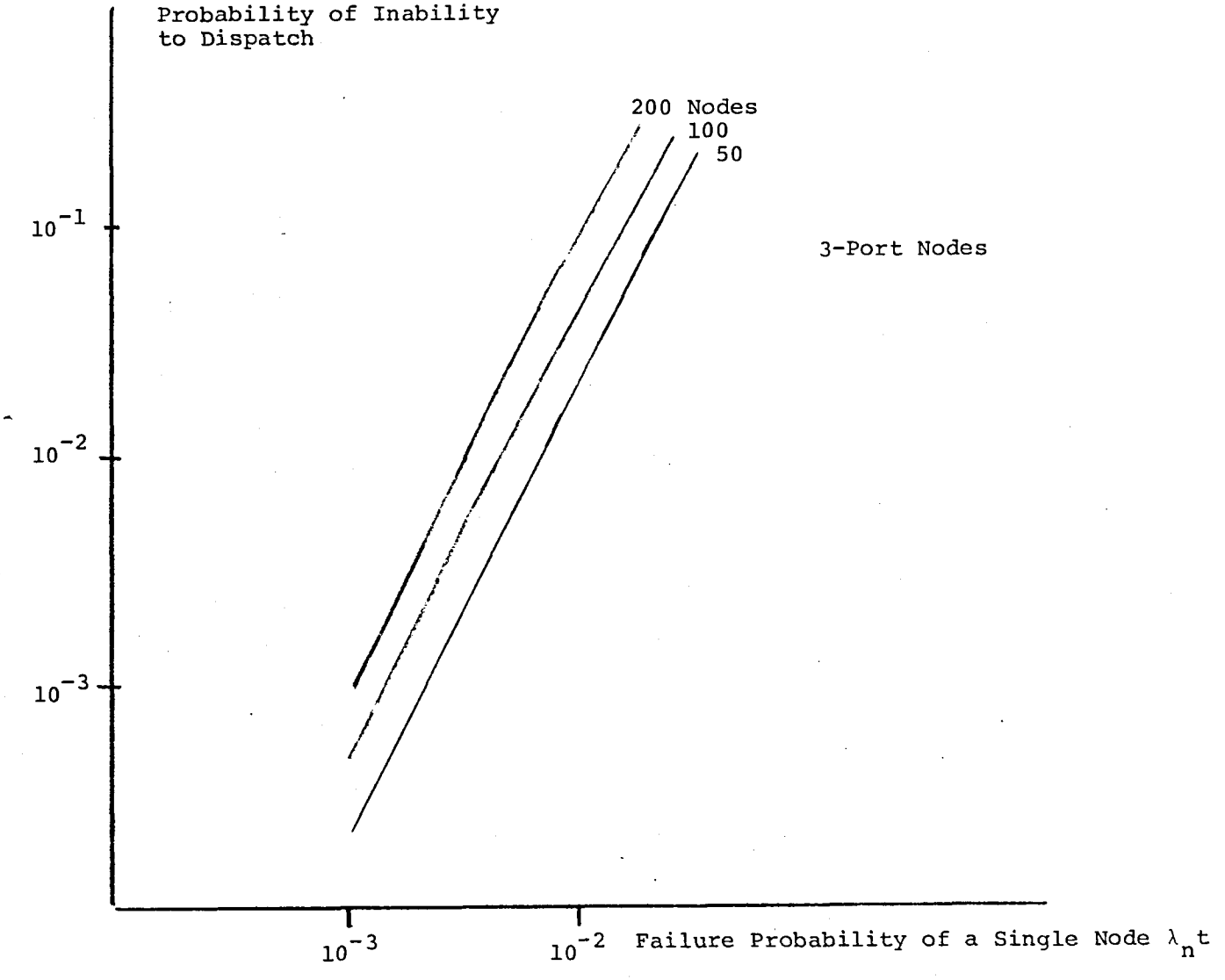


Figure 6.2-2. Plot of $1 - P_D$.

As part of our study of aircraft communication systems, a computer program has been developed which provides to the designer an analysis tool for determining the level of connectivity of large networks and to assist him in identifying regions of proposed designs which are weak from this standpoint and therefore require improvements. The remaining part of this section describes this particular effort.

6.3.1 Problem Description

The problem which needs to be solved is the following: What is the minimum number of failures in a network which will isolate an innocent node or nodes from a particular node? This special node represents the location of the fault-tolerant multiprocessor which is assumed to be an element in the total system design. The designer wants to ensure that this minimum number is large enough for his design. There are additional pieces of information which are also useful to the designer. For example, he may like to know what combination of failures caused the isolation, and which nodes were isolated. He may also have requirements, as alluded to earlier, that certain regions of the system have a higher minimum number of failures which can cause node isolation.

The algorithm described below provides an engineering solution to these questions.

6.3.2 Algorithm Description

There are several general approaches to the problem of determining connectivity, and these are well-documented in the literature. For large networks, all of the available analysis methods consume a great deal of computer time. Thus, we have attempted to take advantage of certain problem-dependent features to design a program which, most of the time, will run efficiently. This dependence on the particular network being analyzed means that there are networks which can be constructed which will take extremely large amounts of computer time for solution. But in the development of a design tool, we are not interested in program performance for worst cases, but wish to take advantage of features unique to this aircraft network problem.

Figure 6.3.2-1 is a flow chart of the entire connectivity algorithm. The central element in the process is an algorithm which determines whether or not a path exists between a node and the special node. This determination is made numerous times for each network

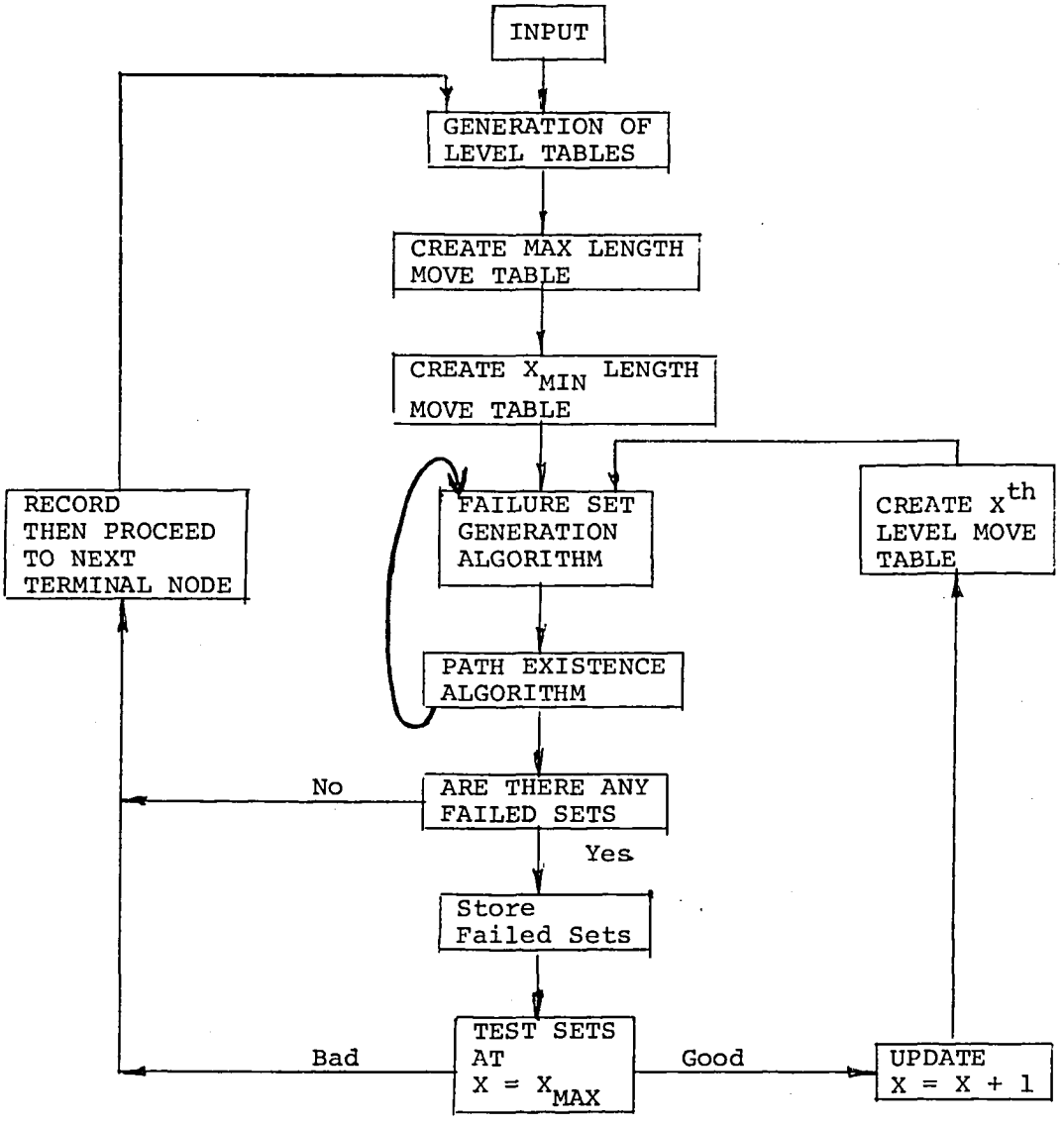


Figure 6.3.2-1. Network Failure Program.

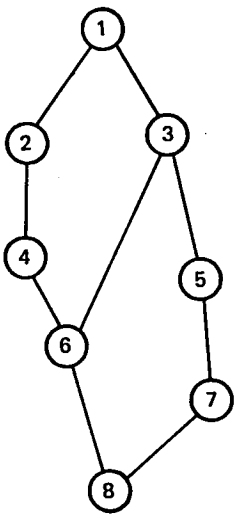
design, on a set of modified "networks" which are generated by the program by varying the length of pathways considered, the number of failures in the failure sets, and the particular nodes which are considered to have failed. The failure set generation algorithm and the general structure of the path existence determination process are designed to take advantage of network structure, and the program can provide flexibility of parameters controlling these elements which allows the user to modify program performance depending on the particular networks being analyzed.

Figures 6.3.2-2 and 6.3.2-3 are examples to illustrate the process described below.

Networks are represented by sets of nodes and links, and the input required by the program consists of numbered nodes and an indication of whether or not a link exists between any two nodes in the network. This information is contained in the connectivity table. A particular terminal node or destination is chosen. From the connectivity table, the program generates a level table consisting of a set of lists, the first of which consists of the special, or reference node. The second consists of all of the nodes which can be reached from the reference node by traversing one link. The third list contains all of the nodes which are a distance of two links away from the origin. This level table is constructed in a similar manner until the list is of maximum length. The maximum length is governed by the network structure and can never be more than N-1 long, where N is the number of nodes in the network, since it is impossible to form a path longer than this length without passing through the reference or terminal nodes more than once. The maximum length is actually shorter than N-1 when nodes which are not on any paths between reference and terminal nodes are removed from the network. These irrelevant nodes consist of nodes with only one link connection or entire clumps of nodes which form isolated areas of the networks linked to the remainder of the network by a single connecting link.

For a given terminal node, a start level table and an end level table are formed. The end level table is constructed in the same manner as the start level table described above, with the roles of the reference node and terminal node reversed. The tables are of maximum length, also described above, with the irrelevant nodes being determined during the formation of the level tables. The process of forming the level tables will

NETWORK



CONNECTIVITY TABLE

1	2, 3
2	1, 4
3	1, 5, 6
4	2, 6
5	3, 7
6	3, 4, 8
7	5, 8
8	6, 7

START LEVEL TABLE

	LEVEL	NODES
L_0	0	1
L_1	1	2, 3
L_2	2	4, 5, 6
L_3	3	2, 3, 4, 6, 7, 8
L_4	4	2, 3, 4, 5, 6, 8
L_5	5	2, 3, 4, 5, 6, 7, 8
L_6	6	2, 3, 4, 5, 6, 7, 8

END LEVEL TABLE

	LEVEL	NODES
E_0	0	8
E_1	1	6, 7
E_2	2	3, 4, 5
E_3	3	1, 2, 3, 5, 6, 7
E_4	4	1, 2, 3, 4, 5, 6, 7
E_5	5	1, 2, 3, 4, 5, 6, 7
E_6	6	1, 2, 3, 4, 5, 6, 7

Figure 6.3.2-2. Connectivity and Level Tables.

MOVE TABLE
LENGTH = N - 1

NODES		ALGORITHM
M_0	1	$M_0 = L_0 \cap (E_6 \cup E_5 \cup E_4 \cup E_3 \cup E_2 \cup E_1 \cup E_0)$
M_1	2, 3	$M_1 = L_1 \cap (E_5 \cup E_4 \cup E_3 \cup E_2 \cup E_1 \cup E_0)$
M_2	4, 5, 6	$M_2 = L_2 \cap (E_4 \cup E_3 \cup E_2 \cup E_1 \cup E_0)$
M_3	2, 3, 4, 6, 7, 8	$M_3 = L_3 \cap (E_3 \cup E_2 \cup E_1 \cup E_0)$
M_4	3, 4, 5, 6, 8	$M_4 = L_4 \cap (E_2 \cup E_1 \cup E_0)$
M_5	6, 7, 8	$M_5 = L_5 \cap (E_1 \cup E_0)$
M_6	8	$M_6 = L_6 \cap (E_0)$

Figure 6.3.2-3. Move Table.

automatically eliminate irrelevant nodes which are connected directly to the reference and terminal nodes, but the identification of irrelevant nodes in other parts of the network is ignored because of the added complexity and time penalty associated with such searching.

The next step in the process of determining path existence is the creation of move tables. (See Figures 6.3.2-2 and 6.3.2-3 again.) Move tables are of particular length X and are formed by the logical intersection of the first X lists of the start and end level tables, as demonstrated in the example. If S_i is the set of nodes contained in the i^{th} list of the start level table and E_j is the set of nodes contained in the j^{th} list of the end level table, then the i^{th} set of the move table, M_i is:

$$M_i = S_i \cap \left(\bigcup_{j=0}^{X-i} E_j\right) \quad \text{for } 0 \leq i \leq X.$$

Therefore, the resulting move table presents information about paths of length X or less between the reference and terminal nodes.

The program generates a move table of maximum length to be utilized at various stages throughout the process. The program then generates move tables of variable length beginning with the minimum length necessary for a path to exist between reference and terminal nodes. The length of this move table is increased, as the process progresses, in order to determine whether longer paths exist which restore connections caused by particular failures simulated by the algorithm.

The stage is now set for determining path existence for a particular terminal node with the reference node. The process proceeds as follows:

1. The failure set generation algorithm produces a set of nodes to be failed. (This process will be described later.)

2. The nodes which are failed cause changes to connectivity and to the move table, which starts out to be of minimum length. To determine if there are any paths remaining in the network, the connections between nodes in adjacent lists of the move table are checked by using the failure-modified connectivity table. When a node is no longer connected, it is deleted from the move table. If an empty list in the move table is produced, then the terminal node is isolated from the reference

node, but only by this particular length of path.

3. Another failure set is generated and path existence is tested as in 2. All failure sets which cause isolation are stored.

4. If all failure set candidates have been tested and no isolation has been caused, the entire process is restarted with a new node as the terminal node.

5. The maximum length move table is used to see if any of the stored failure sets cause isolation in the complete network. If any sets are found to cause isolation and this stage, much computer time has been saved.

6. If some stored failure sets do not cause isolation in the maximum length move table, the length of the move table is increased. The increase shown in Figure 6.3.2-1 is one unit but the designer can use his judgement in selecting this number. Experience has shown the increase of one unit performs well. The added computation time due to increased-length move tables more than counter-balances the extra time cost of checking some failure sets several times as the move tables are increased by only one unit at a time. This is due to the fact that for long move tables, many more possible paths exist.

7. The procedure continues until all nodes but the reference node have been terminal nodes. Upon completion the user knows the minimum number of node failures which will isolate an innocent node and also what nodes are contained in failure sets causing such isolation. The nodes which have been isolated by the failure sets can also be identified.

6.3.3 Failure Generation

Little has been said thus far about the selection of the failure sets. It is in this aspect of the problem that the most significant gains have been made in reducing program running time for certain network structures. If one simply wants to determine the number of failures which will isolate an innocent node, then prudent choice of the order in which failed nodes are simulated can have a great effect on how "early" this minimum number is found. If, however, a more complete presentation of information is requested by the designer, this program running time can be quite a bit longer.

The user selects the maximum number of nodes which can be failed at one time for which he wishes to check a particular network. The failure generation algorithm, based on a set of design principles, produces failure sets according to a system of priorities. These principles are as follows:

1. Of course, single failures are tested first, then double failures, etc.

2. Only relevant nodes need to be selected.

3. Nodes which are linked directly to the reference and terminal nodes are given priority. These nodes are elements of the cutsets of the reference and terminal nodes and with higher probability will be entries of failure sets which cause isolation of innocent nodes.

4. In a way similar to 3, nodes closer to the end points are given higher priority.

5. If a failure set is found to cause isolation, then all failure sets which include that set as a subset need not be considered further, unless particular design information is sought.

6. If a failure set causes isolation, increasing the length of the move table may restore a path.

7. If a failure set causes isolation at maximum length, no path exists between the reference and terminal nodes.

8. If a failure set does not cause isolation, one more node is added on to this failure set to create a new failure set. This is repeated until the maximum number of failed nodes per set, selected by the user, is reached.

6.3.4 Performance

The performance of this design tool has not been evaluated extensively. However, the sensitivity of the length of running time to such heuristic changes as described in the principles used for failure set ordering is important, and significant performance increases have been made by implementation of these principles. Further program development would undoubtedly bring about still better performance.

Figures 6.3.4-1 and 6.3.4-2 are two examples used to assess program performance. The first is a regular, hexagonal network of 96 nodes each with three links. An Amdahl 470/V8 took 11.5 minutes to verify that this network can survive any two failed nodes at one time. The second example is a 66-node network which took only 4.25 minutes to check.

6.4 Bus Reliability and Dispatch Probability

Bus reliability modeling is fairly straightforward compared to network reliability modeling. Network hazards include complex topological factors absent from bus hazards. When network terminals fail passively, they do not disturb the bus. When a terminal port creates a short circuit, it is tolerated if the bus is a 1553 data bus, but not if it is an unprotected bus such as a local parallel bus, and not if the bus is a power bus. Two or more short circuits bring down one or all copies of a redundant bus. An open circuit fault in a bus makes it partly or totally useless.

It is difficult to estimate the failure rates associated with the hazards listed above. Passive, short-circuit, and active faults depend on the technologies and designs of the terminals. Bus faults depend on routing and environmental factors. Once these rates are known, bus reliability can be expressed combinatorially using a failure modes and effects analysis (FMEA). An examples of a top-level FMEA is given in Table 6.4-2 for a 1553 bus system.

The dispatch criterion for a redundant bus system must be to have at least two unfailed copies. Thus a dual bus system may not be dispatched if either bus is faulty. One terminal on each may be short circuited, and fewer than K terminals may have passive faults. For redundancies greater than dual, obviously, a greater variety and extent of failures is permissible. As before, calculation of dispatch probability involves simple combinatorics once the failure rates are estimated.

96-NODE NETWORK
(11.5 min TO EVALUATE)

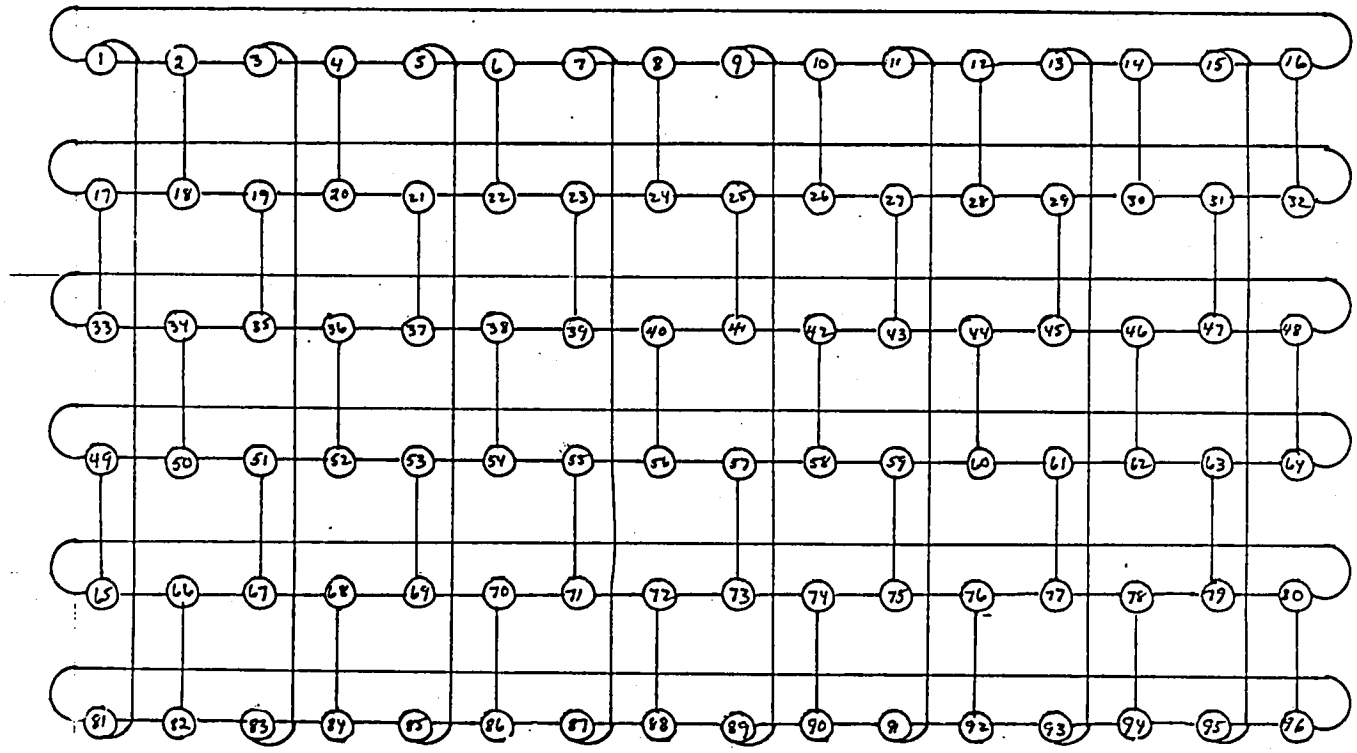


Figure 6.3.4-1

66-NODE NETWORK
(4.25 min TO EVALUATE)

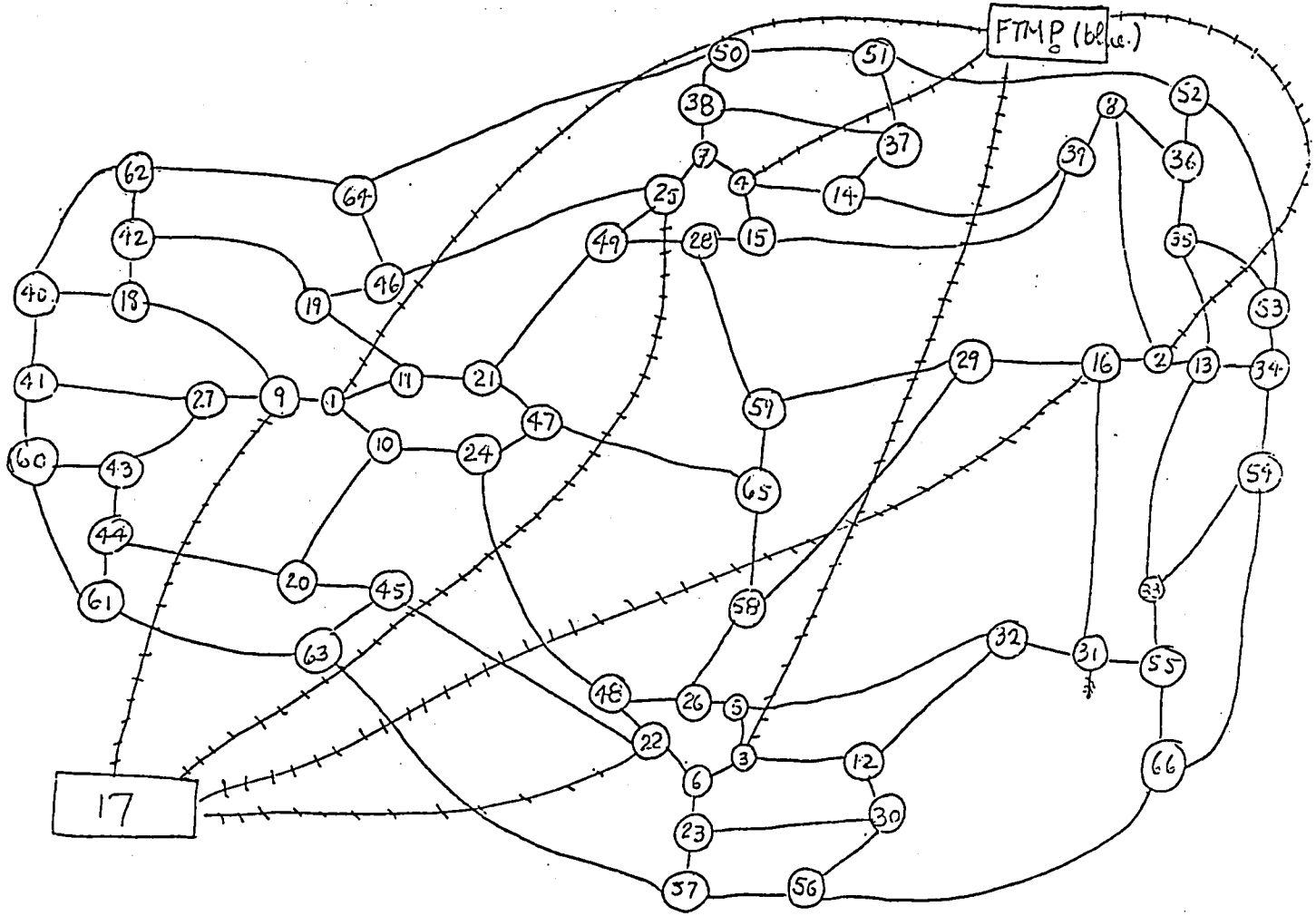


Figure 6.3.4-2

TABLE 6.4-1

FMEA OF 1553 BUS

Loss of Communication Can Result from:

1. Active fault on all copies of bus or
2. Loss of all copies of bus, where
   Loss of one copy can result from
   a. Open or short bus fault or
   b. Active terminal fault on one copy of bus or
   c. Two short terminal faults or
   d. K passive terminal faults, where
      K is determined by relationship of specific fault set to minimum equipment criterion of subscribers.

6.5 Remote Power Control Reliability

Remote power control has the advantages of substituting buses for dedicated links, and removing a single-point system damage vulnerability. The potential hazard of a remote system is the failure of remote switches or limiters, or of the communications that support them. It is desired to determine the necessary levels of switch reliability, as well as reliability of other elements, to support flight safety.

Figure 6.5-1 shows one possible model of a power distribution system. Four prime power sources, A, are protected by breakers, B. These breakers may be remote or dispersed in the cockpit. Buses, C, distribute power to remote switches, D, serving subscribers, E. The remote switches are controlled by communication, F, with redundant controllers, not shown.

The failure modes are as follows. First, both buses fail. The failure of a single bus can happen from loss of two generators A or two breakers B or a combination of the two, or an open or short fault of the bus C, or a short fault of any subscriber with a switch stuck on. The second failure mode is where too many subscribers become unpowered. Say that K subscribers have both switches stuck open. The third failure mode is a combination of the first two.

The failure probability can be approximated in terms of the

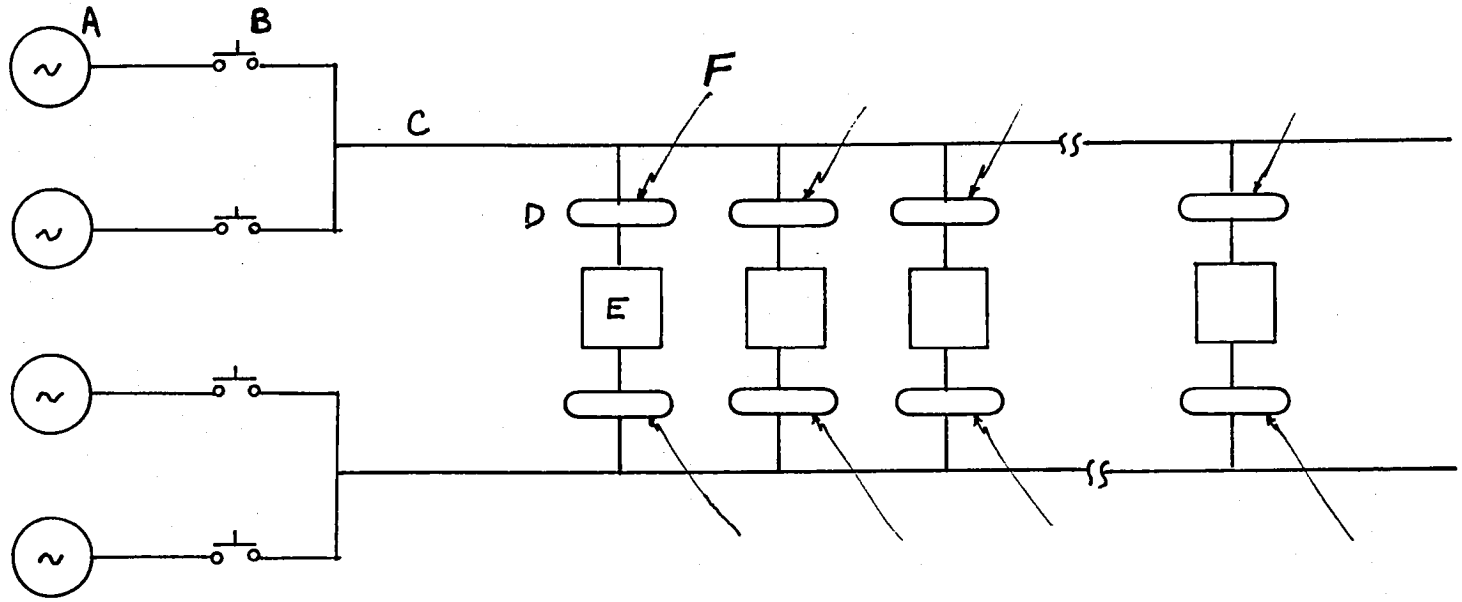


Figure 6.5-1. Remote Power Distribution Model.

respective falure probabilities of the elements, as listed below.

A = probability of generator failure

B = probability of breaker failure

C = probability of bus open or short

D_0 = probability of an open switch

D_1 = probability of a switch stuck on

E_1 = probability of a shorted subscriber

F_0 = probability of a switch command failure causing all switches to open

F_1 = probability of a switch command failure causing all switches to close.

Let N be the number of subscribers. Now the probability of one failed bus is approximately as follows:

$$P_1 = (A+B)^2 + C + NE_1 (D_1 + F_1)$$

The probability of K subscribers not able to be powered from a single bus is approximately:

$$P_2 = \binom{N}{K} D_0^K + F_0$$

The probability of K subscribers unable to receive power from either bus is approximately:

$$P_3 = \binom{N}{K} D_0^{2K} + \binom{N}{K} D_0^K F_0 + F_0^2$$

The probability of having two failed buses is approximately:

$$P_4 = NE_1 (D_1+F_1)^2 + 2 NE_1 (D_1+F_1) [(A+B)^2+C] + [(A+B)^2 + C]^2$$

P_3 and P_4 roughly resemble the squares of P_2 and P_1, respectively. They differ from the squares because subscribers are common to both buses.

The overall failure probability is given approximately by:

$$P = P_4 + P_3 + P_1 P_2$$

Table 6.5-1 contains some sample evaluations of the model. Case 1 lists parameter values chosen at random, and the resultant values of P_1 through P_4 and P. Case 1 values were chosen well on the

TABLE 6.5-1

REMOTE POWER MODEL EVALUATIONS

PARAMETER	CASE					
	1	2	3	4	5	6
K	3					
N	100					
A	10^{-3}					
B	10^{-3}					
C	10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-5}	10^{-5}
D_0	10^{-3}	10^{-3}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
D_1	10^{-3}					
E_1	10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-5}	10^{-6}
F_0	10^{-3}	10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-5}
F_1	10^{-3}					
P_1	1.2×10^{-4}	1.2×10^{-4}	1.2×10^{-4}	1.6×10^{-5}	1.6×10^{-5}	1.4×10^{-5}
P_2	1.2×10^{-3}	2.7×10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-5}
P_3	1.2×10^{-6}	2.7×10^{-8}	10^{-8}	10^{-8}	10^{-10}	10^{-10}
P_4	5.5×10^{-8}	5.5×10^{-8}	5.5×10^{-8}	4.3×10^{-9}	4.3×10^{-9}	6.1×10^{-10}
P	1.4×10^{-6}	1.2×10^{-7}	7.7×10^{-8}	1.6×10^{-8}	4.5×10^{-9}	8.5×10^{-10}

pessimistic side. Succeeding cases changed one or two parameters at a time. The value of any parameter for any case is either shown explicitly or is the same as for case 1. Case 6 results in P being below 10^{-9}. To reach this value, E_1, the subscriber power short probability was put at 10^{-6} to reduce the first term of P_4 without having to assume anything more stringent than 10^{-3} for D_1 and F_1. Meanwhile, D_0 and F_0 are assumed to be 10^{-5}. This corresponds to switches and controllers being designed to be biased to fail in the on position rather than off.

6.6 Power Network Using Current Limiters

Data for Tables 4.5-1 and 4.5-2 was obtained by numerical solutions of a mathematical model of the network in Figure 4.5-1 with the piecewise linear approximation of a limiter characteristic shown in Figure 4.4-2. Thirty nonlinear nodal equations were solved by relaxation.

6.7 Reliability Analysis Tool

This section describes the reliability analysis tool used in Volume 2 to compare the reliabilities of the various communication structure alternatives. This analysis tool is an interactive computer program which provides a convenient technique for constructing the combinatorial reliability equations for complex, highly interconnected systems. The equations are constructed and presented in the form of a diagram. This diagram facilitates system analysis by clearly showing the constituent terms that make up the probability of failure of the total system. The diagram makes it easy to identify the most critical failure modes and provides a means for evaluating, quantitatively, the relative contributions to the unreliability of the system made by the different parts of the system and by the different failure modes.

The basic concept for this analysis method was developed during a reliability analysis study of the NASA F-8 Digital Fly-by-Wire system [8]. Its development was motivated by a perceived inadequacy and/or inconvenience of most of the existing reliability analysis techniques. The technique proved to be valuable in the F-8 DFBW study. It allowed the equations describing the unreliability of the total system to be developed and displayed as a diagram which clearly showed the interrrelationships and relative importance of the various system elements. For example, it was easy to identify the most critical failure modes and show what modifications could be made in the design

to improve reliability and what this improvement would be quantitatively.

During this present study, the technique was developed further to provide a more general user interface. The original F-8 DFBW program was written just for that system. The modified program allows the definition of the element failure models and the generation of the reliability equation diagram by successive interactive computer terminal sessions.

This section first gives the mathematical foundations for this technique. A description is then given of the operation of the program which is then illustrated by an example.

6.7.1 Mathematical Basis for the Analysis Technique

The mathematical basis for this technique is the repeated application of the basic conditional probability equation. This equation can be written in the form

$$Q(S) = Q(S/A)\ P(A) + Q(S/B)\ P(B) + Q(S/C)\ P(C) + \ldots$$

where

$Q(S)$ = unreliability of the system
$Q(S/A)$ = unreliability of the system given event A
$A,B,C,\ldots$ = events describing the state of a particular system element
$P(A)$ = probability of event A

and the following conditions must be met for the events

$$P(A) + P(B) + P(C) + \ldots = 1 \text{ (exhaustive)}$$
$$P(AB) + P(AC) + P(BC) + \ldots = 0 \text{ (mutually exclusive)}.$$

This equation can be represented graphically as shown in Figure 6.7-1. It should be noted that the equation is written for unreliability instead of reliability in order to avoid the necessity of multiplying reliability terms that are almost equal to unity.

The reliability of the total system is found by successively applying this equation, considering each basic system element until all of the conditional probabilities are defined and the point is reached where the probability of failure of the total system can be defined as either one or zero. The process of applying this technique can be described by the following five steps:

Step 1: Partition the System into Basic Elements.

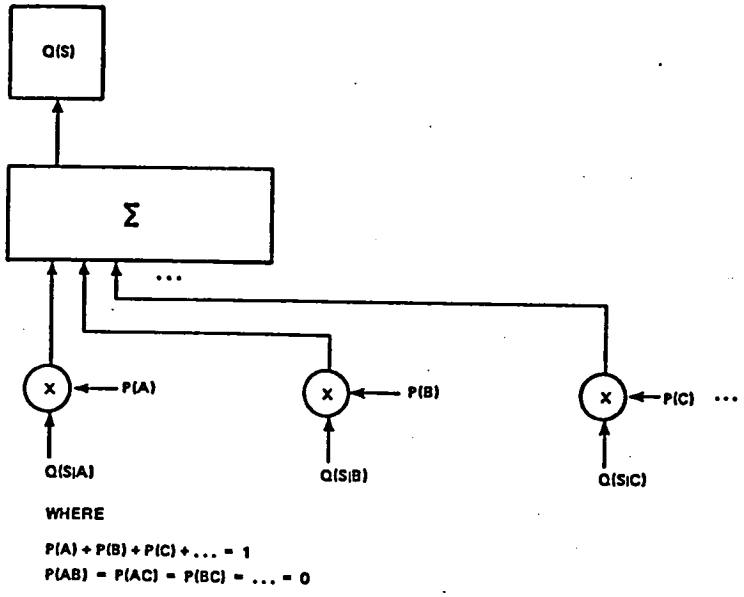


Figure 6.7-1. Graphical equivalent of basic equation.

The system must be divided into a number of basic elements. The number of elements should be kept as small as possible as long as the total system is accurately represented. The elements are defined by random failure containment boundaries that are made as large as possible subject to the constraint that any failure within the element prevents any other part of the element from being used.

Step 2: Identify Events that Define the State of Each Element

The events which describe the operational state of each element must be described and must meet the requirements of the basic reliability equation: they must form an exhaustive and mutually exclusive set. For simple elements, there may only be two events: the element is good or the element is bad. In other cases, it is necessary to distinguish failure modes. For example, an element such as an actuator may fail passively or fail hardover. In other cases, it may be necessary to distinguish whether the failure of the element was covered or not, i.e., whether the failure was detected by the system and the effects of the failure were automatically accounted for.

In many cases, the analysis can be simplified by grouping elements together as, for example, all three channels of a triplex actuator or three redundant electrical power supplies. If the elements are grouped, the events must include all combinations of failures, such as all channels good, one failed and the others good, etc. The important consideration is the identification of all events that have significance in defining the state of the other elements in the system. In some cases, it may be useful to use a Markov process to define the states of an element. Along with defining the events, it is also necessary to calculate the probability of each event. In most cases, these probabilities will be functions of the reliability of the elements calculated by an FMEA or similar technique.

Step 3: Select an Order for the Application of the Equations

An order must be selected whereby the reliability equations will be applied to the basic system elements. The order in many cases can be arbitrary, but the analysis can be simplified with a prudent choice of order. The primary factor determining order is dependency. Elements on which other elements depend for proper operation, such as power supplies, are placed first. The states of these elements must be

determined before it is known how to analyze the elements that depend on them. The order will normally follow the signal flow through the system. The best order, however, will be based on practical experience with the system combined with trial and error.

Step 4: Construct a Diagram of the Equations

The equation giving the probability of failure for the total system is now generated by interconnecting the diagram segments representing all of the basic elements of the system. The diagram for the first element is constructed first. The inputs to this diagram come from the diagrams for the next element, and so forth, until the success or failure of the system can be defined. When enough elements are defined as good to guarantee the success of the system independent of the state of the other elements, the input to the diagram will be an unreliability of zero. In other words, the conditional unreliability of the system given that sufficient elements are good is zero. On the other hand, if enough elements have failed such that the system fails, independent of the state of any of the other elements in the system, the unreliability is one. The unreliability of the system is thus determined by interconnected diagram segments which define the state of the system and eventually have inputs of ones or zeros.

The inputs to the diagram segments are the unreliabilities of the system due to failures in all following elements. The unreliability of the system due to these following elements is based on the state of the system as defined by the previous elements. The diagram grows geometrically with one diagram for the first element followed by a diagram for the second element for each event used to define the first element. The total diagram for a real system with a reasonably large number of elements could become completely unmanageable if it were not for the fact that many of the system states are equivalent and do not have to be repeated. For example, if communication fails to a device by one path but can be re-established by another, the unreliability of the rest of the system is likely to be equivalent.

Step 5: Compute the System Unreliability

The final step is to insert the values for the probabilities of the events for each element and then perform the required arithmetic operations. The equations must be solved from the bottom up. For simple systems, the entire process can be done by hand with a calculator. For more realistic systems, a computer program is useful. An

interactive program was developed that aids in the construction of the diagram and computes the resulting unreliabilities.

6.7.2 Operation of the Computer Program

A program was written as a part of this effort that greatly simplifies the process of constructing the equation diagram. This program is not a fully developed software product, but it has served very well during this study as a tool for determining the reliability of the various communications structure options. The operation of the program is described briefly in the following paragraphs. This description is not intended to be a users guide but only to provide a background for the analysis presented in Volume 2.

The program is intended for interactive use at a terminal. It is presumed that any reasonably sized system would not be analyzed in one session. The program is thus designed to be interrupted at any time and started again where left off.

Inputs are entered into the program in response to its request for a command. Commands to the program start with the letter 'z'. The first step is to establish the definition, failure states, and failure probabilities of the basic elements into which the system has been partitioned. The command for establishing this model for the basic system elements is 'zm'. The program then asks whether you want a previously stored model, to modify the stored model, or create a new model. For a new model the program first asks for the time parameter that is to be used with the failure rates that are to be specified. It then asks for the name of the first element. The order in which elements are entered is arbitrary. The first three letters of the name of the element is used as the identifier for that element. The program then asks for two letter codes to identify the failure states of that element and the associated failure rate for that state. The program assumes that the first state for each element is the good state and automatically computes the probability as one minus the probability of the failure states.

After the model is entered the process of constructing the equation diagram begins. In response to the request for a command the analyst enters the three letter code for the first element in the order determined in Step 3. The program responds with the current state of the system. At this point the only term will be that this first element is good. The failed states for each element are automatically

entered by the program into a push-down list to be considered later. The analyst then enters the next element in the sequence. The program responds with the new state which is these first two elements good. The process of adding elements continues until it can be determined that the probability of system failure is zero regardless of the state of any remaining elements that have not yet been considered. At this point a 'q0' is entered as a command. This tells the program that the system in the currently defined state has a unreliability of zero. The program then responds with a new current state of the system. This new state is determined by taking from the push-down list the first failure state for the last element entered. The analyst then enters commands to tell the program the implications of the new state. If additional resources are available in the form of other elements, which is normally the situation in a redundant fault-tolerant system, the analyst enters the codes for these additional elements as before until a state is reached where the system unreliability is zero. Eventually a point will be reached where the system probability of failure is equal to one, independent of the state of any remaining elements. At this point the analyst enters 'q1' to indicate that the unreliability is one. This process continues until all system states have been accounted for.

The program automatically computes the unreliabilities for the various subparts of the system as it has sufficient information to make those computations. The form of the equation diagram is stored as matrices which contain the alphanumeric names which identify the elements and their associated states. These identifiers are arranged in the matrix in a way that reflects the form of the equation diagram. The computed unreliability numbers are arranged in matrices of the same form. These matrices can be viewed at the terminal or printed on a printer. The equation diagram for any realistic system is, of course, too big to be viewed all at once on a terminal. If a command 'zdnn,nn' is given, the terminal displays the part of the diagram starting at the upper left hand corner with the row and column defined by 'nn,nn'. The command 'zp' causes the diagram to be printed. If the diagram is too big for one page, it is printed on multiple pages and can be assembled later.

As was discussed in Step 4 above, the size of the diagram for any realistic system would get completely unmanageable if it were not for the fact that many of the different system states are equivalent and do not have to be reconstructed or recomputed. This is handled by

the program by referring to the matrix location of the point at which the diagram is equivalent. The command 'q(nn,nn)' is typed, which means that the unreliability of the system with the current state is the same as that at the point defined by 'nn,nn' that has already been computed.

A particular terminal session can be terminated at any time by entering the command 'zz'. The analysis process can be continued by reinstating the stored element model (or modifying the model if desired) and entering 'zx' which regenerates the previously entered diagram. This process is also used to evaluate the effects of changes in failure rates or system design.

6.7.3 Example Illustrating the Equation Diagram Analysis Technique

This analysis method can best be illustrated with an example that is typical of the systems that it will be used to analyze in Volume 2. This example is a simplified version of a flight control system which uses a small communications network to connect a fault-tolerant computer system to two dual actuators. A schematic diagram of the system is shown in Figure 6.7-2. A diagram showing the interconnections in the communication network is shown in Figure 6.7-3. The links into the computer are designated by the points A and B. The function of the system is assumed to be the control of split aerodynamic control surfaces, one by each of the dual actuators. The flight control function is assumed to be performed if either actuator is operating. Each of the channels in the dual actuators has two failure modes; a passive one and a hard-over failure. If one channel fails passive and the other is good, the actuator can still perform its function. However if either channel fails hard-over, that actuator fails. If both actuators fail, the system fails.

There is a node of the communication network attached to each actuator channel. Each node has three communication links that go to other nodes or to the computer. Communications fail to an actuator channel if there is no operational path of good links and nodes between the computer and the channel.

The first step in applying this analysis technique is to partition the system into basic elements. The system is divided into three different types of elements: actuator channels, communication nodes, and communication links. The actuator includes the hydraulic channel itself, the servo electronics, and all of the interface

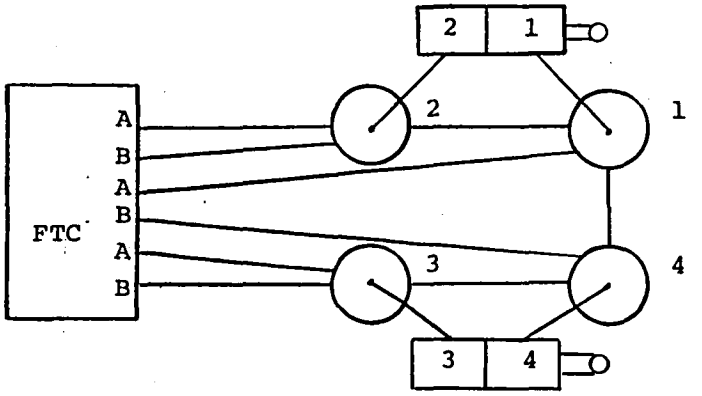


Figure 6.7-2. Schematic Diagram of the Example System.

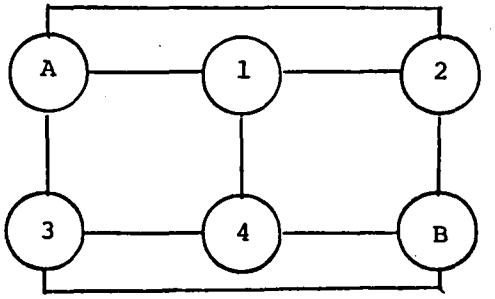


Figure 6.7-3. Network Interconnections.

electronics and wiring that is unique to that channel. The link includes all parts of the communication equipment that are unique to that path between nodes or between a node and the computer. Included are the wires, connectors, and the interface electronics at each end. The node element includes all parts of the communication terminal that are common to all paths. This partitioning of the elements is shown in Figure 6.7-4. The fault tolerant computer itself is assumed to be much more reliable than the other parts of the system. Also there are no complex interactions between computer failure modes and the rest of the system. The unique electronics in the computer for each link is included in the link element. If the computer is working it is assumed to be able to service all links. If it fails, the whole system fails and this unreliability can be simply added to that resulting from the rest of the system.

The next step is the identification of the states of each of the elements. The nodes and links can be adequately described by two states: good and failed. The actuators are described by three states: good, failed passive (soft), or failed active (hard). For this example it is assumed that the probability of being in a failure state as a function of time is given simply by a failure rate times time. More complex descriptions can be accommodated by the model if necessary. The numbers assumed by this example are:

node failure rate	2.0 x 10-4
link failure rate	1.0 x 10-5
actuator channel failure rate	
soft	3.0 x 10-4
hard	3.0 x 10-6

Step three in this analysis process is to determine the order in which the equations are to be constructed. The order for this example is chosen by selecting the minimum set of elements that is necessary to communicate with one actuator channel plus assure that the system can operate. A node must be good before any of the links through it can be used. Then a link must be established to the computer. Next the actuator channel itself must be good. Finally it must be assured that the other channel of that dual set has not failed hard. If it has not, the unreliability of the system is zero independent of the state of any of the other elements. If it has failed

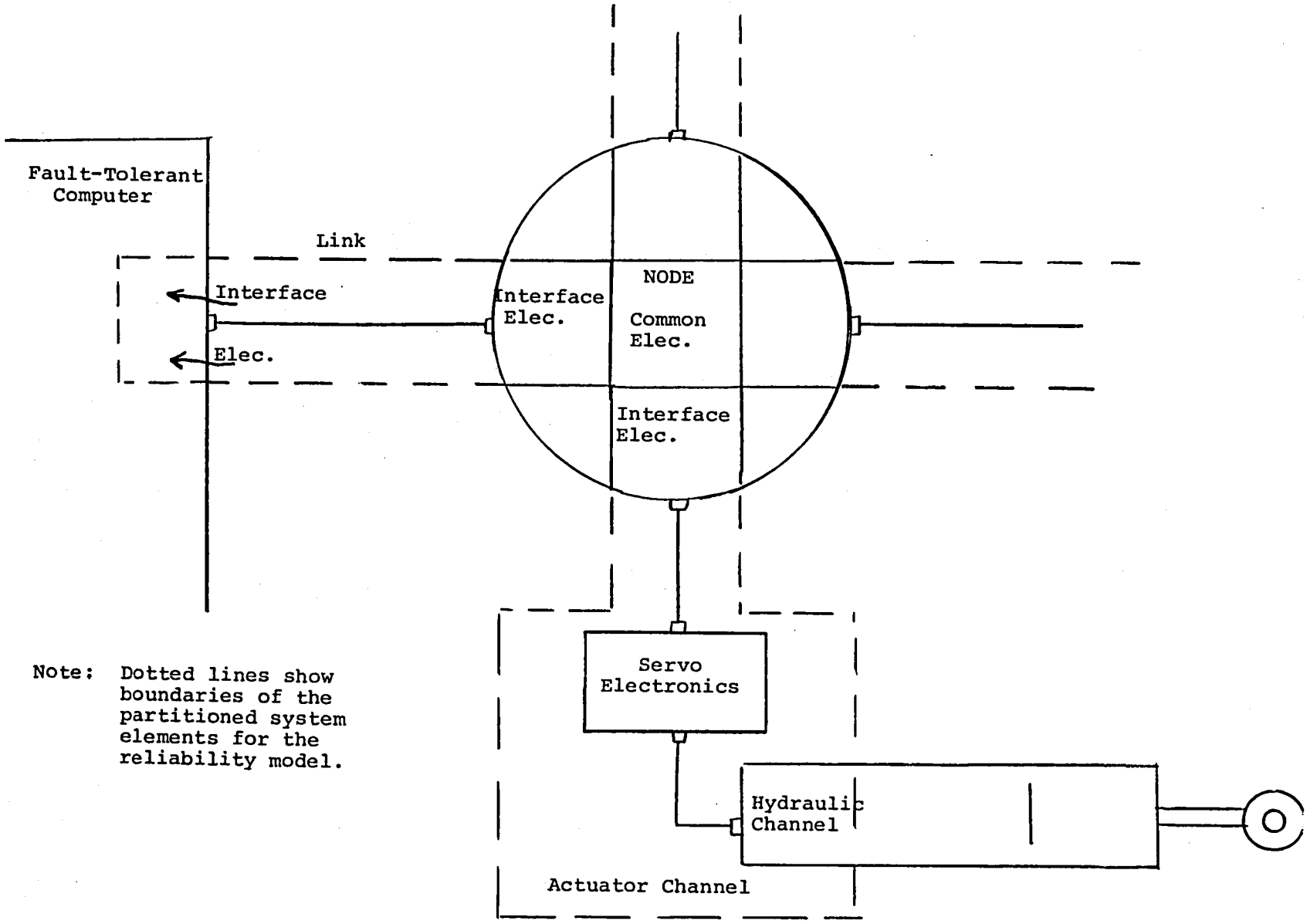


Figure 6.7-4. Partitioning of System Elements.

hard, a path must be established to the other actuator and it must be determined if it is good or failed. This process is continued until all states are accounted for.

The equation diagram is then constructed with the aid of the computer program. Several stages in the construction of the diagram are shown in figures starting with Figure 6.7-5. The elements and their states are defined by the alpha-numeric symbols. Typical definitions are:

nodl G	node 1 good
nodl fl	node 1 failed
lnkAl G	link from computer port A to node 1 good
lnkAl fl	same link failed
actl G	actuator 1 good
actl sf	actuator 1 failed soft
actl hd	actuator 1 failed hard

The diagram is read from the top left corner. Lines have been added to the computer print-out to show the relationship between elements. The first state considered is 'node 1 good' followed by 'link from the computer to node 1 good', 'actuator channel 1 good', and 'actuator channel 2 good'. At this point 'q0' is entered to indicate that the unreliability of the system is zero. The next system state that is automatically considered by the program keeps the first three elements the same (good) and adds 'actuator 2 failed soft'. This system state also has reliability of zero since channel 1 is still good. The next state changes actuator 2 to failed hard and at this point the first dual actuator has failed. It is now necessary to consider other elements to determine the state of the other actuator.

Node 3, link from A to 3, actuator 3, and actuator 4 are then considered in a similar way as before in next state in the development of the diagram as shown in Figure 6.7-6. Again for the states containing actuator 4 good and failed soft, the unreliability of the system is zero. For the state including actuator 4 hard failed, however, the system unreliability is one since the other dual dual actuator has already been determined to be failed.

At this point in the example there is enough information for the program to start computing numbers as shown in Figure 6.7-7. The numbers below the alpha-numeric labels are the products of the probability of the state of the element times the appropriate conditional

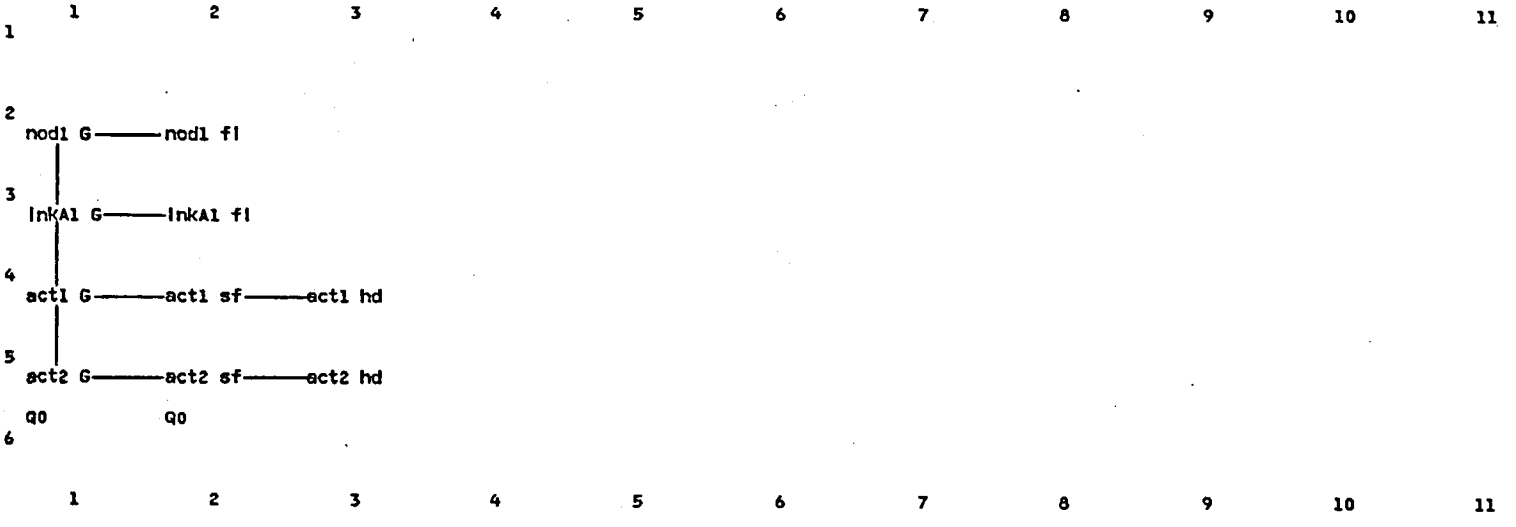


Figure 6.7-5 Equation Diagram Stage 1.

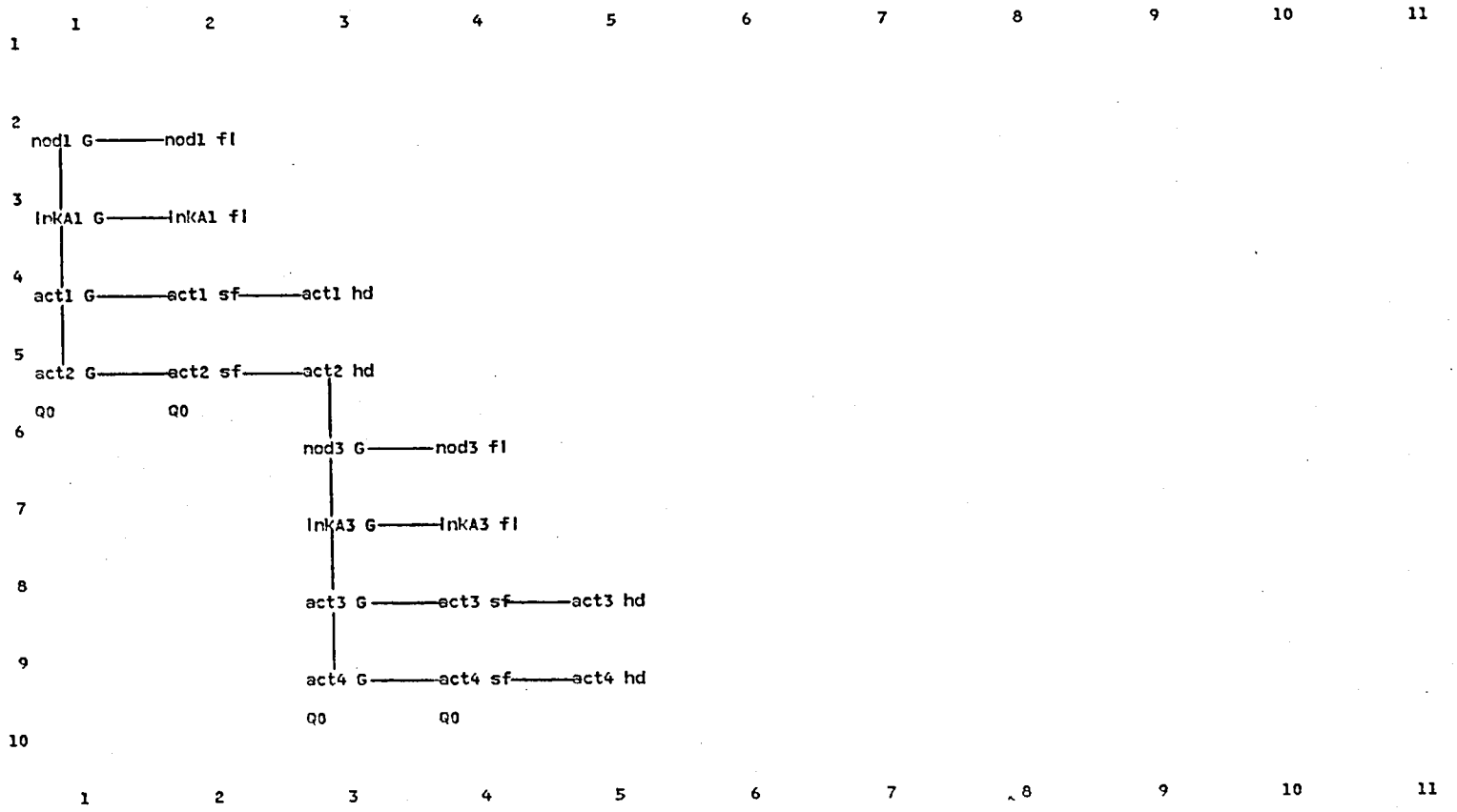


Figure 6.7-6 Equation Diagram Stage 2.

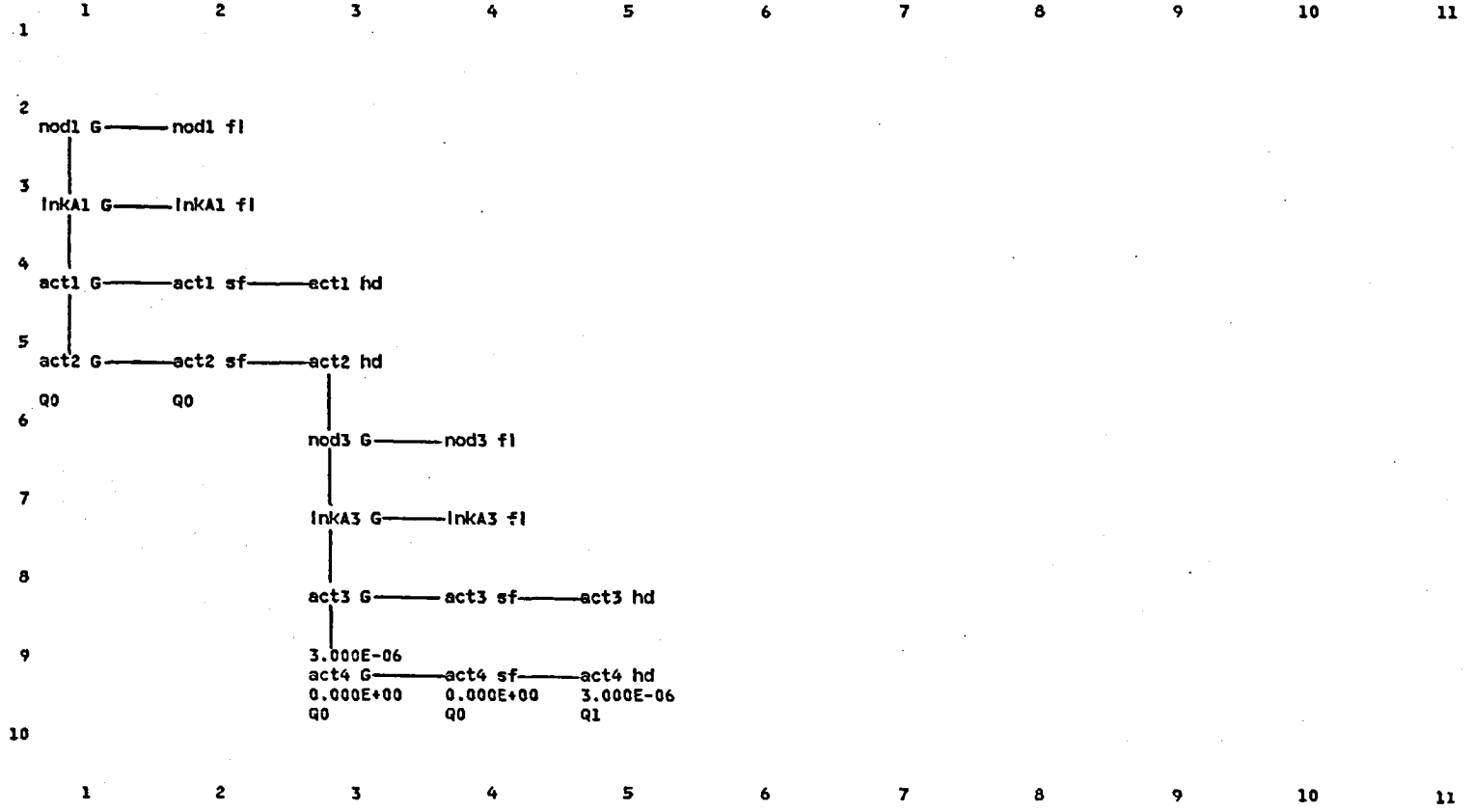


Figure 6.7-7 Equation Diagram Stage 3.

unreliability. This conditional unreliability is the unreliability of the system given the states of the elements that have already been established. Its value is dependent on the states that have not yet been considered, based on that particular system state. In general this conditional unreliability will depend on further development of the equation diagram. At this point in this example, however, these conditional unreliabilities are known to be 0, 0, and 1 respectively independent of the state of any other system elements, so the appropriate products can be computed as shown. The number above the 'act4 g' is the sum of the three act4 terms below and is the conditional unreliability for that substate of the system and serves as the input to the next level.

The next state that is drawn automatically by the program from its push-down list adds 'actuator 3 failed soft' with all previous elements in the same state. Now it must be determined not only that actuator 4 is good but that it is in communication with the computer. Thus node 4, link from B to 4, and actuator 4 are added to the diagram. If the actuator is good the system unreliability is 0, but if it has failed soft or hard, the unreliability is 1. The added diagram and numbers is shown in Figure 6.7-8.

The next state considered ends with the link from B to 4 failed. An alternate link from 3 to 4 can be used. (The fact that node 3 and link B to 3 are good is already established by the system state.) If this link is good, an equivalent system state exists to the one where link B to 4 is good thus the previously computed conditional unreliability can be transferred by entering its coordinates '(11,6)' as shown in Figure 6.7-9. If the link from 3 to 4 fails, the link from 1 to 4 can be used. If this link fails, the system fails due to the failure of communication to actuator channel 4, the soft failure of actuator channel 3 and the hard failure of actuator channel 2.

The same process is continued until all system states are considered. The entire diagram is shown in Figure 6.7-10. This diagram does not model all failure modes exactly. Certain approximations have been made to simplify the analysis wherever it can be shown that the effects of the approximation on the analysis are completely insignificant. These points are marked by appending the letter 'X' to the alphanumeric designator to remind the analyst that an approximation has been made.

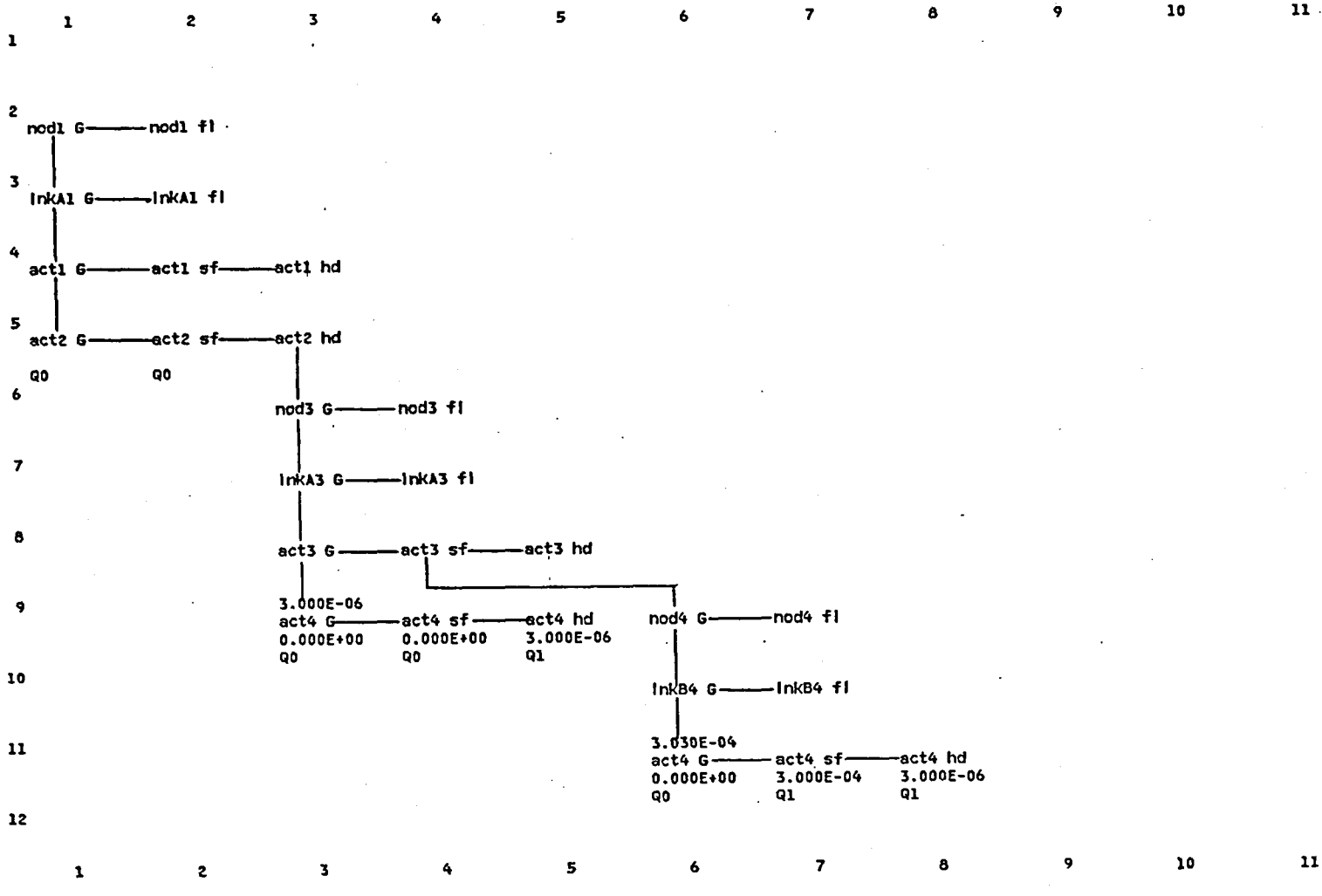


Figure 6.7-8 Equation Diagram Stage 4.

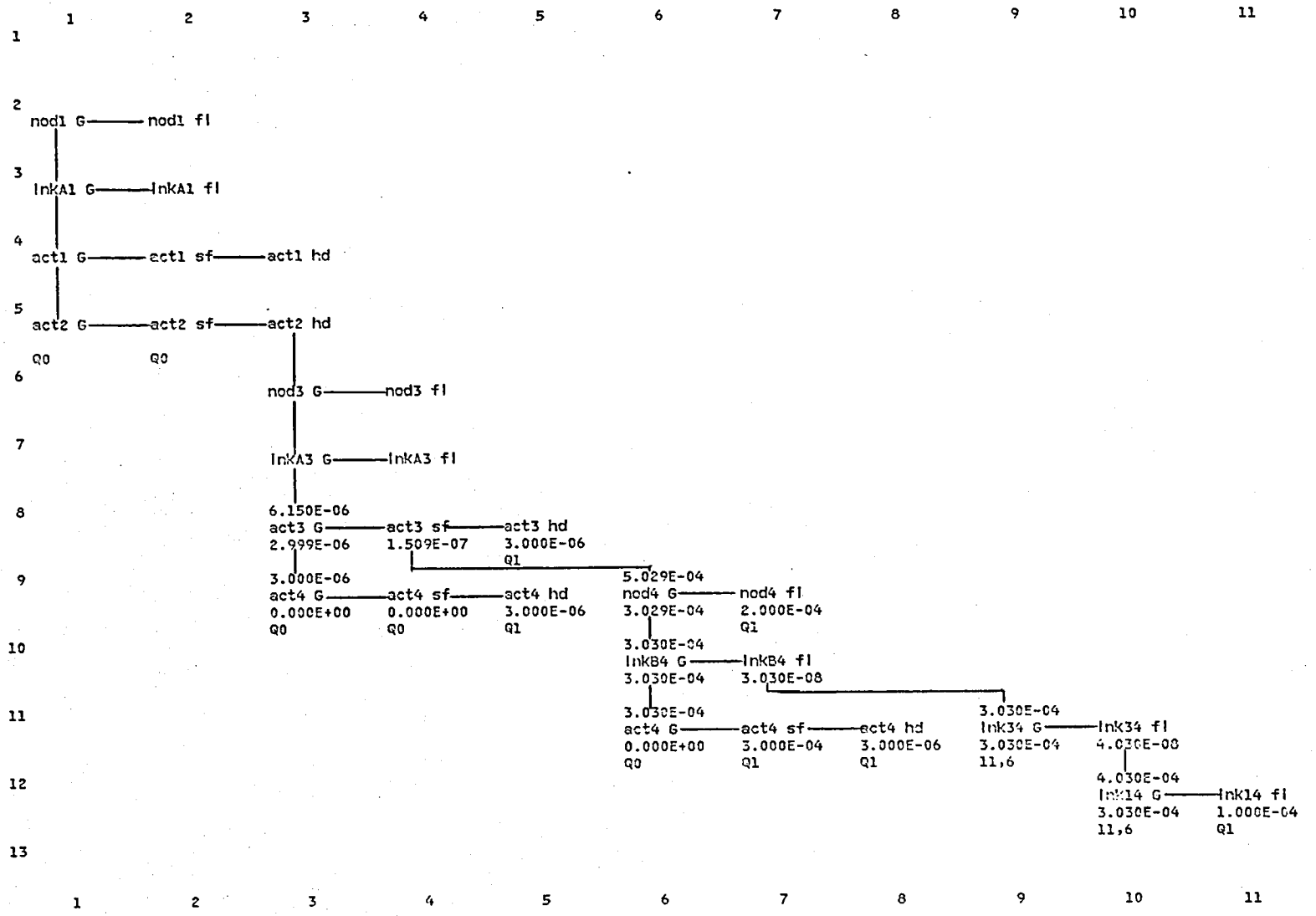


Figure 6.7-9 Equation Diagram Stage 5.

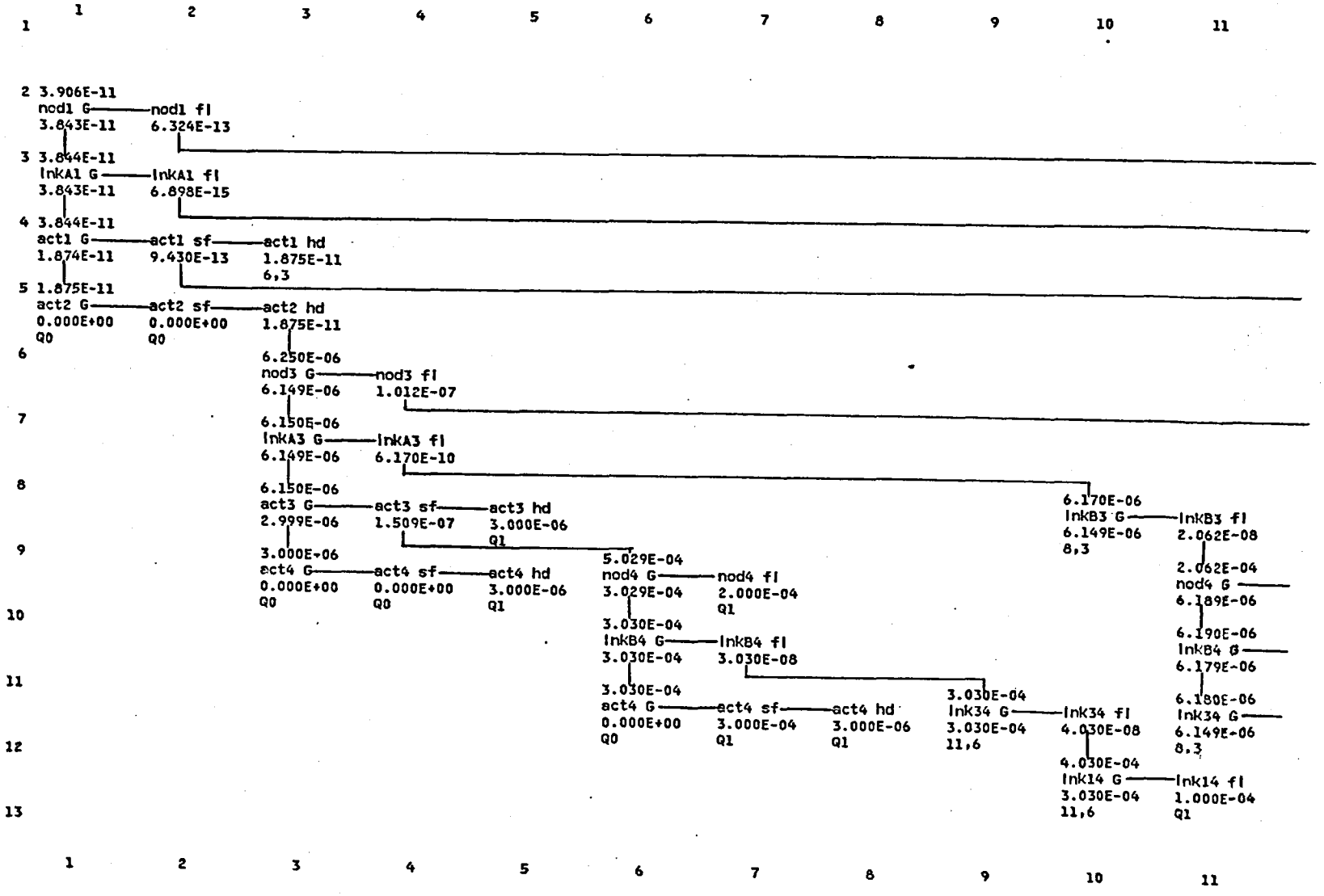


Figure 6.7-10 Equation Diagram Stage 6.

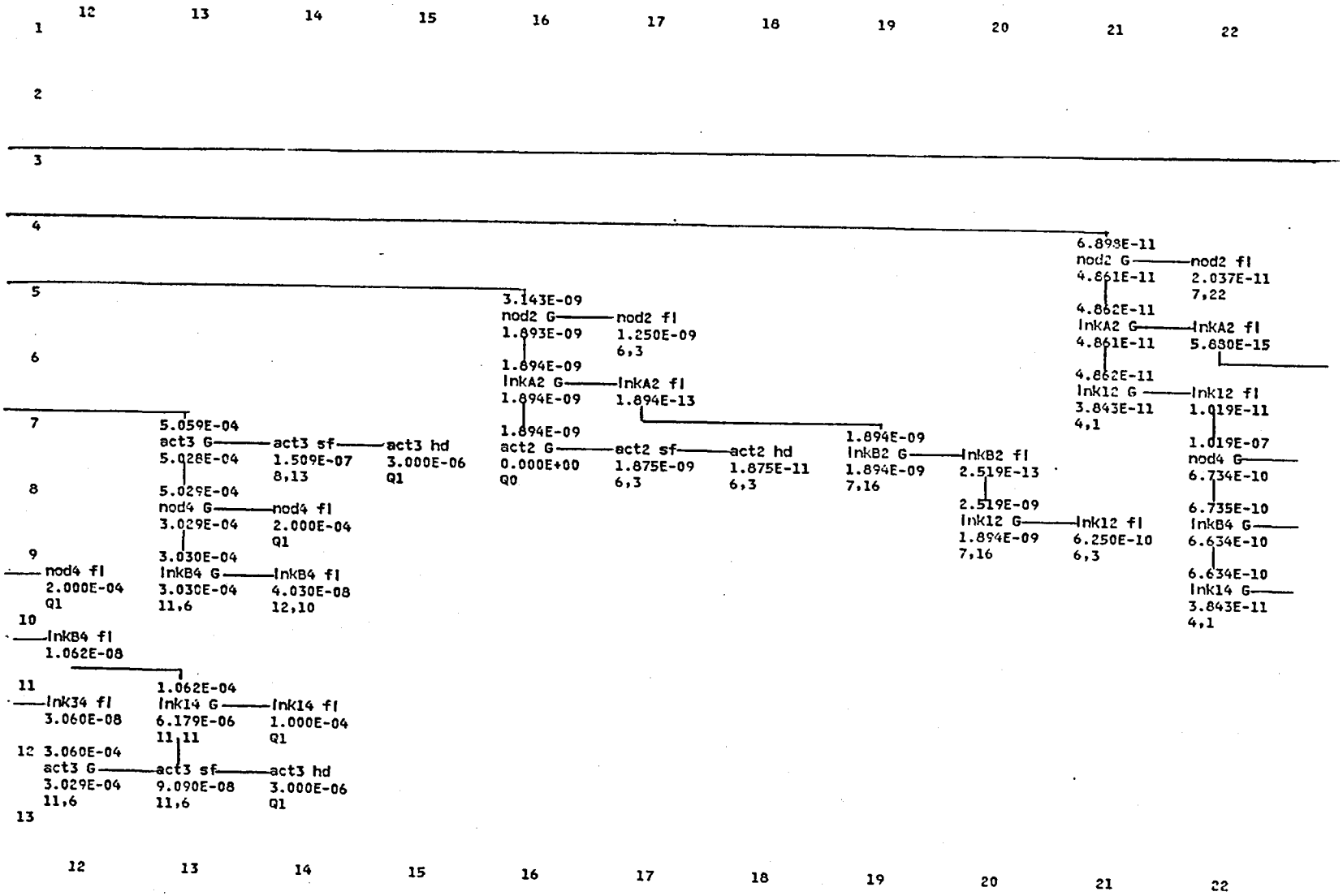


Figure 6.7-10 Cont'd

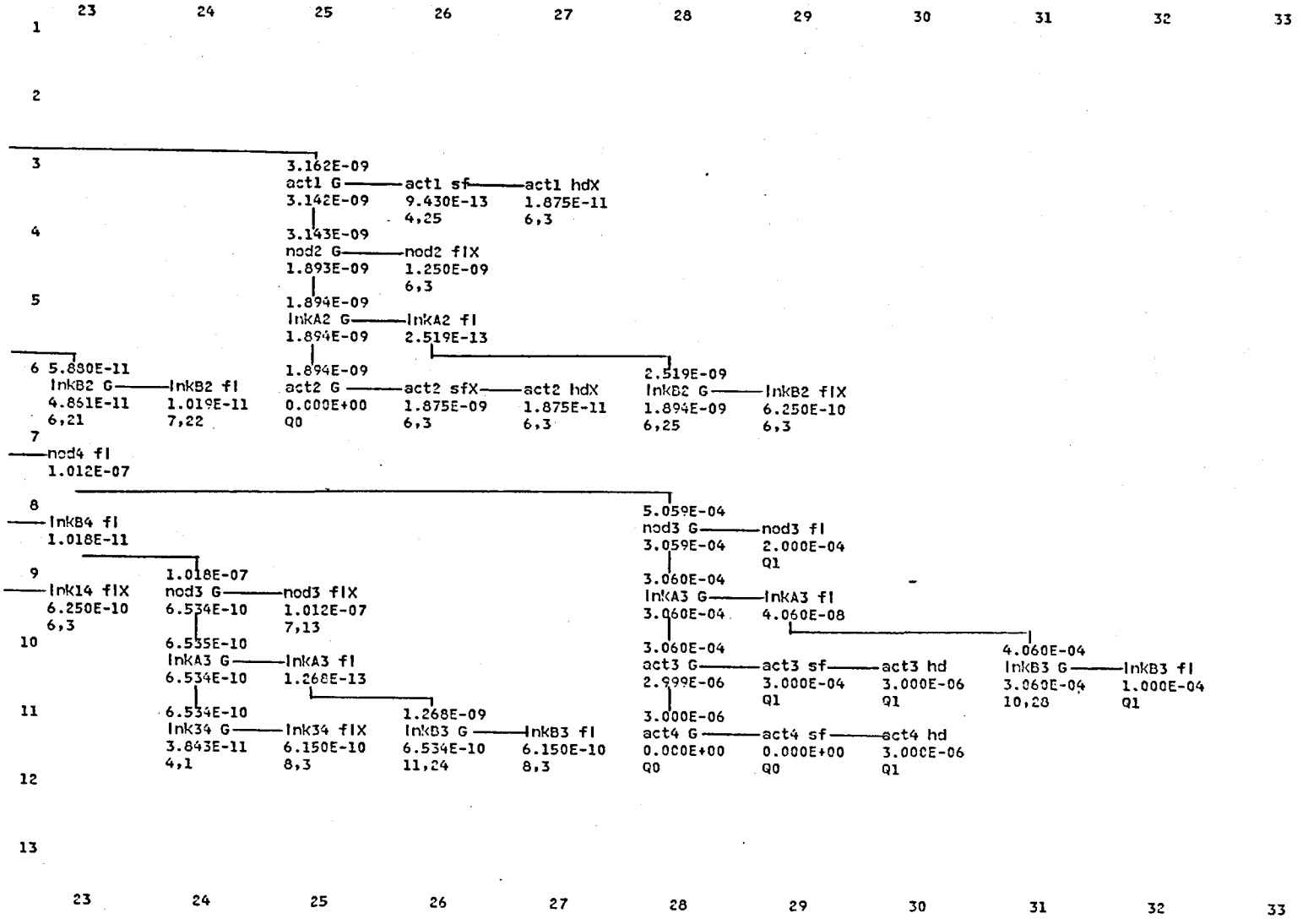


Figure 6.7-10 Concluded

The ability to make these approximations with a high degree of confidence illustrates one of the strengths of this analysis technique. This strength is the easy ability to judge quantitatively the significance of changes in the system configuration. The ability to make these judgements and use them to make simplifying assumptions can be illustrated by the process used to construct the upper left part of the equation diagram as shown in Figure 6.7-10.

The diagram in the upper left deals with the case where node 1 is failed. The first step in this analysis is to consider actuator 1. If it is good or failed soft, node 2, link 2, and actuator 2 are added. If they are good the system is good. If there is a failure of any of these elements, this dual actuator can not be used and the state of the other dual actuator must be considered to determine system unreliability. A very similar process was done earlier in the analysis as shown by the diagram starting at the point '6,3'. There is one difference in the state of the system, however. This earlier analysis was based on the fact that node 1 was good. The new analysis should be based on the fact that node 1 is failed. The significance of this difference can be determined by tracing through the diagram starting at the point '6,3' to see where the state of node 1 is significant. The only time node 1 was used is where communication links have failed and the link 1 to 4 is used. These points are at '12, 10' and '11,13'. With node 1 failed this link would be useless and a 'ql' should be entered instead. The significance of this change can be estimated by looking at the appropriate numbers. If a 'ql' had been entered at '11,10' the associated number would have been 1.0E-04, the failure rate of a link. The resulting (upper) number at point '11,9' would now be 4.03E-04 which feeds into the point '10,7' which becomes 4.03E-08. It can now be seen that this change has no significance since the resulting number at '10,6' would not change at all in the four significant numbers shown. A similar process can be used at the other point that link 1 to 4 is used which will also show that the loss of node 1 is insignificant. Thus it is possible to transfer the result from point '6,3' to the points where it is needed in the upper right of the diagram and avoid having to reproduce this whole diagram for the minor change that would not have changed the numbers.

The process described in the previous paragraphs has illustrated some of the strengths of this analysis technique. Other advantages can be seen by reviewing the numbers shown in Figure 6.7-10. It can be seen, starting with the first line, that very little of the final

unreliability of the system is contributed by system states that include the failure of node 1. The number added by the failure of node 1 is almost two orders of magnitude smaller. On the next line it can be seen that the contribution of link A to 1 failed is even less (almost four orders of magnitude less than the state which includes the failure of that link). On the other hand, lines three and four show that almost all of the system unreliability comes from failure modes that include hard failure of one or the other of the actuator channels. This analysis thus shows that the unreliability of this sample system is dominated by the actuators. The richness of interconnections in the communication structure makes it very reliable and thus an insignificant contributer to the system unreliability.

This analysis technique has several other features that have proven to be useful. The unreliability of the system as a function of time can be evaluated by entering different times in the element model. The sensitivity of the unreliability to change in the elements failure rates can easily be determined by modifying the model and rerunning the program using the stored data defining the diagram. Changes in the design of the system can be evaluated in a similar way. Many times the effects of changes in failure rates or the design can be estimated by manually tracing the appropriate paths in the diagram without needing to rerun the analysis.

REFERENCES

1. N.D. Murray, A.L. Hopkins, and J.H. Wensley, "Highly Reliable Multiprocessors" in AGARDograph #224, Integrity in Electronic Flight Control Systems, P. Kurzhals, Ed., AGARD-NATO Neuilly-Sur-Seine, France, April 1977.

2. J.H. Wensley, L. Lamport, J. Goldberg, M.W. Green, K.N. Levitt, P.M. Melliar-Smith, R.E. Shostak, and C.B. Weinstock, "SIFT: Design and Analysis of a Fault-Tolerant Computer for Aircraft Control," Proc. IEEE Vol. 66, No. 10, October, 1978, pp 1240-1255.

3. A.L. Hopkins, T.B. Smith, and J.H. Lala, "FTMP-A Highly Reliable Fault-Tolerant Multiprocessor for Aircraft," Proc. IEEE Vol. 66, No. 10, October, 1978, pp 1221-1239.

4. B. Vandecasteele, "French Utilization of the Digital Multiplex Bus," Proc. 2nd AFSC Multiplex DAta Bus Conference, 10-12 October 1978, Dayton, Ohio, USAF Aeronautical Systems Division, Volume II pp. 45-50.

5. T.B. Smith, "A damage- and Fault-Tolerant Input/Output Network," IEEE Trans. Computers, Vol. C-24, No. 5, May 1975.

6. H. Ernest, I. Mehdi and E. Reiquam, "YC-14 Electrical Power Systems - Unique Features and Problem Solutions," NAECON '77 Record p. 171.

7. J.R. Perkins and A.J. Marek, "Overview of the Advanced Aircraft Electrical System (AAES) A-7E Prototype Design," NAECON '77 Record p. 178.

8. L.D. Brock and H.A. Goodman, "Reliability Analysis of the F-8 Digital Fly-By-Wire System," The Charles Stark Draper Laboratory, Inc., Cambridge, MA, Report R-1324, Nov. 1979.

9. F. Fisher and J.A. Plummer, "Lightning Protection of Aircraft," NASA Reference Publication 1008, October 1977.