NOT MEASUREMENT SENSITIVE

MIL-HDBK-338B
1 October 1998
SUPERSEDING
MIL-HDBK-338A
12 October 1988

MILITARY HANDBOOK

ELECTRONIC RELIABILITY DESIGN HANDBOOK

This handbook is for guidance only. Do not cite this document as a requirement

AMSC N/A **AREA RELI**

FOREWORD

1. This handbook is approved for use by all Departments and Agencies of the Department of Defense (DoD). It was developed by the DoD with the assistance of the military departments, federal agencies, and industry and replaces in its entirety MIL-HDBK-338A. The handbook is written for reliability managers and engineers and provides guidance in developing and implementing a sound reliability program for all types of products.

2. This Handbook is for guidance only. This Handbook cannot be cited as a requirement. If it is, the contractor does not have to comply.

3. Reliability is a discipline that continues to increase in importance as systems become more complex, support costs increase, and defense budgets decrease. Reliability has been a recognized performance factor for at least 50 years. During World War II, the V-1 missile team, led by Dr. Wernher von Braun, developed what was probably the first reliability model. The model was based on a theory advanced by Eric Pieruschka that if the probability of survival of an element is 1/x, then the probability that a set of n identical elements will survive is (1/x)n . The formula derived from this theory is sometimes called Lusser's law (Robert Lusser is considered a pioneer of reliability) but is more frequently known as the formula for the reliability of a series system: Rs = R1 x R2 x . . x Rn.

4. Despite the long gestation period for reliability, achieving the high levels needed in military systems is too often an elusive goal. System complexity, competing performance requirements, the rush to incorporate promising but immature technologies, and the pressures of acquisition budget and schedule contribute to this elusiveness. In the commercial sector, high levels of reliability are also necessary. Recently, American products once shunned in favor of foreign alternatives have made or are making a comeback. This shift in consumer preferences is directly attributable to significant improvements in the reliability and quality of the American products.

5. Noting these improvements, and facing a shrinking defense budget, the Department of Defense began the process of changing its acquisition policies to buy more commercial off-the-shelf products and to use commercial specifications and standards. The objective is to capitalize on the "best practices" that American business has developed or adopted, primarily in response to foreign competitive pressures. When combined with the knowledge and expertise of military contractors in building complex and effective military systems (soundly demonstrated during the conflict with Iraq), it is hoped that these commercial practices will allow the Department of Defense to acquire world-class systems on time and within budget.

FOREWORD

6. The information in this Handbook reflects the move within the military to incorporate best commercial practices and the lessons learned over many years of acquiring weapon systems “by the book”. Military as well as commercial standards and handbooks are cited for reference because they are familiar to both military and commercial companies. Many of the military documents are being rescinded, so copies may be difficult to obtain. For those who have copies or can obtain them, the military documents provide a wealth of valuable information.

7. Beneficial comments (recommendations, additions, deletions) and any pertinent data which may be useful in improving this document should be addressed to: Air Force Research Laboratory/IFTB, 525 Brooks Road, Rome, NY 13441-4505. Comments should be submitted using the self-addressed Standardization Document Improvement Proposal (DD Form 1426) appearing at the end of this document or by letter.

TABLE OF CONTENTS

Section		Page
1.0	SCOPE	1-1
1.1	Introduction	1-1
1.2	Application	1-1
1.3	Organization	1-1
2 .0	REFERENCED DOCUMENTS	2-1
2.1	Government Documents	2-1
	2.1.1 Specifications, Standards and Handbooks	2-1
2.2	Other Referenced Documents	2-3
3.0	DEFINITIONS OF TERMS AND ACRONYMS AND ABBREVIATIONS	3-1
3.1	Introduction	3-1
3.2	Definitions	3-1
3.3	List of Abbreviations and Acronyms	3-21
4.0	GENERAL STATEMENTS	4-1
4.1	Introduction and Background	4-1
4.2	The System Engineering Process	4-2
	4.2.1 Systems Engineering and IPTs	4-3
	4.2.2 The Four Steps of Systems Engineering	4-3
4.3	System Effectiveness	4-7
	4.3.1 R/M Considerations in System Effectiveness	4-8
4.4	Factors Influencing System Effectiveness	4-8
	4.4.1 Equipment of New Design	4-8
	4.4.2 Interrelationships Among Various System Properties	4-9
4.5	Optimization of System Effectiveness	4-11
5 .0	RELIABILITY/MAINTAINABILITY/AVAILABILITY THEORY	5-1
5.1	Introduction	5-1
5.2	Reliability Theory	5-1
	5.2.1 Basic Concepts	5-2
5.3	Statistical Distributions Used in Reliability Models	5-8
	5.3.1 Continuous Distributions	5-8
	5.3.1.1 Normal (or Gaussian) Distribution	5-8
	5.3.2 Examples of Reliability Calculations Using the Normal Distribution	5-14
	5.3.2.1 Microwave Tube Example	5-14
	5.3.2.2 Mechanical Equipment Example	5-15
	5.3.3 Lognormal Distribution	5-16
	5.3.3.1 Fatigue Failure Example	5-17

TABLE OF CONTENTS

Section | Page
5.3.4 Exponential Distribution 5-17
5.3.4.1 Airborne Fire Control System Example 5-18
5.3.4.2 Computer Example 5-18
5.3.5 Gamma Distribution 5-19
5.3.5.1 Missile System Example 5-21
5.3.6 Weibull Distribution 5-22
5.3.6.1 Example of Use of Weibull Distribution 5-23
5.3.7 Discrete Distributions 5-24
5.3.7.1 Binomial Distribution 5-24
5.3.7.1.1 Quality Control Example 5-24
5.3.7.1.2 Reliability Example 5-25
5.3.8 Poisson Distribution 5-26
5.3.8.1 Example With Permissible Number of Failures 5-27
5.4 Failure Modeling 5-28
5.4.1 Typical Failure Rate Curve 5-28
5.4.2 Reliability Modeling of Simple Structures 5-30
5.4.2.1 Series Configuration 5-31
5.4.2.2 Parallel Configuration 5-32
5.4.2.3 K-Out-Of-N Configuration 5-35
5.5 Bayesian Statistics in Reliability Analysis 5-37
5.5.1 Bayes' Theorem 5-38
5.5.1.1 Bayes' Example (Discrete Distribution) 5-39
5.5.1.2 Bayes' Example (Continuous Distribution) 5-42
5.6 Maintainability Theory 5-44
5.6.1 Basic Concepts 5-45
5.6.2 Statistical Distributions Used in Maintainability Models 5-48
5.6.2.1 Lognormal Distribution 5-49
5.6.2.1.1 Ground Electronic System Maintainability Analysis Example 5-51
5.6.2.2 Normal Distribution 5-63
5.6.2.2.1 Equipment Example 5-65
5.6.2.3 Exponential Distribution 5-67
5.6.2.3.1 Computer Example 5-68
5.6.2.4 Exponential Approximation 5-70
5.7 Availability Theory 5-70
5.7.1 Basic Concepts 5-72
5.7.2 Availability Modeling (Markov Process Approach) 5-73
5.7.2.1 Single Unit Availability Analysis (Markov Process Approach) 5-75

TABLE OF CONTENTS

Section					Page
5.8	R&M Trade-Off Techniques				5-83
	5.8.1	Reliability vs Maintainability			5-83
5.9	References For Section 5				5-88
6 .0	RELIABILITY SPECIFICATION, ALLOCATION, MODELING AND PREDICTION				6-1
6.1	Introduction				6-1
6.2	Reliability Specification				6-1
	6.2.1	Methods of Specifying the Reliability Requirement			6-1
	6.2.2	Description of Environment and/or Use Conditions			6-3
	6.2.3	Time Measure or Mission Profile			6-5
	6.2.4	Clear Definition of Failure			6-6
	6.2.5	Description of Method(s) for Reliability Demonstration			6-7
6.3	Reliability Apportionment/Allocation				6-7
	6.3.1	Introduction			6-7
	6.3.2	Equal Apportionment Technique			6-10
	6.3.3	ARINC Apportionment Technique (Ref. [6])			6-11
	6.3.4	Feasibility-Of-Objectives Technique (Ref. [7])			6-13
	6.3.5	Minimization of Effort Algorithm			6-16
6.4	Reliability Modeling and Prediction				6-19
	6.4.1	Introduction			6-19
	6.4.2	General Procedure			6-21
		6.4.2.1	Item Definition		6-22
		6.4.2.2	Service Use Profile		6-22
		6.4.2.3	Reliability Block Diagrams		6-24
		6.4.2.4	Mathematical/Simulation Models		6-24
		6.4.2.5	Part Description		6-24
		6.4.2.6	Environmental Data		6-24
		6.4.2.7	Stress Analysis		6-24
		6.4.2.8	Failure Distributions		6-25
		6.4.2.9	Failure Rates		6-25
		6.4.2.10	Item Reliability		6-25
	6.4.3	Tailoring Reliability Models and Predictions			6-25
	6.4.4	Reliability Modeling			6-26
		6.4.4.1	Reliability Block Diagrams		6-26
		6.4.4.2	Reliability Modeling Methods		6-29
			6.4.4.2.1	Conventional Probability Modeling Method	6-29
			6.4.4.2.1.1	Series Model	6-29
			6.4.4.2.1.2	Parallel Models	6-30
			6.4.4.2.1.3	Series-Parallel Models	6-32
			6.4.4.2.2	Boolean Truth Table Modeling Method	6-33

TABLE OF CONTENTS

Section		Page
6.4.4.2.3	Logic Diagram Modeling Method	6-38
6.4.4.2.4	Complex System Modeling Methods	6-41
6.4.4.2.4.1	Markov Modeling (Ref. [9])	6-41
6.4.4.2.4.2	Monte Carlo Simulation Method	6-42
6.4.5	Reliability Prediction	6-44
6.4.5.1	General	6-46
6.4.5.2	Mathematical Models for Reliability Prediction	6-48
6.4.5.3	Reliability Prediction Methods	6-50
6.4.5.3.1	Similar Item Prediction Method	6-50
6.4.5.3.2	Parts Count Prediction Method	6-52
6.4.5.3.3	Parts Stress Analysis Prediction Method	6-54
6.4.5.3.3.1	Stress Analysis Techniques	6-57
6.4.5.3.3.2	Sample Calculation	6-59
6.4.5.3.3.3	Modification for Non-Exponential Failure Densities (General Case)	6-63
6.4.5.3.3.4	Nonoperating Failure Rates	6-66
6.4.5.3.4	Reliability Physics Analysis (Ref. [17] and [18])	6-68
6.4.5.4	Computer Aided Reliability Prediction	6-71
6.5	Step-By-Step Procedure for Performing Reliability Prediction and Allocation	6-71
6.6	References for Section 6	6-72
7.0	RELIABILITY ENGINEERING DESIGN GUIDELINES	7-1
7.1	Introduction	7-1
7.2	Parts Management	7-2
7.2.1	Establishing a Preferred Parts List (PPL)	7-3
7.2.2	Vendor and Device Selection	7-5
7.2.2.1	Critical Devices/Technology/Vendors	7-8
7.2.2.1.1	ASIC Devices	7-9
7.2.2.1.2	GaAs and MMIC Devices	7-9
7.2.2.2	Plastic Encapsulated Microcircuits (PEMs)	7-10
7.2.2.3	Hidden Hybrids	7-10
7.2.2.4	Device Specifications	7-11
7.2.2.5	Screening	7-12
7.2.2.6	Part Obsolescence and Diminishing Manufacturer Sources (DMS)	7-12
7.2.2.7	Failure Reporting, Analysis, And Corrective Action System (FRACAS)	7-15
7.2.3	Design for Reliability	7-15
7.2.3.1	Electronic Part Reliability Assessment / Life Analysis	7-16
7.2.4	Design for Manufacturability	7-19

TABLE OF CONTENTS

Section				Page
	7.2.5	Parts Management Plan Evaluation Criteria		7-20
		7.2.5.1	Quality Improvement Program	7-20
		7.2.5.2	Quality Assurance	7-20
			7.2.5.2.1 Part Qualification	7-21
			7.2.5.2.2 Production Quality Assurance	7-24
		7.2.5.3	Assembly Processes	7-26
		7.2.5.4	Design Criteria	7-28
7.3	Derating			7-30
	7.3.1	Electronic Part Derating		7-30
	7.3.2	Derating of Mechanical and Structural Components		7-32
7.4	Reliable Circuit Design			7-38
	7.4.1	Transient and Overstress Protection		7-38
		7.4.1.1	On-Chip Protection Networks	7-40
		7.4.1.2	Metal Oxide Varistors (MOVs)	7-42
		7.4.1.3	Protective Diodes	7-43
		7.4.1.4	Silicon Controlled Rectifier Protection	7-43
		7.4.1.5	Passive Component Protection	7-44
		7.4.1.6	Protective Devices Summary	7-47
		7.4.1.7	Protection Design For Parts, Assemblies and Equipment	7-48
		7.4.1.8	Printed Wiring Board Layout	7-49
		7.4.1.9	Shielding	7-50
		7.4.1.10	Grounding	7-52
		7.4.1.11	Protection With MOVs	7-54
		7.4.1.12	Protection With Diodes	7-57
	7.4.2	Parameter Degradation and Circuit Tolerance Analysis		7-62
	7.4.3	Computer Aided Circuit Analysis		7-70
		7.4.3.1	Advantages of Computer Aided Circuit Analysis/Simulation	7-71
		7.4.3.2	Limitations of Computer-Aided Circuit Analysis/Simulation Programs	7-71
		7.4.3.3	The Personal Computer (PC) as a Circuit Analysis Tool	7-71
	7.4.4	Fundamental Design Limitations		7-74
		7.4.4.1	The Voltage Gain Limitation	7-75
		7.4.4.2	Current Gain Limitation Considerations	7-78
		7.4.4.3	Thermal Factors	7-79
7.5	Fault Tolerant Design			7-80
	7.5.1	Redundancy Techniques		7-81
		7.5.1.1	Impact on Testability	7-81
	7.5.2	Reliability Role in the Fault Tolerant Design Process		7-84
		7.5.2.1	Fault Tolerant Design Analysis	7-86

TABLE OF CONTENTS

Section Page

7.5.3 Redundancy as a Design Technique 7-88
7.5.3.1 Levels of Redundancy 7-92
7.5.3.2 Probability Notation for Redundancy Computations 7-93
7.5.3.3 Redundancy Combinations 7-94
7.5.4 Redundancy in Time Dependent Situations 7-96
7.5.5 Redundancy Considerations in Design 7-98
7.5.5.1 Partial Redundancy 7-105
7.5.5.2 Operating Standby Redundancy 7-109
7.5.5.2.1 Two Parallel Elements 7-109
7.5.5.2.2 Three Parallel Elements 7-111
7.5.5.2.3 Voting Redundancy 7-112
7.5.5.3 Inactive Standby Redundancy 7-113
7.5.5.4 Dependent Failure Probabilities 7-117
7.5.5.5 Optimum Allocation of Redundancy 7-118
7.5.6 Reliability Analysis Using Markov Modeling 7-119
7.5.6.1 Introduction 7-119
7.5.6.2 Markov Theory 7-121
7.5.6.3 Development of the Markov Model Equation 7-123
7.5.6.4 Markov Model Reduction Techniques 7-125
7.5.6.5 Application of Coverage to Markov Modeling 7-127
7.5.6.6 Markov Conclusions 7-128
7.6 Environmental Design 7-128
7.6.1 Environmental Strength 7-128
7.6.2 Designing for the Environment 7-129
7.6.3 Temperature Protection 7-140
7.6.4 Shock and Vibration Protection 7-142
7.6.5 Moisture Protection 7-144
7.6.6 Sand and Dust Protection 7-145
7.6.7 Explosion Proofing 7-146
7.6.8 Electromagnetic Radiation Protection 7-147
7.6.9 Nuclear Radiation 7-149
7.6.10 Avionics Integrity Program (AVIP) 7-151
7.6.10.1 MIL-STD-1670: Environmental Criteria and Guidelines for Air Launched Weapons 7-153
7.7 Human Performance Reliability 7-159
7.7.1 Introduction 7-159
7.7.2 Reliability, Maintainability, and Availability Parameters for Human - Machine Systems 7-161
7.7.3 Allocating System Reliability to Human Elements 7-165
7.7.3.1 Qualitative Allocation 7-165
7.7.3.2 Quantitative Allocation 7-167

TABLE OF CONTENTS

Section		Page
7.7.4	Sources of Human Performance Reliability Data	7-169
7.7.5	Tools for Designing Man-Machine Systems	7-172
7.7.5.1	Task Analysis	7-173
7.7.5.2	General Design Tools	7-173
7.7.5.3	Computer-Based Design Tools	7-175
7.7.5.3.1	Parametric Design Tools	7-176
7.7.5.3.2	Interface Design Tools	7-176
7.7.5.3.3	Work Space Design Tools	7-176
7.7.6	Reliability Prediction for Human-Machine Systems	7-177
7.7.6.1	Probability Compounding	7-178
7.7.6.2	Stochastic Models	7-183
7.7.6.3	Digital Simulation	7-184
7.7.6.4	Expert Judgment Techniques	7-186
7.7.7	Verification of Human Performance Reliability	7-187
7.8	Failure Mode and Effects Analysis (FMEA)	7-187
7.8.1	Introduction	7-187
7.8.2	Phase 1	7-190
7.8.3	Phase 2	7-201
7.8.4	Example	7-203
7.8.5	Risk Priority Number	7-206
7.8.5.1	Instituting Corrective Action	7-209
7.8.6	Computer Aided FMEA	7-209
7.8.7	FMEA Summary	7-210
7.9	Fault Tree Analysis	7-210
7.9.1	Discussions of FTA Methods	7-221
7.10	Sneak Circuit Analysis (SCA)	7-222
7.10.1	Definition of Sneak Circuit	7-222
7.10.2	SCA: Definition and Traditional Techniques	7-223
7.10.3	New SCA Techniques	7-224
7.10.4	Examples of Categories of SNEAK Circuits	7-225
7.10.5	SCA Methodology	7-229
7.10.5.1	Network Tree Production	7-229
7.10.5.2	Topological Pattern Identification	7-229
7.10.5.3	Clue Application	7-231
7.10.6	Software Sneak Analysis	7-231
7.10.7	Integration of Hardware/Software Analysis	7-234
7.10.8	Summary	7-235
7.11	Design Reviews	7-236
7.11.1	Introduction and General Information	7-236
7.11.2	Informal Reliability Design Review	7-239
7.11.3	Formal Design Reviews	7-240

TABLE OF CONTENTS

Section		Page
7.11.4	Design Review Checklists	7-246
7.12	Design for Testability	7-250
7.12.1	Definition of Testability and Related Terms	7-251
7.12.2	Distinction between Testability and Diagnostics	7-251
7.12.3	Designing for Testability	7-251
7.12.4	Developing a Diagnostic Capability	7-255
7.12.5	Designing BIT	7-256
7.12.6	Testability Analysis	7-257
7.12.6.1	Dependency Analysis	7-258
7.12.6.1.1	Dependency Analysis Tools	7-260
7.12.6.2	Other Types of Testability Analyses	7-260
7.13	System Safety Program	7-262
7.13.1	Introduction	7-262
7.13.2	Definition of Safety Terms and Acronyms	7-267
7.13.3	Program Management and Control Elements	7-268
7.13.3.1	System Safety Program	7-268
7.13.3.2	System Safety Program Plan	7-268
7.13.3.3	Integration/Management of Associate Contractors, Subcontractors, and Architect and Engineering Firms	7-269
7.13.3.4	System Safety Program Reviews/Audits	7-269
7.13.3.5	System Safety Group/System Safety Working Group Support	7-269
7.13.3.6	Hazard Tracking and Risk Resolution	7-269
7.13.3.7	System Safety Progress Summary	7-269
7.13.4	Design and Integration Elements	7-269
7.13.4.1	Preliminary Hazard List	7-269
7.13.4.2	Preliminary Hazard Analysis	7-270
7.13.4.3	Safety Requirements/Criteria Analysis	7-270
7.13.4.4	Subsystem Hazard Analysis	7-270
7.13.4.5	System Hazard Analysis	7-270
7.13.4.6	Operating and Support Hazard Analysis	7-270
7.13.4.7	Occupational Health Hazard Assessment	7-270
7.13.5	Design Evaluation Elements	7-270
7.13.5.1	Safety Assessment	7-270
7.13.5.2	Test and Evaluation Safety	7-271
7.13.5.3	Safety Review of Engineering Change Proposals and Requests for Deviation/Waiver	7-271

TABLE OF CONTENTS

Section					Page
	7.13.6	Compliance and Verification			7-271
		7.13.6.1	Safety Verification		7-271
		7.13.6.2	Safety Compliance Assessment		7-271
		7.13.6.3	Explosive Hazard Classification and Characteristics Data		7-271
		7.13.6.4	Explosive Ordinance Disposal Source Data		7-271
	7.13.7	Tailoring Guidelines			7-272
7.14	Finite Element Analysis				7-272
	7.14.1	Introduction and General Information			7-272
	7.14.2	Finite Element Analysis Application			7-272
	7.14.3	Finite Element Analysis Procedure			7-276
	7.14.4	Applications			7-278
	7.14.5	Limitations			7-278
7.15	References for Section 7				7-279
8.0	RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION AND GROWTH				8-1
8.1	Introduction				8-1
8.2	Failure Reporting, Analysis, and Corrective Action System (FRACAS) and Failure Review Board (FRB)				8-2
	8.2.1	Failure Reporting, Analysis and Corrective Action System (FRACAS)			8-2
		8.2.1.1	Closed Loop Failure Reporting/Corrective Actions System		8-3
		8.2.1.2	Failure Reporting Systems		8-7
		8.2.1.3	Failure Reporting Forms		8-7
		8.2.1.4	Data Collection and Retention		8-7
	8.2.2	Failure Review Board			8-9
8.3	Reliability Data Analysis				8-10
	8.3.1	Graphical Methods			8-10
		8.3.1.1	Examples of Graphical Methods		8-13
	8.3.2	Statistical Analysis			8-21
		8.3.2.1	Introduction		8-21
		8.3.2.2	Treatment of Failure Data		8-22
		8.3.2.3	Reliability Function (Survival Curves)		8-29
			8.3.2.3.1	Computation of Theoretical Exponential Reliability Function	8-31
			8.3.2.3.2	Computation For Normal Reliability Function	8-33
		8.3.2.4	Censored Data		8-36
		8.3.2.5	Confidence Limits and Intervals		8-37
			8.3.2.5.1	Confidence Limits - Normal Distribution	8-39
			8.3.2.5.2	Confidence Limits - Exponential Distribution	8-43
			8.3.2.5.3	Confidence-Interval Estimates for the Binomial Distribution	8-50

TABLE OF CONTENTS

Section Page

8.3.2.6 Tests for Validity of the Assumption Of A Theoretical Reliability Parameter Distribution 8-52
8.3.2.6.1 Kolmogorov-Smirnov (K-S) Goodness-of-Fit Test (also called “d” test) 8-53
8.3.2.6.2 Chi-Square Goodness-of-Fit Test 8-60
8.3.2.6.3 Comparison of K-S and Chi-Square Goodness-of-Fit Tests 8-67
8.4 Reliability Demonstration 8-68
8.4.1 Introduction 8-68
8.4.2 Attributes and Variables 8-75
8.4.3 Fixed Sample and Sequential Tests 8-75
8.4.4 Determinants of Sample Size 8-75
8.4.5 Tests Designed Around Sample Size 8-76
8.4.6 Parameterization of Reliability 8-76
8.4.7 Instructions on the Use of Reliability Demonstration Test Plans 8-76
8.4.7.1 Attributes Demonstration Tests 8-77
8.4.7.1.1 Attributes Plans for Small Lots 8-77
8.4.7.1.2 Attributes Plans for Large Lots 8-81
8.4.7.2 Attributes Demonstration Test Plans for Large Lots, Using the Poisson Approximation Method 8-84
8.4.7.3 Attributes Sampling Using ANSI/ASQC Z1.4-1993 8-87
8.4.7.4 Sequential Binomial Test Plans 8-89
8.4.7.5 Variables Demonstration Tests 8-93
8.4.7.5.1 Time Truncated Demonstration Test Plans 8-93
8.4.7.5.1.1 Exponential Distribution (H-108) 8-93
8.4.7.5.1.2 Normal Distribution 8-95
8.4.7.5.1.3 Weibull Distribution (TR-3, TR-4, TR-6) 8-100
8.4.7.5.2 Failure Truncated Tests 8-103
8.4.7.5.2.1 Exponential Distribution (MIL-HDBK-H108) 8-103
8.4.7.5.2.2 Normal Distribution, σ Known 8-105
8.4.7.5.2.3 Normal Distribution, σ Unknown (MIL-STD-414) 8-110
8.4.7.5.2.4 Weibull Distribution 8-113
8.4.7.5.3 Sequential Tests 8-116
8.4.7.5.3.1 Exponential Distribution (MIL-HDBK-781) 8-116
8.4.7.5.3.2 Normal Distribution 8-119
8.4.7.6 Interference Demonstration Tests 8-123
8.4.7.7 Bayes Sequential Tests 8-127
8.4.8 Reliability Demonstration Summary 8-131

TABLE OF CONTENTS

Section				Page
8.5	Reliability Growth			8-132
	8.5.1	Reliability Growth Concept		8-132
	8.5.2	Reliability Growth Modeling		8-135
		8.5.2.1	Application Example	8-142
	8.5.3	Comparison of the Duane and AMSAA Growth Models		8-144
		8.5.3.1	Other Growth Models	8-147
	8.5.4	Reliability Growth Testing		8-147
		8.5.4.1	When Reliability Growth Testing is Performed	8-147
		8.5.4.2	Reliability Growth Approach	8-148
		8.5.4.3	Economics of Reliability Growth Testing	8-153
	8.5.5	Reliability Growth Management		8-154
		8.5.5.1	Management of the Reliability Growth Process	8-154
		8.5.5.2	Information Sources That Initiate Reliability Growth	8-156
		8.5.5.3	Relationships Among Growth Information Sources	8-157
8.6	Summary of the Differences Between Reliability Growth Testing and Reliability Demonstration Testing			8-159
8.7	Accelerated Testing			8-160
	8.7.1	Accelerated Life Testing		8-162
	8.7.2	Accelerated Stress Testing		8-162
	8.7.3	Equipment Level Accelerated Tests		8-162
	8.7.4	Component Level Accelerated Test		8-163
	8.7.5	Accelerated Test Models		8-163
		8.7.5.1	The Inverse Power Law Acceleration Model	8-164
		8.7.5.2	The Arrhenius Acceleration Model	8-165
		8.7.5.3	Miner's Rule - Fatigue Damage	8-167
	8.7.6	Advanced Concepts In Accelerated Testing		8-169
		8.7.6.1	Step Stress Profile Testing	8-170
		8.7.6.2	Progressive Stress Profile Testing	8-171
		8.7.6.3	HALT Testing	8-171
		8.7.6.4	HASS Testing	8-173
		8.7.6.5	HAST (Highly Accelerated Temperature and Humidity - Stress Test)	8-174
	8.7.7	Accelerated Testing Data Analysis and Corrective Action Caveats		8-174
8.8	References for Section 8			8-176
9.0	SOFTWARE RELIABILITY			9-1
9.1	Introduction			9-1
9.2	Software Issues			9-4

TABLE OF CONTENTS

Section			Page
9.3	Software Design		9-12
	9.3.1	Preliminary Design	9-12
		9.3.1.1 Develop the Architecture	9-13
		9.3.1.2 Physical Solutions	9-13
		9.3.1.3 External Characteristics	9-14
		9.3.1.4 System Functional Decomposition	9-15
	9.3.2	Detailed Design	9-15
		9.3.2.1 Design Examples	9-15
		9.3.2.2 Detailed Design Tools	9-16
		9.3.2.3 Software Design and Coding Techniques	9-16
9.4	Software Design and Development Process Model		9-17
	9.4.1	Ad Hoc Software Development	9-19
	9.4.2	Waterfall Model	9-19
	9.4.3	Classic Development Model	9-20
	9.4.4	Prototyping Approach	9-22
	9.4.5	Spiral Model	9-24
	9.4.6	Incremental Development Model	9-26
	9.4.7	Cleanroom Model	9-28
9.5	Software Reliability Prediction and Estimation Models		9-30
	9.5.1	Prediction Models	9-31
		9.5.1.1 In-house Historical Data Collection Model	9-31
		9.5.1.2 Musa's Execution Time Model	9-32
		9.5.1.3 Putnam's Model	9-33
		9.5.1.4 Rome Laboratory Prediction Model: RL-TR-92-52 (Ref. [16])	9-35
		9.5.1.5 Rome Laboratory Prediction Model: RL-TR-92-15 (Ref. [17])	9-38
	9.5.2	Estimation Models	9-40
		9.5.2.1 Exponential Distribution Models	9-40
		9.5.2.2 Weibull Distribution Model (Ref. [19])	9-46
		9.5.2.3 Bayesian Fault Rate Estimation Model	9-46
		9.5.2.4 Test Coverage Reliability Metrics	9-48
	9.5.3	Estimating Total Number of Faults Using Tagging	9-49
9.6	Software Reliability Allocation		9-51
	9.6.1	Equal Apportionment Applied to Sequential Software CSCIs	9-53
	9.6.2	Equal Apportionment Applied to Concurrent Software CSCIs	9-54
	9.6.3	Allocation Based on Operational Criticality Factors	9-54
	9.6.4	Allocation Based on Complexity Factors	9-56

TABLE OF CONTENTS

Section				Page
9.7	Software Testing			9-58
	9.7.1	Module Testing		9-58
	9.7.2	Integration Testing		9-59
	9.7.3	System Testing		9-61
	9.7.4	General Methodology for Software Failure Data Analysis		9-61
9.8	Software Analyses			9-62
	9.8.1	Failure Modes		9-64
	9.8.2	Failure Effects		9-64
	9.8.3	Failure Criticality		9-65
	9.8.4	Fault Tree Analysis		9-66
	9.8.5	Failure Modes and Effects Analysis		9-67
9.9	References			9-69
10.0	SYSTEMS RELIABILITY ENGINEERING			10-1
10.1	Introduction			10-1
	10.1.1	Commercial-Off-The-Shelf (COTS) and Nondevelopmental Item (NDI) Considerations		10-2
	10.1.2	COTS/NDI as the End Product		10-8
	10.1.3	COTS/NDI Integrated with Other Items		10-8
	10.1.4	Related COTS/NDI Issues		10-9
10.2	System Effectiveness Concepts			10-9
	10.2.1	The ARINC Concept of System Effectiveness (Ref. [1])		10-9
	10.2.2	The Air Force (WSEIAC) Concept (Ref. [2])		10-10
	10.2.3	The Navy Concept of System Effectiveness (Ref. [4])		10-14
	10.2.4	An Illustrative Model of a System Effectiveness Calculation		10-16
10.3	System R&M Parameters			10-20
	10.3.1	Parameter Translation Models		10-21
		10.3.1.1	Reliability Adjustment Factors	10-21
		10.3.1.2	Reliability Prediction of Dormant Products	10-24
	10.3.2	Operational Parameter Translation		10-25
		10.3.2.1	Parameter Definitions	10-27
		10.3.2.2	Equipment Operating Hour to Flight Hour Conversion	10-27
	10.3.3	Availability, Operational Readiness, Mission Reliability, and Dependability - Similarities and Differences		10-28
10.4	System, R&M Modeling Techniques			10-30
	10.4.1	Availability Models		10-33
		10.4.1.1	Model A - Single Unit System (Point Availability)	10-33
		10.4.1.2	Model B - Average or Interval Availability	10-38
		10.4.1.3	Model C - Series System with Repairable/Replaceable Units	10-40
		10.4.1.4	Model D - Redundant Systems	10-43

TABLE OF CONTENTS

Section Page

10.4.1.5 Model E - R&M Parameters Not Defined in Terms of Time 10-55
10.4.2 Mission Reliability and Dependability Models 10-58
10.4.3 Operational Readiness Models 10-60
10.4.3.1 Model A - Based Upon Probability of Failure During Previous Mission and Probability of Repair Before Next Mission Demand 10-61
10.4.3.2 Model B - Same As Model A Except Mission Duration Time, t is Probabilistic 10-63
10.4.3.3 Model C - Similar To Model A But Includes Checkout Equipment Detectability 10-64
10.4.3.4 Model D - For a Population of N Systems 10-66
10.5 Complex Models 10-73
10.6 Trade-off Techniques 10-74
10.6.1 General 10-74
10.6.2 Reliability - Availability - Maintainability Trade-offs 10-75
10.7 Allocation of Availability, Failure and Repair Rates 10-86
10.7.1 Availability Failure Rate and Repair Rate Allocation for Series Systems 10-87
10.7.1.1 Case (1) 10-87
10.7.1.2 Case (2) 10-88
10.7.2 Failure and Repair Rate Allocations For Parallel Redundant Systems 10-93
10.7.3 Allocation Under State-of-the-Art Constraints 10-99
10.8 System Reliability Specification, Prediction and Demonstration 10-100
10.8.1 Availability Demonstration Plans 10-100
10.8.1.1 Fixed Sample Size Plans 10-101
10.8.1.2 Fixed-Time Sample Plans 10-104
10.9 System Design Considerations 10-106
10.10 Cost Considerations 10-109
10.10.1 Life Cycle Cost (LCC) Concepts 10-109
10.11 References for Section 10 10-117

11.0 PRODUCTION AND USE (DEPLOYMENT) R&M 11-1
11.1 Introduction 11-1
11.2 Production Reliability Control 11-3
11.2.1 Quality Engineering (QE) and Quality Control (QC) 11-4
11.2.1.1 Quality System Requirements 11-6
11.2.1.1.1 ISO 9000 11-6
11.2.1.1.1.1 Comparing ISO 9000 to MIL-Q-9858 11-8
11.2.1.1.1.2 Why ISO 9000? 11-9
11.2.1.2 Quality Control 11-10

TABLE OF CONTENTS

Section Page

11.2.2 Production Reliability Degradation Assessment & Control 11-14
11.2.2.1 Factors Contributing to Reliability Degradation During Production: Infant Mortality 11-15
11.2.2.2 Process Reliability Analysis 11-19
11.2.3 Application of Environmental Stress Screening (ESS) During Production to Reduce Degradation and Promote Growth 11-26
11.2.3.1 Part Level Screening 11-28
11.2.3.2 Screening at Higher Levels of Assembly 11-30
11.2.3.3 Screen Test Planning and Effectiveness 11-32
11.2.3.3.1 Environmental Stress Screening per MIL-HDBK-344 11-32
11.2.3.3.2 Tri-Service ESS Guidelines 11-36
11.2.3.3.2.1 Types of Flaws to be Precipitated 11-37
11.2.3.3.2.2 Levels of Assembly at which ESS May be Performed 11-37
11.2.3.3.2.3 Types and Severities of Stresses 11-40
11.2.3.3.2.4 Failure Detection Measurements During Thermal Cycling and Random Vibration 11-41
11.2.3.3.2.5 Baseline ESS Profiles 11-41
11.2.3.3.2.6 Optimizing/Tailoring of ESS 11-44
11.2.4 Production Reliability Acceptance Testing (MIL-HDBK-781) 11-45
11.2.5 Data Collection and Analysis (During Production) 11-52
11.2.6 Monitor/Control of Subcontractors and Suppliers 11-54
11.2.6.1 Major Subcontractor and Manufacturer Monitoring 11-54
11.2.6.2 Establishing Vendor Capability and Program Reviews 11-54
11.2.6.3 Supplier Monitoring 11-55
11.3 Production Maintainability Control 11-55
11.4 Reliability and Quality During Shipment and Storage 11-55
11.4.1 Factors Contributing to Reliability Degradation During Shipment & Storage 11-56
11.4.2 Protection Methods 11-58
11.4.3 Shipment and Storage Degradation Control (Storage Serviceability Standards) 11-62
11.4.3.1 Application of Cyclic Inspection During Storage to Assure Reliability and Material Readiness 11-72
11.4.4 Data Collection and Analysis (During Storage) 11-72
11.5 Operational R&M Assessment and Improvement 11-74
11.5.1 Factors Contributing to R&M Degradation During Field Operation 11-75
11.5.2 Maintenance Degradation Control (During Depot Storage) 11-76
11.5.3 Maintenance Documentation Requirements 11-79
11.5.4 Data Collection and Analysis (During Field Deployment) 11-80

TABLE OF CONTENTS

Section			Page
	11.5.5	System R&M Assessment	11-82
	11.5.6	System R&M Improvement	11-85
11.6	References For Section 11		11-87
12.0	RELIABILITY MANAGEMENT CONSIDERATIONS		12-1
12.1	Impacts of Acquisition Reform		12-1
	12.1.1	Acquisition Reform History	12-1
		12.1.1.1 Performance-based Specifications	12-1
		12.1.1.2 Other Standardization Documents	12-3
		12.1.1.3 Overall Acquisition Policy and Procedures	12-4
		12.1.1.4 Impacts on Reliability Management	12-4
12.2	Reliability Program Management Issues		12-5
12.3	Reliability Specification Requirements		12-6
	12.3.1	Template for Preparing Reliability Section of Solicitation	12-7
	12.3.2	Guidance for Selecting Sources	12-15
12.4	Reliability Program Elements		12-17
12.5	Phasing of Reliability Program Activities		12-19
	12.5.1	Reliability Activities by Life Cycle Phase	12-20
		12.5.1.1 Phase 0 - Concept Exploration	12-22
		12.5.1.2 Phase I - Program Definition and Risk Reduction	12-22
		12.5.1.3 Phase II - Engineering and Manufacturing Development	12-23
		12.5.1.4 Phase III - Production, Deployment, and Operational Support	12-24
12.6	R&M Planning and Budgeting		12-25
	12.6.1	Conceptual Exploration Phase Planning	12-26
	12.6.2	Program Definition and Risk Reduction	12-26
	12.6.3	Engineering and Manufacturing Development (EMD) Phase Planning	12-27
	12.6.4	Production, Deployment, and Operational Support Phase Planning	12-28
12.7	Trade-offs		12-28
	12.7.1	Concept Exploration Phase Trade-off Studies	12-29
	12.7.2	Program Definition and Risk Reduction Phase Trade-off Studies	12-30
	12.7.3	Trade-offs During Engineering Manufacturing Development (EMD), Production, Deployment and Operational Support Phases	12-31
12.8	Other Considerations		12-32
	12.8.1	Software Reliability	12-32
		12.8.1.1 Requirements Definition	12-35
		12.8.1.2 System Analysis	12-35
		12.8.1.3 Package Design	12-37
		12.8.1.4 Unit Design, Code and Debug	12-37
		12.8.1.5 Module Integration and Test	12-37
		12.8.1.6 System Integration and Test	12-38

TABLE OF CONTENTS

Section			Page
		12.8.1.7 Acceptance Test	12-38
		12.8.1.8 Program Plan	12-38
		12.8.1.9 Specifications	12-38
		12.8.1.10 Data System	12-39
		12.8.1.11 Program Review	12-39
		12.8.1.12 Test Plan	12-40
		12.8.1.13 Technical Manuals	12-40
	12.8.2	Cost Factors and Guidelines	12-40
		12.8.2.1 Design-To-Cost Procedures	12-43
		12.8.2.2 Life Cycle Cost (LCC) Concepts	12-45
	12.8.3	Product Performance Agreements	12-45
		12.8.3.1 Types of Product Performance Agreements	12-47
		12.8.3.2 Warranty/Guarantee Plans	12-51
	12.8.4	Reliability Program Requirements, Evaluation and Surveillance	12-53
		12.8.4.1 Reliability Program Requirements Based Upon the Type of Procurement	12-53
		12.8.4.2 Reliability Program Evaluation and Surveillance	12-55
12.9	References for Section 12		12-56

LIST OF FIGURES

		Page
FIGURE 3-1:	INTERVALS OF TIME	3-19
FIGURE 4.2-1:	SYSTEM MANAGEMENT ACTIVITIES	4-4
FIGURE 4.2-2:	FUNDAMENTAL SYSTEM PROCESS CYCLE	4-6
FIGURE 4.5-1:	FLOW DIAGRAM FOR A GENERAL OPTIMIZATION PROCESS	4-12
FIGURE 5.2-1:	SUMMARY OF BASIC RELIABILITY CONCEPTS	5-7
FIGURE 5.3-1:	SHAPES OF FAILURE DENSITY, RELIABILITY AND HAZARD RATE FUNCTIONS FOR COMMONLY USED CONTINUOUS DISTRIBUTIONS	5-9
FIGURE 5.3-2:	SHAPES OF FAILURE DENSITY AND RELIABILITY FUNCTIONS OF COMMONLY USED DISCRETE DISTRIBUTIONS	5-10
FIGURE 5.3-3:	FIVE CHANNEL RECEIVER WITH TWO FAILURES ALLOWED	5-25
FIGURE 5.4-1:	HAZARD RATE AS A FUNCTION OF AGE	5-28
FIGURE 5.4-2:	STABILIZATION OF FAILURE FREQUENCY	5-30
FIGURE 5.4-3:	SERIES CONFIGURATION	5-31
FIGURE 5.4-4:	PARALLEL CONFIGURATION	5-33
FIGURE 5.4-5:	COMBINED CONFIGURATION NETWORK	5-33
FIGURE 5.5-1:	SIMPLE PRIOR DISTRIBUTION	5-40
FIGURE 5.5-2:	SIMPLE POSTERIOR DISTRIBUTION	5-41
FIGURE 5.5-3:	TREE DIAGRAM EXAMPLE	5-42
FIGURE 5.6-1:	BASIC METHODS OF MAINTAINABILITY MEASUREMENT	5-47
FIGURE 5.6-2:	EXAMPLE MAINTAINABILITY FUNCTION DERIVED FROM TIME-TO-REPAIR DISTRIBUTION	5-47
FIGURE 5.6-3:	PLOT OF THE LOGNORMAL OF THE TIMES-TO-RESTORE DATA GIVEN IN TABLE 5.6-5 IN TERMS OF THE STRAIGHT t'S	5-56
FIGURE 5.6-4:	PLOT OF THE LOGNORMAL PDF OF THE TIMES-TO-RESTORE DATA GIVEN IN TABLE 5.6-5 IN TERMS OF THE LOGARITHMS OF T, OR *ln* t'	5-58
FIGURE 5.6-5:	PLOT OF THE MAINTAINABILITY FUNCTION FOR THE TIMES-TO-REPAIR DATA OF EXAMPLE 2	5-61
FIGURE 5.6-6:	EXPONENTIAL APPROXIMATION OF LOGNORMAL MAINTAINABILITY FUNCTIONS	5-71
FIGURE 5.7-1:	THE RELATIONSHIP BETWEEN INSTANTANEOUS, MISSION, AND STEADY STATE AVAILABILITIES AS A FUNCTION OF OPERATING TIME	5-74
FIGURE 5.7-2:	MARKOV GRAPH FOR SINGLE UNIT	5-75
FIGURE 5.7-3:	SINGLE UNIT AVAILABILITY WITH REPAIR	5-81
FIGURE 5.8-1:	BLOCK DIAGRAM OF A SERIES SYSTEM	5-84
FIGURE 5.8-2	RELIABILITY-MAINTAINABILITY TRADE-OFFS	5-87

LIST OF FIGURES

		Page
FIGURE 6.2-1:	SATISFACTORY PERFORMANCE LIMITS FOR EXAMPLE RADAR	6-4
FIGURE 6.2-2:	TEMPERATURE PROFILE	6-5
FIGURE 6.2-3:	TYPICAL OPERATIONAL SEQUENCE FOR AIRBORNE FIRE CONTROL SYSTEM	6-6
FIGURE 6.2-4:	EXAMPLE DEFINITION OF RELIABILITY DESIGN REQUIREMENTS IN A SYSTEM SPECIFICATION FOR (1) AVIONICS, (2) MISSILE SYSTEM AND (3) AIRCRAFT	6-8
FIGURE 6.4-1:	SERVICE USE EVENTS IN THE LOGISTIC AND OPERATIONAL CYCLES	6-23
FIGURE 6.4-2:	PROGRESSIVE EXPANSION OF RELIABILITY BLOCK DIAGRAM AS DESIGN DETAIL BECOMES KNOWN	6-27
FIGURE 6.4-3:	RADAR SYSTEM HIERARCHY (PARTIAL LISTING)	6-45
FIGURE 6.4-4:	SAMPLE RELIABILITY CALCULATION	6-56
FIGURE 7.2-1:	VENDOR SELECTION METHODOLOGIES	7-6
FIGURE 7.2-2:	PART OBSOLESCENCE AND DMS PROCESS FLOW	7-14
FIGURE 7.2-3:	REDUCED SCREEN FLOW	7-25
FIGURE 7.3-1:	STRESS-STRENGTH DISTRIBUTIONS AND UNRELIABILITY IN DESIGN	7-35
FIGURE 7.3-2:	NORMAL (GAUSSIAN) STRESS-STRENGTH DISTRIBUTIONS AND UNRELIABILITY IN DESIGN	7-36
FIGURE 7.3-3:	FACTORS AFFECTING UNRELIABILITY	7-37
FIGURE 7.4-1:	ON-CHIP DIODE PROTECTION CIRCUIT	7-41
FIGURE 7.4-2:	(A) FOUR-LAYER STRUCTURE OF AN SCR (B) CURRENT - VOLTAGE CHARACTERISTIC	7-44
FIGURE 7.4-3:	GROUNDING PRACTICE AT A SINGLE PHASE SERVICE ENTRANCE	7-52
FIGURE 7.4-4:	CIRCUIT SUBSYSTEMS WITH GROUND CONNECTIONS “DAISY-CHAINED” INVITES PROBLEMS	7-53
FIGURE 7.4-5:	GROUND TRACES RETURNED TO A COMMON POINT	7-54
FIGURE 7.4-6:	DIODE PROTECTION OF A BIPOLAR TRANSISTOR	7-58
FIGURE 7.4-7:	DIODE PROTECTION FOR A DISCRETE MOSFET TRANSISTOR	7-58
FIGURE 7.4-8:	DIODE PROTECTION FOR SILICON CONTROLLED RECTIFIERS	7-59
FIGURE 7.4-9:	TRANSIENT PROTECTION FOR A TTL CIRCUIT USING DIODES	7-59
FIGURE 7.4-10:	TRANSIENT PROTECTION FOR A CMOS CIRCUIT	7-60
FIGURE 7.4-11:	INPUT PROTECTION FOR POWER SUPPLIES	7-60
FIGURE 7.4-12:	PROTECTION OF DATA LINES OR POWER BUSES USING A DIODE ARRAY	7-61

LIST OF FIGURES

		Page
FIGURE 7.4-13:	FUSE PROTECTION FOR A TRANSIENT VOLTAGE SUPPRESSOR DIODE	7-62
FIGURE 7.4-14:	RESISTOR PARAMETER VARIATION WITH TIME (TYPICAL)	7-64
FIGURE 7.4-15:	CAPACITOR PARAMETER VARIATION WITH TIME (TYPICAL)	7-65
FIGURE 7.4-16:	RESISTOR PARAMETER CHANGE WITH STRESS AND TIME (TYPICAL)	7-66
FIGURE 7.4-17:	OUTPUT VOLTAGE VERSUS TRANSISTOR GAIN BASED ON A FIGURE APPEARING IN TAGUCHI TECHNIQUES FOR QUALITY ENGINEERING (REFERENCE [21])	7-69
FIGURE 7.4-18:	RATIO OF I_{CO} OVER TEMPERATURE T TO I_{CO} AT T = 25°C	7-79
FIGURE 7.5-1:	HARDWARE REDUNDANCY TECHNIQUES	7-82
FIGURE 7.5-2:	EFFECT OF MAINTENANCE CONCEPT ON LEVEL OF FAULT TOLERANCE	7-85
FIGURE 7.5-3:	PARALLEL NETWORK	7-88
FIGURE 7.5-4:	SIMPLE PARALLEL REDUNDANCY: SUMMARY	7-91
FIGURE 7.5-5:	SERIES-PARALLEL REDUNDANCY NETWORK	7-92
FIGURE 7.5-6:	RELIABILITY BLOCK DIAGRAM DEPICTING REDUNDANCY AT THE SYSTEM, SUBSYSTEM, AND COMPONENT LEVELS	7-93
FIGURE 7.5-7:	SERIES-PARALLEL CONFIGURATION	7-94
FIGURE 7.5-8:	PARALLEL-SERIES CONFIGURATION	7-95
FIGURE 7.5-9:	DECREASING GAIN IN RELIABILITY AS NUMBER OF ACTIVE ELEMENTS INCREASES	7-103
FIGURE 7.5-10:	RELIABILITY GAIN FOR REPAIR OF SIMPLE PARALLEL ELEMENT AT FAILURE	7-104
FIGURE 7.5-11:	PARTIAL REDUNDANT ARRAY	7-106
FIGURE 7.5-12:	RELIABILITY FUNCTIONS FOR PARTIAL REDUNDANT ARRAY OF FIGURE 7.5-11	7-108
FIGURE 7.5-13:	REDUNDANCY WITH SWITCHING	7-109
FIGURE 7.5-14:	THREE-ELEMENT REDUNDANT CONFIGURATIONS WITH SWITCHING	7-111
FIGURE 7.5-15:	THREE-ELEMENT VOTING REDUNDANCY	7-112
FIGURE 7.5-16:	MAJORITY VOTING REDUNDANCY	7-115
FIGURE 7.5-17:	SYSTEM RELIABILITY FOR N STANDBY REDUNDANT ELEMENTS	7-116
FIGURE 7.5-18:	LOAD SHARING REDUNDANT CONFIGURATION	7-117
FIGURE 7.5-19:	SUCCESS COMBINATIONS IN TWO-ELEMENT LOAD-SHARING CASE	7-118
FIGURE 7.5-20:	POSSIBLE REDUNDANT CONFIGURATIONS RESULTING FROM ALLOCATION STUDY	7-120
FIGURE 7.5-21:	MARKOV MODELING PROCESS	7-122

LIST OF FIGURES

		Page
FIGURE 7.5-22:	MARKOV FLOW DIAGRAM	7-124
FIGURE 7.5-23:	TWO CHANNEL EXAMPLE	7-126
FIGURE 7.5-24:	COVERAGE EXAMPLE	7-127
FIGURE 7.6-1:	EFFECTS OF COMBINED ENVIRONMENTS	7-130
FIGURE 7.7-1:	THE HUMAN IN SYSTEM RELIABILITY AND MAINTAINABILITY [44]	7-162
FIGURE 7.7-2:	THE COGNITIVE HUMAN MODEL	7-163
FIGURE 7.7-3:	FACTORS THAT AFFECT HUMAN FUNCTION RELIABILITY	7-163
FIGURE 7.7-4:	ZONES OF HUMAN PERFORMANCE FOR LONGITUDINAL VIBRATION (ADAPTED FROM MIL-STD-1472)	7-164
FIGURE 7.7-5:	HIERARCHICAL STRUCTURE OF FUNCTIONAL ANALYSIS (EXAMPLE)	7-166
FIGURE 7.7-6:	SIMPLIFIED DYNAMIC PROGRAMMING	7-170
FIGURE 7.7-7:	TOOLS FOR DESIGNING HUMAN-MACHINE SYSTEMS	7-172
FIGURE 7.7-8:	GOAL-SUCCESS TREE	7-175
FIGURE 7.7-9:	CATEGORIES OF HUMAN PERFORMANCE RELIABILITY PREDICTION METHODS	7-177
FIGURE 7.7-10:	THERP PROBABILITY TREE [62]	7-180
FIGURE 7.8-1:	TYPICAL SYSTEM SYMBOLIC LOGIC BLOCK DIAGRAM	7-191
FIGURE 7.8-2:	TYPICAL UNIT SYMBOLIC LOGIC BLOCK DIAGRAM	7-192
FIGURE 7.8-3:	FAILURE EFFECTS ANALYSIS FORM	7-200
FIGURE 7.8-4:	SYMBOLIC LOGIC DIAGRAM OF RADAR EXAMPLE	7-203
FIGURE 7.8-5:	DETERMINATION OF PREAMPLIFIER CRITICALITY	7-205
FIGURE 7.9-1:	FAULT TREE ANALYSIS SYMBOLS	7-213
FIGURE 7.9-2:	TRANSFORMATION OF TWO-ELEMENT SERIES RELIABILITY BLOCK DIAGRAM TO “FAULT TREE” LOGIC DIAGRAMS	7-214
FIGURE 7.9-3:	TRANSFORMATION OF SERIES/PARALLEL BLOCK DIAGRAM TO EQUIVALENT FAULT TREE LOGIC DIAGRAM	7-215
FIGURE 7.9-4:	RELIABILITY BLOCK DIAGRAM OF HYPOTHETICAL ROCKET MOTOR FIRING CIRCUIT	7-216
FIGURE 7.9-5:	FAULT TREE FOR SIMPLIFIED ROCKET MOTOR FIRING CIRCUIT	7-217
FIGURE 7.10-1:	AUTOMOTIVE SNEAK CIRCUIT	7-223
FIGURE 7.10-2:	SNEAK PATH ENABLE	7-226
FIGURE 7.10-3:	REDUNDANT CIRCUIT SWITCHED GROUND	7-226
FIGURE 7.10-4:	EXAMPLES OF CATEGORIES OF SNEAK CIRCUITS	7-228
FIGURE 7.10-5:	BASIC TOPOGRAPHS	7-230
FIGURE 7.10-6:	SOFTWARE TOPOGRAPHS	7-232
FIGURE 7.10-7:	SOFTWARE SNEAK EXAMPLE	7-234
FIGURE 7.11-1:	DESIGN REVIEW AS A CHECK VALVE IN THE SYSTEM ENGINEERING CYCLE	7-237

LIST OF FIGURES

		Page
FIGURE 7.11-2:	BASIC STEPS IN THE PRELIMINARY DESIGN REVIEW (PDR) CYCLE	7-242
FIGURE 7.11-3:	DESIGN RELIABILITY TASKS FOR THE PDR	7-243
FIGURE 7.11-4:	BASIC STEPS IN THE CDR CYCLE	7-244
FIGURE 7.11-5:	DESIGN RELIABILITY TASKS FOR THE CRITICAL DESIGN REVIEW (CDR)	7-245
FIGURE 7.11-6:	TYPICAL AREAS TO BE COVERED IN A DESIGN REVIEW	7-246
FIGURE 7.11-7:	TYPICAL QUESTIONS CHECKLIST FOR THE DESIGN REVIEW	7-249
FIGURE 7.12-1:	SIMPLE SYSTEM SHOWING TEST DEPENDENCIES	7-258
FIGURE 7.12-2:	REDUNDANCY BIT (SOURCE: RADC-TR-89-209, VOL. II)	7-261
FIGURE 7.12-3:	WRAP-AROUND BIT (SOURCE: RADC-TR-89-209, VOL II)	7-261
FIGURE 7.14-1:	NODAL ANALYSIS	7-276
FIGURE 7.14-2:	DISPLACEMENT/STRESS INTERPRETATION	7-277
FIGURE 7.14-3:	DETERMINISTIC ANALYSIS	7-277
FIGURE 7.14-4:	LIFETIME ESTIMATE	7-278
FIGURE 8.2-1:	CLOSED LOOP FAILURE REPORTING AND CORRECTIVE ACTION SYSTEM	8-4
FIGURE 8.2-2:	EXAMPLE OF FAILURE REPORT FORM	8-8
FIGURE 8.2-3:	CLOSED LOOP FAILURE REPORTING AND CORRECTIVE ACTION SYSTEM WITH FAILURE REVIEW BOARD	8-9
FIGURE 8.3-1:	GRAPHICAL POINT ESTIMATION FOR THE NORMAL DISTRIBUTION	8-14
FIGURE 8.3-2:	GRAPHICAL POINT ESTIMATION FOR THE WEIBULL DISTRIBUTION	8-20
FIGURE 8.3-3:	DISTRIBUTION GRAPHICAL EVALUATION	8-21
FIGURE 8.3-4:	HAZARD AND DENSITY FUNCTIONS FOR TABLE 8.3-3	8-25
FIGURE 8.3-5:	RELIABILITY FUNCTIONS FOR THE EXAMPLE GIVEN IN TABLE 8.3-4	8-28
FIGURE 8.3-6:	NORMAL DISTRIBUTION OF FAILURE IN TIME	8-30
FIGURE 8.3-7:	CALCULATION AND PRESENTATION OF A NORMAL SURVIVAL CURVE	8-30
FIGURE 8.3-8:	EXPONENTIAL DISTRIBUTION OF FAILURES IN TIME	8-30
FIGURE 8.3-9:	CALCULATION AND PRESENTATION OF AN EXPONENTIAL CURVE	8-30
FIGURE 8.3-10:	OBSERVED AND THEORETICAL EXPONENTIAL SURVIVAL CURVES	8-32
FIGURE 8.3-11:	OBSERVED AND THEORETICAL NORMAL SURVIVAL CURVES	8-32
FIGURE 8.3-12:	ACTUAL RELIABILITY FUNCTION AND THEORETICAL EXPONENTIAL RELIABILITY FUNCTION	8-34

LIST OF FIGURES

		Page
FIGURE 8.3-13:	NON-PARAMETRIC AND THEORETICAL NORMAL RELIABILITY FUNCTIONS	8-36
FIGURE 8.3-14:	GEOMETRICAL INTERPRETATION OF THE CONCEPT OF A CONFIDENCE INTERVAL	8-39
FIGURE 8.3-15:	TWO-SIDED CONFIDENCE INTERVAL AND LIMITS	8-41
FIGURE 8.3-16:	MULTIPLICATION RATIOS FOR DETERMINING UPPER AND LOWER CONFIDENCE LIMITS VS. NUMBER OF FAILURES FOR TEST TRUNCATED AT A FIXED TIME	8-49
FIGURE 8.3-17:	CHART FOR 95% CONFIDENCE LIMITS ON THE PROBABILITY S/N	8-51
FIGURE 8.3-18:	EXAMPLE OF THE APPLICATION OF THE "d" TEST	8-57
FIGURE 8.3-19:	FUEL SYSTEM FAILURE TIMES	8-62
FIGURE 8.3-20:	COMPUTATION	8-63
FIGURE 8.4-1:	NORMAL DISTRIBUTION	8-69
FIGURE 8.4-2A:	HYPOTHESIS TEST A	8-70
FIGURE 8.4-2B:	HYPOTHESIS TEST B	8-70
FIGURE 8.4-3A:	IDEAL OPERATING CHARACTERISTIC (OC) CURVE	8-71
FIGURE 8.4-3B:	TYPICAL OPERATING CHARACTERISTIC CURVE	8-71
FIGURE 8.4-4:	ACTUAL OPERATING CHARACTERISTIC CURVE	8-72
FIGURE 8.4-5:	OC CURVE CHARACTERISTICS	8-73
FIGURE 8.4-6:	GRAPHICAL SOLUTION OF SEQUENTIAL BINOMIAL TEST	8-92
FIGURE 8.5-1:	RELIABILITY GROWTH PROCESS	8-134
FIGURE 8.5-2:	RELIABILITY GROWTH PLOTS	8-136
FIGURE 8.5-3:	UP-IS-GOOD DUANE CHART WITH PLOT OF CURRENT MTBF	8-138
FIGURE 8.5-4:	FAILURE RATE VS. DEVELOPMENT TIME FOR WEIBULL FAILURE RATE	8-141
FIGURE 8.5-5:	FAILURE RATE VS. DEVELOPMENT TEST TIME FOR WEIBULL FAILURE RATE	8-144
FIGURE 8.5-6:	RELIABILITY GROWTH ANALYSIS (AMSAA MODEL)	8-146
FIGURE 8.5-7:	RELIABILITY GROWTH PLOTS	8-150
FIGURE 8.5-8:	COMPARISON OF CUMULATIVE LIFE CYCLE COSTS WITH AND WITHOUT SPECIFIED RELIABILITY GROWTH TEST REQUIREMENTS	8-153
FIGURE 8.5-9:	RELIABILITY GROWTH MANAGEMENT MODEL (ASSESSMENT)	8-155
FIGURE 8.5-10:	EXAMPLE OF A RELIABILITY GROWTH CURVE	8-156
FIGURE 8.5-11:	INFORMATION SOURCES THAT INITIATE RELIABILITY GROWTH	8-157
FIGURE 8.6-1:	RELIABILITY TESTING OPTIONS	8-160
FIGURE 8.7-1:	ARRHENIUS ACCELERATION MODEL	8-167

LIST OF FIGURES

		Page
FIGURE 8.7-2:	STEP STRESS PROFILE	8-170
FIGURE 8.7-3:	PROGRESSIVE STRESS PROFILE	8-171
FIGURE 9.1-1:	SOFTWARE ENVIRONMENT TIMELINE	9-2
FIGURE 9.1-2:	HARDWARE/SOFTWARE SYSTEM LIFE CYCLE RELATIONSHIP (REF. [2])	9-4
FIGURE 9.2-1:	BATHTUB CURVE FOR HARDWARE RELIABILITY	9-9
FIGURE 9.2-2:	REVISED BATHTUB CURVE FOR SOFTWARE RELIABILITY	9-11
FIGURE 9.3-1:	HIGH-LEVEL SOFTWARE ARCHITECTURE EXAMPLE	9-14
FIGURE 9.4-1:	WATERFALL MODEL (REF. [6])	9-20
FIGURE 9.4-2:	THE CLASSIC DEVELOPMENT MODEL (REF. [7])	9-21
FIGURE 9.4-3:	STEPS IN THE PROTOTYPING APPROACH	9-23
FIGURE 9.4-4:	SPIRAL MODEL (REF. [7])	9-25
FIGURE 9.4-5:	INCREMENTAL DEVELOPMENT MODEL (REF. [7])	9-27
FIGURE 9.4-6:	THE CLEANROOM DEVELOPMENT PROCESS (REF. [10])	9-29
FIGURE 9.5-1:	EXPECTED PROPORTION OF THE TOTAL NUMBER OF DEFECTS	9-35
FIGURE 9.5-2:	EXPONENTIAL MODEL BASIS	9-41
FIGURE 9.6-1:	RELIABILITY ALLOCATION PROCESS (REF. [2])	9-52
FIGURE 9.7-1:	STRUCTURAL REPRESENTATION OF A SOFTWARE SYSTEM	9-60
FIGURE 9.7-2:	FLOWCHART FOR SOFTWARE FAILURE DATA ANALYSIS AND DECISION-MAKING	9-63
FIGURE 9.8-1:	EXAMPLE OF SOFTWARE FMECA	9-68
FIGURE 10.1-1:	THE COMMERCIAL/NDI DECISION PROCESS	10-7
FIGURE 10.2-1:	SYSTEM EFFECTIVENESS MODELS	10-15
FIGURE 10.3-1:	PART DATABASE DISTRIBUTION	10-22
FIGURE 10.4-1:	PRINCIPAL STEPS REQUIRED FOR EVALUATION OF SYSTEM EFFECTIVENESS	10-32
FIGURE 10.4-2:	THE AVAILABILITY OF A SINGLE UNIT	10-35
FIGURE 10.4-3:	AVERAGE AND POINTWISE AVAILABILITY	10-39
FIGURE 10.4-4:	BLOCK DIAGRAM OF A SERIES SYSTEM	10-42
FIGURE 10.4-5:	HYPOTHETICAL HISTORY OF MACHINE GUN USAGE	10-56
FIGURE 10.4-6:	RENEWAL PROCESS IN TERMS OF ROUNDS FIRED	10-57
FIGURE 10.4-7:	OPERATIONAL READINESS PROBABILITY VERSUS QUEUING FACTOR ρ. FOR POPULATION SIZE N = 15; NUMBER OF REPAIR CHANNELS k	10-72
FIGURE 10.4-8:	OPERATIONAL READINESS PROBABILITY VERSUS QUEUING FACTOR ρ. FOR POPULATION SIZE N = 20; NUMBER OF REPAIR CHANNELS k	10-73
FIGURE 10.6-1:	RELIABILITY - MAINTAINABILITY - AVAILABILITY RELATIONSHIPS	10-77

LIST OF FIGURES

Page

FIGURE 10.6-2: AVAILABILITY AS A FUNCTION OF λ/μ 10-78
FIGURE 10.6-3: AVAILABILITY AS A FUNCTION OF MTBF AND 1/MTTR 10-78
FIGURE 10.6-4: AVAILABILITY NOMOGRAPH 10-79
FIGURE 10.6-5: RELIABILITY-MAINTAINABILITY TRADE-OFFS 10-82
FIGURE 10.6-6: BLOCK DIAGRAM OF A SERIES SYSTEM 10-84
FIGURE 10.7-1: PERMISSIBLE EQUIPMENT FAILURE AND REPAIR RATES FOR λ/μ = .25 10-97
FIGURE 10.7-2: UNAVAILABILITY CURVES 10-98
FIGURE 10.10-1: LCC CATEGORIES VS. LIFE CYCLE 10-111
FIGURE 10.10-2: R&M AND COST METHODS 10-114
FIGURE 10.10-3: LIFE CYCLE COSTS VS. RELIABILITY 10-116
FIGURE 11.1-1: RELIABILITY LIFE CYCLE DEGRADATION & GROWTH CONTROL 11-2
FIGURE 11.2-1: QUALITY ENGINEERING AND CONTROL OVER TIME 11-5
FIGURE 11.2-2: ISO 9000 FAMILY OF STANDARDS 11-7
FIGURE 11.2-3: LIFE CHARACTERISTIC CURVE 11-16
FIGURE 11.2-4: IMPACT OF DESIGN AND PRODUCTION ACTIVITIES ON EQUIPMENT RELIABILITY 11-18
FIGURE 11.2-5: “STEP” MTBF APPROXIMATION 11-19
FIGURE 11.2-6: MTBF (OUTGOING FROM PRODUCTION) ESTIMATING PROCESS 11-23
FIGURE 11.2-7: SAMPLE PROCESS FLOW DIAGRAM 11-24
FIGURE 11.2-8: A TYPICAL PRODUCTION PROCESS, FINDING DEFECTS AT THE LOWEST LEVEL OF MANUFACTURE IS THE MOST COST-EFFECTIVE 11-28
FIGURE 11.2-9: APPLICATION OF SCREENING WITHIN THE MANUFACTURING PROCESS 11-29
FIGURE 11.2-10: EFFECTIVENESS OF ENVIRONMENTAL SCREENS 11-31
FIGURE 11.2-11: MIL-HDBK-344 ESS PROCESS 11-35
FIGURE 11.2-12: SAMPLE ENVIRONMENTAL TEST CYCLE 11-49
FIGURE 11.2-13: REJECT-ACCEPT CRITERIA FOR TEST PLAN XVIIIC 11-50
FIGURE 11.4-1: PROTECTIVE CONTROL DURING SHIPMENT AND STORAGE . 11-60
FIGURE 11.4-2: TECHNICAL APPROACH TO STORAGE SERVICEABILITY STANDARDS (SSS)... 11-64
FIGURE 11.4-3: STORAGE SERVICEABILITY STANDARD PREPARATION PROCESS.. 11-68
FIGURE 11.4-4: DETERIORATION CLASSIFICATION OF MATERIAL 11-69
FIGURE 11.4-5: INSPECTION FREQUENCY MATRIX 11-71
FIGURE 11.4-6: CODED QUALITY INSPECTION LEVELS 11-73
FIGURE 12.3-1: CHECKLIST FOR EVALUATING RELIABILITY PORTION OF A PROPOSAL 12-16

LIST OF FIGURES

		Page
FIGURE 12.5-1:	LIFE CYCLE PHASES OF A PRODUCT	12-21
FIGURE 12.8-1:	CONCURRENT SYSTEM DEVELOPMENT PROCESS FOR BOTH HARDWARE AND SOFTWARE (REF. [6])	12-33
FIGURE 12.8-2:	SOFTWARE RELIABILITY PROGRAM ELEMENTS BY PROGRAM PHASE	12-36
FIGURE 12.8-3:	BALANCED DESIGN APPROACH	12-41
FIGURE 12.8-4:	EXPENDITURES DURING LIFE CYCLE	12-42
FIGURE 12.8-5:	EFFECT OF EARLY DECISION ON LIFE CYCLE COST	12-42
FIGURE 12.8-6:	LIFE CYCLE COST ACTIVITIES	12-46

LIST OF TABLES

		Page
TABLE 4.5-1:	PARTIAL LIST OF OPTIMIZATION TECHNIQUES	4-13
TABLE 5.3-1:	VALUES OF THE STANDARD NORMAL DISTRIBUTION FUNCTION	5-12
TABLE 5.3-2:	ORDINATES F(z) OF THE STANDARD NORMAL CURVE AT z ..	5-13
TABLE 5.3-3:	GAMMA FUNCTION $\Gamma(n)$	5-20
TABLE 5.6-1:	COMPARISON OF BASIC RELIABILITY AND MAINTAINABILITY FUNCTIONS	5-46
TABLE 5.6-2:	VALUES OF ϕ OR $Z(T'_{(1-\alpha)})$ MOST COMMONLY USED IN MAINTAINABILITY ANALYSIS	5-51
TABLE 5.6-3:	TIME-TO-REPAIR DATA ON A GROUND ELECTRONIC SYSTEM	5-52
TABLE 5.6-4:	CALCULATIONS TO DETERMINE $\overline{t'}$ AND σ_T FOR THE DATA IN TABLE 5.6-3	5-54
TABLE 5.6-5:	THE PROBABILITY DENSITY OF TIME-TO-REPAIR DATA (FROM TABLE 5.6.2.1.1-1 BASED ON THE STRAIGHT TIMES TO REPAIR AND THE NATURAL LOGARITHM OF THE TIMES TO REPAIR USED TO PLOT FIGURES 5.6-3 AND 5.6-4, RESPECTIVELY.*)	5-57
TABLE 5.6-6:	VALUES OF ϕ FOR SPECIFIED α	5-65
TABLE 5.6-7:	VALUES OF k_E FOR SPECIFIED α	5-68
TABLE 5.7-1:	THE AVAILABILITY OF A SINGLE SYSTEM OR UNIT	5-82
TABLE 6.3-1:	MECHANICAL-ELECTRICAL SYSTEM	6-16
TABLE 6.4-1:	USES OF RELIABILITY MODELS AND PREDICTIONS	6-21
TABLE 6.4-2:	TRUTH TABLE CALCULATION FOR THE SYSTEM RELIABILITY DIAGRAM	6-35
TABLE 6.4-3:	REDUCTION TABULATION	6-37
TABLE 6.4-4:	LOGIC DIAGRAM EXAMPLES	6-39
TABLE 6.4-5:	PROS AND CONS OF PHYSICS-OF-FAILURE PREDICTION MODELS	6-46
TABLE 6.4-6:	ENVIRONMENTAL SYMBOL IDENTIFICATION AND DESCRIPTION	6-47
TABLE 6.4-7:	RELIABILITY ANALYSIS SIMILAR ITEM	6-52
TABLE 6.4-8:	GENERIC FAILURE RATE - λ_G (FAILURES PER 10^6 HOURS) FOR DISCRETE SEMICONDUCTORS........................	6-55
TABLE 6.4-9:	DISCRETE SEMICONDUCTOR QUALITY FACTORS - π_Q	6-56
TABLE 6.4-10:	MAJOR INFLUENCE FACTORS ON PART RELIABILITY	6-57
TABLE 6.4-11:	FORMULAS FOR CALCULATING MICROCIRCUIT RELIABILITY	6-58
TABLE 6.4-12:	BIPOLAR COMPLEXITY FAILURE RATE C1	6-60
TABLE 6.4-13:	ENVIRONMENTAL FACTOR - π_E	6-61

LIST OF TABLES

		Page
TABLE 6.4-14:	QUALITY FACTORS - π_Q	6-61
TABLE 6.4-15:	BASIC APPROACH TO RELIABILITY PHYSICS ANALYSIS	6-69
TABLE 6.4-16:	EXAMPLE OF A PINION RELIABILITY ANALYSIS	6-70
TABLE 7.2-1:	QUESTIONS FOR PART SUPPLIERS	7-7
TABLE 7.2-2:	HIDDEN HYBRID CHECKLIST	7-11
TABLE 7.2-3:	GENERIC PART APPLICATION FACTORS	7-17
TABLE 7.3-1:	PRINCIPLE RELIABILITY DEPENDENT STRESS FACTORS/DERATING FACTORS	7-31
TABLE 7.3-2:	DERATING VALUES FOR TRANSISTORS	7-32
TABLE 7.4-1:	COMPARISON OF PROTECTION DEVICES	7-48
TABLE 7.4-2:	0.5µS - 100 KHZ RING WAVE	7-56
TABLE 7.4-3:	8/20µS, 1.2/50µS COMBINATION WAVE	7-57
TABLE 7.4-4:	COMPARISON OF VARIABILITY ANALYSIS METHODS	7-68
TABLE 7.5-1:	DIAGNOSTIC IMPLICATIONS OF FAULT TOLERANT DESIGN APPROACHES	7-83
TABLE 7.5-2:	REDUNDANCY TECHNIQUES	7-100
TABLE 7.5-3:	RELIABILITY CALCULATIONS FOR EXAMPLE 2	7-107
TABLE 7.6-1:	ENVIRONMENTAL COVERAGE CHECKLIST (TYPICAL)	7-129
TABLE 7.6-2:	VARIOUS ENVIRONMENTAL PAIRS	7-131
TABLE 7.6-3:	ENVIRONMENTAL EFFECTS	7-135
TABLE 7.6-4:	LOW TEMPERATURE PROTECTION METHODS	7-142
TABLE 7.6-5:	ENVIRONMENTAL STRESSES IMPROVEMENT TECHNIQUES IN ELECTRONIC EQUIPMENT	7-150
TABLE 7.6-6:	SYSTEM USE CONDITIONS CHECKLIST (TYPICAL)	7-154
TABLE 7.6-7:	ENVIRONMENTAL ANALYSIS (INDUCED ENVIRONMENT)	7-156
TABLE 7.6-8:	ASSOCIATION OF FACTOR IMPORTANCE WITH REGION OF ENVIRONMENT	7-158
TABLE 7.7-1:	COMPARISON BETWEEN HARDWARE AND HUMAN RELIABILITY [39]	7-160
TABLE 7.7-2:	HUMAN-MACHINE COMPARATIVE CAPABILITIES	7-167
TABLE 7.7-3:	DATA BANKS AND THEIR AFFILIATIONS [55]	7-171
TABLE 7.7-4:	DATA CATEGORIES OF NATIONAL DATA BANKS [55]	7-172
TABLE 7.7-5:	MAPPS SCOPE	7-185
TABLE 7.8-1:	FAILURE MODE DISTRIBUTION OF PARTS	7-193
TABLE 7.8-2:	COLUMN DESCRIPTIONS FOR FIGURE 7.8-3	7-200
TABLE 7.8-3:	SEVERITY CLASSIFICATION	7-207
TABLE 7.8-4:	OCCURRENCE RANKING	7-207
TABLE 7.8-5:	DETECTION RANKING	7-209
TABLE 7.11-1:	DESIGN REVIEW GROUP, RESPONSIBILITIES AND MEMBERSHIP SCHEDULE	7-241
TABLE 7.11-2:	RELIABILITY ACTIONS CHECKLIST	7-247

LIST OF TABLES

		Page
TABLE 7.12-1:	RISKS AND CONSEQUENCES OF NOT MAKING BIT PART OF PRODUCT DESIGN	7-257
TABLE 7.12-2:	FIRST ORDER DEPENDENCY MODEL FOR SIMPLE SYSTEM	7-258
TABLE 7.12-3:	INHERENT TESTABILITY CHECKLIST	7-263
TABLE 7.13-1:	APPLICATION MATRIX FOR SYSTEM PROGRAM DEVELOPMENT	7-273
TABLE 7.13-2:	APPLICATION MATRIX FOR FACILITIES ACQUISITION	7-274
TABLE 8.3-1:	DATA ON TIMES TO FAILURE OF 20 ITEMS	8-12
TABLE 8.3-2:	MEDIAN RANKS	8-15
TABLE 8.3-3:	FAILURE DATA FOR TEN HYPOTHETICAL ELECTRONIC COMPONENTS	8-23
TABLE 8.3-4:	COMPUTATION OF DATA FAILURE DENSITY AND DATA HAZARD RATE	8-24
TABLE 8.3-5:	FAILURE DATA FOR 1,000 B-52 AIRCRAFT	8-26
TABLE 8.3-6:	TIME-TO-FAILURE DATA FOR S = 1000 MISSION HOURS	8-27
TABLE 8.3-7:	COMPUTATION OF THEORETICAL EXPONENTIAL RELIABILITY FUNCTION FOR MTBF = 1546 HOURS	8-34
TABLE 8.3-8:	OBSERVED FAILURE DATA	8-35
TABLE 8.3-9:	CONFIDENCE LIMITS - NORMAL DISTRIBUTION	8-40
TABLE 8.3-10:	CONFIDENCE INTERVAL	8-42
TABLE 8.3-11:	DISTRIBUTION OF χ^2 (CHI-SQUARE)	8-44
TABLE 8.3-12:	FACTORS FOR CALCULATION OF MEAN LIFE CONFIDENCE INTERVALS FROM TEST DATA (FACTORS = $2/\chi^2_{P,D}$)	8-48
TABLE 8.3-13:	CRITICAL VALUES $d_{\alpha;n}$ OF THE MAXIMUM ABSOLUTE DIFFERENCE BETWEEN SAMPLE AND POPULATION RELIABILITY FUNCTIONS	8-54
TABLE 8.7-1:	ACTIVATION ENERGIES ASSOCIATED WITH VARIOUS SILICON SEMICONDUCTOR FAILURE MECHANISMS	8-166
TABLE 9.2-1:	ASSESSING THE ORGANIZATIONAL COMMUNICATIONS GAP	9-7
TABLE 9.2-2:	SUMMARY: LIFE CYCLE DIFFERENCES	9-12
TABLE 9.3-1:	SOFTWARE DESIGN TECHNIQUES	9-17
TABLE 9.3-2:	SOFTWARE CODING TECHNIQUES	9-17
TABLE 9.4-1:	SOFTWARE DEVELOPMENT PROCESS SELECTION	9-18
TABLE 9.4-2:	CLEANROOM PERFORMANCE MEASURES (REF. [11])	9-30
TABLE 9.5-1:	COMPARING PREDICTION AND ESTIMATION MODELS	9-31
TABLE 9.5-2:	SOFTWARE RELIABILITY PREDICTION TECHNIQUES	9-32
TABLE 9.5-3:	TERMS IN MUSA’S EXECUTION TIME MODEL	9-33
TABLE 9.5-4:	PUTNAM’S TIME AXIS MILESTONES	9-34
TABLE 9.5-5:	RL-TR-92-52 TERMINOLOGY	9-36

LIST OF TABLES

		Page
TABLE 9.5-6:	AMOUNT OF HISTORICAL DATA INCLUDED	9-36
TABLE 9.5-7:	SUMMARY OF THE RL-TR-92-52 MODEL	9-37
TABLE 9.5-8:	REGRESSION EQUATION COEFFICIENTS	9-39
TABLE 9.5-9:	NOTATIONS FOR THE EXPONENTIAL DISTRIBUTION MODEL	9-41
TABLE 9.5-10:	VARIOUS EXPONENTIAL MODELS	9-42
TABLE 9.6-1:	SOFTWARE RELIABILITY ALLOCATION TECHNIQUES (REF. [2])	9-52
TABLE 9.6-2:	SOFTWARE FUNCTIONS BY SYSTEM MODE - EXAMPLE	9-51
TABLE 9.6-3:	COMPLEXITY PROCEDURES	9-56
TABLE 9.8-1:	HARDWARE FAILURE SEVERITY LEVELS (REF. [26])	9-65
TABLE 9.8-2:	SOFTWARE FAILURE SEVERITY LEVELS (REF. [5])	9-66
TABLE 9.8-3:	SOFTWARE FAILURE MODES AND CRITICALITY ANALYSIS CATEGORIES	9-67
TABLE 10.1-1:	CONCEPT OF SYSTEM EFFECTIVENESS	10-1
TABLE 10.1-2:	ADVANTAGES AND DISADVANTAGES OF COTS/NDI	10-5
TABLE 10.1-3:	R&M ACTIVITIES FOR NEW DEVELOPMENT ITEMS AND FOR COTS	10-6
TABLE 10.3-1:	SYSTEM R&M PARAMETERS	10-20
TABLE 10.3-2:	PART QUALITY FACTORS (MULTIPLY SERIES MTBF BY)	10-22
TABLE 10.3-3:	ENVIRONMENTAL CONVERSION FACTORS (MULTIPLY SERIES MTBF BY)	10-23
TABLE 10.3-4:	TEMPERATURE CONVERSION FACTORS (MULTIPLY SERIES MTBF BY)	10-24
TABLE 10.3-5:	AIRCRAFT RECEIVER CONVERSION: AIRBORNE OPERATING TO GROUND DORMANT FAILURE RATE (EXAMPLE)	10-25
TABLE 10.3-6:	RELIABILITY TRANSLATION MODELS	10-26
TABLE 10.3-7:	DEFINITIONS OF KEY R&M SYSTEM PARAMETERS	10-29
TABLE 10.4-1:	AVAILABILITY OF SOME REDUNDANT SYSTEMS BASED ON EXPONENTIAL FAILURE AND REPAIR DISTRIBUTIONS	10-48
TABLE 10.6-1:	ALTERNATIVE DESIGN TRADE-OFF CONFIGURATIONS	10-83
TABLE 10.6-2:	COST COMPARISON OF ALTERNATIVE DESIGN CONFIGURATIONS	10-83
TABLE 10.7-1:	PRELIMINARY SYSTEM AND SUBSYSTEM RELIABILITY SPECIFICATIONS	10-95
TABLE 10.10-1:	LIFE CYCLE COST BREAKDOWN	10-115
TABLE 11.2-1:	MIL-Q-9858 QUALITY PROGRAM ELEMENTS	11-9
TABLE 11.2-2:	QUALITY ENGINEERING TASKS	11-12
TABLE 11.2-3:	FOUR TYPES OF FAILURES	11-15

LIST OF TABLES

		Page
TABLE 11.2-4:	SCREENING ENVIRONMENTS VERSUS TYPICAL FAILURE MECHANICS	11-37
TABLE 11.2-5:	RISKS AND RESULTS OF ESS AT VARIOUS LEVELS	11-39
TABLE 11.2-6:	BASELINE VIBRATION PROFILE	11-42
TABLE 11.2-7:	BASELINE THERMAL CYCLE PROFILE	11-43
TABLE 11.2-8:	TEST CONDITIONS MATRIX (TAKEN FROM MIL-HDBK-781)	11-48
TABLE 11.4-1:	FAILURE MODES ENCOUNTERED WITH ELECTRONIC COMPONENTS DURING STORAGE	11-59
TABLE 11.4-2:	STORAGE-INDUCED QUALITY DEFECTS	11-65
TABLE 11.5-1:	DEPOT MAINTENANCE REQUIREMENT AREAS	11-79
TABLE 12.4-1:	COMMON RELIABILITY PROGRAM ELEMENTS	12-18
TABLE 12.5-1:	RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE CONCEPT EXPLORATION PHASE	12-22
TABLE 12.5-2.	RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE PROGRAM DEFINITION AND RISK REDUCTION PHASE	12-23
TABLE 12.5-3:	RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE ENGINEERING AND MANUFACTURING DEVELOPMENT PHASE	12-24
TABLE 12.5-4:	RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE PRODUCTION, DEPLOYMENT, AND OPERATIONAL SUPPORT PHASE	12-25
TABLE 12.8-1:	TYPES OF DESIGN-TO-COST PROGRAMS	12-44
TABLE 12.8-2:	FEATURES OF CURRENT WARRANTY-GUARANTEE PLANS	12-52

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

1.0 SCOPE

1.1 Introduction

This Handbook provides procuring activities and development contractors with an understanding of the concepts, principles, and methodologies covering all aspects of electronic systems reliability engineering and cost analysis as they relate to the design, acquisition, and deployment of DoD equipment/systems.

1.2 Application

This Handbook is intended for use by both contractor and government personnel during the conceptual, validation, full scale development, production phases of an equipment/system life cycle.

1.3 Organization

The Handbook is organized as follows:

SECTION 2	Referenced Documents
SECTION 3	Definitions
SECTION 4	General Statements
SECTION 5	Reliability/Maintainability/Availability Theory
SECTION 6	Reliability Specification, Allocation and Prediction
SECTION 7	Reliability Engineering Design Guidelines
SECTION 8	Reliability Data Collection and Analysis, Demonstration and Growth
SECTION 9	Software Reliability
SECTION 10	Systems Reliability Engineering
SECTION 11	Production and Use (Deployment) R&M
SECTION 12	R&M Management Considerations

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

2 .0 REFERENCED DOCUMENTS

The documents cited in this section are for guidance and information.

2.1 Government Documents

2.1.1 Specifications, Standards and Handbooks

The following specifications, standards, and handbooks form a part of this document to the extent specified herein. Unless otherwise specified, the issues of these documents are those listed in the issue of the Department of Defense Index of Specifications and Standards (DODISS) and applicable supplement thereto.

SPECIFICATIONS

Military

MIL-E-4158	General Specification For Ground Electronic Equipment
MIL-E-5400	General Specifications For Aerospace Electronic Equipment
MIL-E-16400	General Specification For Naval Ship and Shore: Electronic, Interior Communication and Navigation Equipment
MIL-E-17555	Packaging of Electronic and Electrical Equipment, Accessories, and Provisioned Items (Repair Parts)
MIL-M-28787	General Specification For Standard Electronic Modules
MIL-H-38534	General Specification For Hybrid Microcircuits
MIL-I-38535	General Specification For Manufacturing Integrated Circuits
MIL-H-46855	Human Engineering Requirements For Military Systems, Equipment and Facilities
MIL-PRF-19500K	General Specification For Semiconductor Devices
MIL-PRF-3853C	General Specification For Microcircuits
MIL-S-52779	Software Quality Assurance Program Requirements

SECTION 2: REFERENCED DOCUMENTS

STANDARDS

Military

MIL-STD-210	Climatic Extremes For Military Equipment
MIL-STD-414	Sampling Procedures and Tables For Inspection by Variables For Percent
MIL-STD-701	Lists of Standard Semiconductor Devices
MIL-STD-721	Definitions of Terms For Reliability, and Maintainability
MIL-STD-750	Tests Methods For Semiconductor Devices
MIL-STD-756	Reliability Modeling and Prediction
MIL-STD-790	Reliability Assurance Program For Electronic Part Specifications
MIL-STD-810	Environmental Test Methods and Engineering Guidelines
MIL-STD-882	System Safety Program Requirements
MIL-STD-883	Test Methods and Procedures For Microelectronics
MIL-STD-975	Standard Parts Derating Guidelines
MIL-STD-1472	Human Engineering Design Criteria For Military Systems, Equipment and Facilities
MIL-STD-1562	Lists of Standard Microcircuits
MIL-STD-1670	Environmental Criteria and Guidelines for Air Launched Weapons
MIL-STD-1686	Electrostatic Discharge Control Program For Protection of Electrical and Electronic Parts, Assemblies and Equipment (Excluding Electrically Initiated Explosive Devices)
MIL-STD-1772	Certification Requirements For Hybrid Microcircuit Facility and Lines
MIL-STD-2155	Failure Reporting, Analysis and Corrective Action System
MIL-STD-2167	Defense System Software Development

HANDBOOKS

Military

MIL-HDBK-454	Standard General Requirements For Electronic Equipment
MIL-HDBK-470	Maintainability Program Requirements For Systems and Equipment
MIL-HDBK-471	Maintainability Verification/Demonstration/Evaluation
MIL-HDBK-781	Reliability Testing For Engineering Development, Qualification and Production
MIL-HDBK-965	Parts Control Program
MIL-HDBK-1547	Technical Requirements For Parts, Materials, and Processes for Space and Launch Vehicles
MIL-HDBK-2084	General Requirements For Maintainability
MIL-HDBK-2164	Environmental Stress Screening Process For Electronic Equipment
MIL-HDBK-2165	Testability Program For Electronic Systems and Equipment

Unless otherwise indicated, copies of federal and military specification, standards, handbooks and bulletins are available from:

Standardization Documents Order Desk
Bldg. 4D
700 Robbins Avenue
Philadelphia, PA 19110-5094
For Assistance: (215) 697-2667 or 2179
Telephone Order Entry System (Touch-Tone Access Only): (215) 697-1187
FAX: (215) 697-2978

Copies of the DODISS's are also available on a yearly subscription basis from the Standardization Documents Order Desk.

2.2 Other Referenced Documents

Other referenced documents, government and non-government are listed in other sections of this handbook under “REFERENCES.”

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

FAILURE, NON-CHARGEABLE: (a) A non-relevant failure. (b) A relevant failure caused by a condition previously not specified as being the responsibility of a given organizational entity. All relevant failures are chargeable to one organizational entity or another.

FAILURE, NON-RELEVANT: (a) A failure verified as having been caused by a condition not present in the operational environment. (b) A failure verified as peculiar to an item design that will not enter the operational, or active, inventory.

FAILURE, RANDOM: A failure, the occurrence of which cannot be predicted except in a probabilistic or statistical sense.

FAILURE RATE: The total number of failures within an item population, divided by the total number of life units expended by that population, during a particular measurement period under stated conditions.

FALSE ALARM RATE (FAR): The frequency of occurrence of false alarms over a defined period of measure (e.g., time, cycles, etc.).

FALSE ALARM: A fault indicated by BIT or other monitoring circuitry where no fault can be found or confirmed.
FAULT: Immediate cause of failure (e.g., maladjustment, misalignment, defect, etc.).

FAULT DETECTION (FD): A process which discovers the existence of faults.

FAULT ISOLATION (FI): The process of determining the location of a fault to the extent necessary to effect repair.

FAULT ISOLATION TIME: The time spent arriving at a decision as to which items caused the system to malfunction. This includes time spent working on (replacing, attempting to repair, and adjusting) portions of the system shown by subsequent interim tests not to have been the cause of the malfunction.

FAULT LOCALIZATION: The process of determining the approximate location of a fault.

FRACTION OF FAULTS DETECTABLE (FFD): That fraction of all failures that occur over operating time, t, that can be correctly identified through direct observation or other specified means by an operator or by maintenance personnel under stated conditions.

FRACTION OF FAULTS ISOLATABLE (FFI): That fraction of all failures that occur over operating time, t, that can be correctly isolated to n or fewer units at a given maintenance level through the use of specified means by maintenance personnel under stated conditions.

FUNCTIONAL TEST: An evaluation of a product or item while it is being operated and checked under limited conditions without the aid of its associated equipment in order to determine its fitness for use.

-G-

GOVERNMENT-FURNISHED EQUIPMENT (GFE): An item provided for inclusion in or use with a product or service being procured by the Government.

GUIDE SPECIFICATION: This is a type of performance specification prepared by the Government. It identifies standard, recurring requirements that must be addressed when developing new systems, subsystems, equipments, and assemblies. Its structure forces appropriate tailoring to meet user needs.

-H-

HUMAN ENGINEERING (HE): The application of scientific knowledge to the design of items to achieve effective user-system integration (man-machine interface).

HUMAN FACTORS: A body of scientific facts about human characteristics. The term covers all biomedical and psychosocial considerations; it includes, but is not limited to, principles and applications in the areas of human engineering, personnel selection, training, life support, job performance aids, work loads, and human performance evaluation.

-I-

INACTIVE TIME: That time during which an item is in reserve. (In an inactive inventory).

INHERENT AVAILABILITY(Ai): A measure of availability that includes only the effects of an item design and its application, and does not account for effects of the operational and support environment. Sometimes referred to as "intrinsic" availability.

INHERENT R&M VALUE: A measure of reliability or maintainability that includes only the effects of an item's design and application, and assumes an ideal operating and support environment.

INITIAL ISOLATION LEVEL OF AMBIGUITY: The initial number of possible product subunits, identified by the built-in-test, built-in-test equipment, external test equipment, or manual test procedure, which might contain the failed component.

INITIAL ISOLATION: Isolation to the product subunit which must be replaced on line to return the product to operation. A subunit can be a modular assembly, or a component such as a crystal or antenna subsection. In the event that the maintenance concept requires a subunit to be removed, repaired and then replaced in the product, initial isolation includes both isolation to the failed subunit and isolation to the failed and removable portion of the subunit.

INTEGRATED DIAGNOSTICS: A structured process which maximizes the effectiveness of diagnostics by integrating pertinent elements, such as testability, automatic and manual testing, training, maintenance aiding, and technical information as a means for providing a cost effective capability to unambiguously detect and isolate all faults known or expected in items and to satisfy system mission requirements. Products of this process are hardware, software, documentation, and trained personnel.

INTEGRATED PRODUCT TEAM: A concurrent engineering team made up of individuals representing all relevant disciplines associated with a product's design, manufacturing, and marketing. All members work together using shared knowledge and capabilities to develop and manufacture a product in which requirements are balanced. The individuals must be committed to a common purpose, work to a unified set of requirements, and hold themselves accountable for decisions made and actions taken.

INTERCHANGE: Removing the item that is to be replaced, and installing the replacement item.

INTERCHANGEABILITY: The ability to interchange, without restriction, like equipments or portions thereof in manufacture, maintenance, or operation. Like products are two or more items that possess such functional and physical characteristics as to be equivalent in performance and durability, and are capable of being exchanged one for the other without alteration of the items themselves or of adjoining items, except for adjustment, and without selection for fit and performance.

INTERFACE DEVICE: An item which provides mechanical and electrical connections and any signal conditioning required between the automatic test equipment (ATE) and the unit under test (UUT); also known as an interface test adapter or interface adapter unit.

INVENTORY, ACTIVE: The group of items assigned to an operational status.

INVENTORY, INACTIVE: The group of items being held in reserve for possible future assignment to an operational status.

ISOLATION: Determining the location of a failure to the extent possible.

ITEM: A general term used to denote any product, system, material, part, subassembly, set, accessory, shop replaceable assembly (SRA), Shop Replaceable Unit (SRU), Weapon Replaceable Assembly (WRA), Line Replaceable Unit (LRU), etc.

SECTION 3: DEFINITIONS OF TERMS, ACRONYMS AND ABBREVIATIONS

-L-

LEVELS OF MAINTENANCE: The division of maintenance, based on different and requisite technical skill, which jobs are allocated to organizations in accordance with the availability of personnel, tools, supplies, and the time within the organization. Within the DoD, typical maintenance levels are organizational, intermediate and depot.

LIFE CYCLE COST (LCC): The sum of acquisition, logistics support, operating, and retirement and phase-out expenses.

LIFE CYCLE PHASES: Identifiable stages in the life of a product from the development of the first concept to removing the product from service and disposing of it. Within the Department of Defense, four phases are formally defined: Concept Exploration; Program Definition and Risk Reduction; Engineering and Manufacturing Development; and Production, Deployment, and Operational Support. Although not defined as a phase, demilitarization and disposal is defined as those activities conducted at the end of a product's useful life. Within the commercial sector, various ways of dividing the life cycle into phases are used. One way is: Customer Need Analysis, Design and Development, Production and Construction, Operation and Maintenance, and Retirement and Phase-out.

LIFE PROFILE: A time-phased description of the events and environments experienced by an item throughout its life. Life begins with manufacture, continues during operational use (during which the item has one or more mission profiles), and ends with final expenditure or removal from the operational inventory.

LINE REPLACEABLE UNIT (LRU): A unit designed to be removed upon failure from a larger entity (product or item) in the operational environment, normally at the organizational level.

LIFE UNITS: A measure of use duration applicable to the item. Measures include time, cycles, distance, rounds fired, attempts to operate, etc.

LOCALIZATION: Determining the location of a failure to the extent possible, without using accessory test equipment.

-M-

MAINTAINABILITY: The relative ease and economy of time and resources with which an item can be retained in, or restored to, a specified condition when maintenance is performed by personnel having specified skill levels, using prescribed procedures and resources, at each prescribed level of maintenance and repair. Also, the probability that an item can be retained in, or restored to, a specified condition when maintenance is performed by personnel having specified skill levels, using prescribed procedures and resources, at each prescribed level of maintenance and repair.

MAINTAINABILITY, MISSION: Maintainability as measured when maintenance is performed during the course of a specified mission profile. A mission-related system maintainability parameter.

MAINTENANCE: All actions necessary for retaining an item in or restoring it to a specified condition.

MAINTENANCE ACTION: An element of a maintenance event. One or more tasks (i.e., fault localization, fault isolation, servicing and inspection) necessary to retain an item in or restore it to a specified condition.

MAINTENANCE, CORRECTIVE: See Corrective Maintenance.

MAINTENANCE EVENT: One or more maintenance actions required to effect corrective and preventive maintenance due to any type of failure or malfunction, false alarm or scheduled maintenance plan.

MAINTENANCE, MANNING LEVEL: The total number of authorized or assigned personnel to support a given system at specified levels of maintenance.

MAINTENANCE, PREVENTIVE: See Preventive Maintenance.

MAINTENANCE RATIO: A measure of the total maintenance manpower burden required to maintain an item. It is expressed as the cumulative number of labor hours of maintenance expended in direct labor during a given period of the life units divided by the cumulative number of end item life units during the same period.

MAINTENANCE, SCHEDULED: See Scheduled Maintenance

MAINTENANCE, UNSCHEDULED: See Unscheduled Maintenance

MAINTENANCE TASK: The maintenance effort necessary for retaining an item in, or changing/restoring it to a specified condition.

MAINTENANCE TIME: An element of downtime which excludes modification and delay time.

MEAN DOWNTIME (MDT): The average time a system is unavailable for use due to a failure. Time includes the actual repair time plus all delay time associated with a repair person arriving with the appropriate replacement parts.

MEAN MAINTENANCE TIME: A basic measure of maintainability taking into account maintenance policy. The sum of preventive and corrective maintenance times, divided by the sum of scheduled and unscheduled maintenance events, during a stated period of time.

MEAN TIME BETWEEN DEMAND (MTBD): A measure of system reliability related to demand for logistic support. The total number of system life units divided by the total number of system demands on the supply system during a stated period of time.

MEAN TIME BETWEEN DOWNING EVENTS: A measure of system reliability related to readiness and availability. The total number of system life units divided by the total number of events which cause the system to be unavailable to initiate its mission(s), over a stated period of time.

MEAN TIME BETWEEN CRITICAL FAILURE (MTBCF): A measure of mission or functional reliability. The mean number of life units during which the item performs its mission or function within specified limits, during a particular measurement interval under stated conditions.

MEAN TIME BETWEEN FAILURE (MTBF): A basic measure of reliability for repairable items. The mean number of life units during which all parts of the item perform within their specified limits, during a particular measurement interval under stated conditions.

MEAN TIME BETWEEN MAINTENANCE (MTBM): A measure of the reliability taking into account maintenance policy. The total number of life units expended by a given time, divided by the total number of maintenance events (scheduled and unscheduled) due to that item.

MEAN TIME BETWEEN MAINTENANCE ACTIONS (MTBMA): A measure of the product reliability parameter related to demand for maintenance labor. The total number of product life units, divided by the total number of maintenance actions (preventive and corrective) during a stated period of time.

MEAN TIME BETWEEN REMOVALS (MTBR): A measure of the product reliability parameter related to demand for logistic support: The total number of system life units divided by the total number of items removed from that product during a stated period of time. This term is defined to exclude removals performed to facilitate other maintenance and removals for product improvement.

MEAN TIME TO FAILURE (MTTF): A basic measure of reliability for non-repairable items. The total number of life units of an item population divided by the number of failures within that population, during a particular measurement interval under stated conditions.

MEAN TIME TO REPAIR (MTTR): A basic measure of maintainability. The sum of corrective maintenance times at any specific level of repair, divided by the total number of failures within an item repaired at that level, during a particular interval under stated conditions.

MEAN TIME TO RESTORE SYSTEM (MTTRS): A measure of the product maintainability parameter, related to availability and readiness: The total corrective maintenance time, associated with downing events, divided by the total number of downing events, during a stated period of time. (Excludes time for off-product maintenance and repair of detached components.)

MEAN TIME TO SERVICE (MTTS): A measure of an on-product maintainability characteristic related to servicing that is calculated by dividing the total scheduled crew/operator/driver servicing time by the number of times the item was serviced.

MISSION RELIABILITY: The measure of the ability of an item to perform its required function for the duration of a specified mission profile. Mission reliability defines the probability that the system will not fail to complete the mission, considering all possible redundant modes of operation.

MISSION PROFILE: A time-phased description of the events and environments experienced by an item during a given mission. The description includes the criteria for mission success and critical failures.

MISSION TIME: That element of up time required to perform a stated mission profile.

MISSION-TIME-TO-RESTORE-FUNCTIONS (MTTRF): A measure of mission maintainability: The total corrective critical failure maintenance time, divided by the total number of critical failures, during the course of a specified mission profile.

MODIFICATION TIME: That time during which a product is being modified to enhance or expand functionality, correct a design deficiency, improve safety or reliability through design changes, or to bring the product up to the latest configuration.

SECTION 3: DEFINITIONS OF TERMS, ACRONYMS AND ABBREVIATIONS

-N-

NON-DEVELOPMENTAL ITEM (NDI): Any previously developed item used exclusively for governmental purposes by a Federal agency, a State or local government, or a foreign government with which the U.S. has a mutual defense cooperation agreement; any such item with minor modifications; and any item fully developed and in production but not yet in use. (See "Buying Commercial and Non-Developmental Items: A Handbook [SD-2, Apr 1996, OUSD/A&T]" or the Federal Acquisition Regulation Parts 6, 10, 11, 12 and 14, for a complete definition and criteria.)

NON-DESTRUCTIVE INSPECTION (NDI): Any method used for inspecting an item without physically, chemically, or otherwise destroying or changing the design characteristics of the item. However, it may be necessary to remove paint or other external coatings to use the NDI method. A wide range of technology is usually described as nondestructive inspection, evaluation, or testing (collectively referred to as non-destructive evaluation or NDE). The core of NDE is commonly thought to contain ultrasonic, visual, radiographic, eddy current, liquid penetrant, and magnetic particle inspection methods. Other methodologies, include acoustic emission, use of laser interference, microwaves, magnetic resonance imaging, thermal imaging, and so forth.

NON-DETECTABLE FAILURE: Failures at the component, equipment, subsystem, or system (product) level that are identifiable by analysis but cannot be identified through periodic testing or revealed by an alarm or an indication of an anomaly.

NOT-OPERATING TIME: That time during which the product is operable according to all indications or the last functional test, but is not being operated.

-O-

OPERABLE: The state in which an item is able to perform its intended function(s).

OPERATIONAL ENVIRONMENT: The aggregate of all external and internal conditions (such as temperature, humidity, radiation, magnetic and electric fields, shock vibration, etc.) either natural or man made, or self-induced, that influences the form, operational performance, reliability or survival of an item.

OPERATIONAL R&M: A measure of reliability and maintainability that includes the combined effects of design, installation, quality, environment, operation, maintenance, etc. on an item.

OPERATIONAL READINESS: The ability of a military unit to respond to its operation plan(s) upon receipt of an operations order. (A function of assigned strength, item availability, status, or supply, training, etc.).

OPERATIONAL TEST AND EVALUATION (OT&E): Test and evaluation which focuses on the development of optimum tactics, techniques, procedures, and concepts for products and items, evaluation of reliability, maintainability and operational effectiveness, and suitability of products and items under realistic operational conditions.

-P-

PERCENT ISOLATION TO A GROUP OF RIs: The percent of time that detected failures can be fault isolated to a specified ambiguity group of size n or less, where n is the number of replaceable items (RIs).

PERCENT ISOLATION TO A SINGLE RI: The percent of time that detected failures can be fault isolated to exactly one replaceable item (RI).

PERFORMANCE SPECIFICATION (PS): A design document stating the functional requirements for an item.

PERFORMANCE-BASED REQUIREMENTS (SPECIFICATION): Requirements that describe what the product should do, how it should perform, the environment in which it should operate, and interface and interchangeability characteristics. They should not specify how the product should be designed or manufactured.

PREDICTED: That which is expected at some future time, postulated on analysis of past experience and tests.

PROCESS ACTION TEAM (PAT): A group of individuals with complementary skills, committed to a common purpose, set of performance goals, and approach for which they hold themselves accountable, who work together using shared knowledge and capabilities to improve business processes.

PROGRAM-UNIQUE SPECIFICATION. This type of Government specification, also called a system specification, establishes requirements for items used for a particular weapon system or program. Little potential exists for the use of the document in other programs or applications. It is written as a performance specification, but it may include a blend of performance and detail design type requirements.

PREPARATION TIME: The time spent obtaining, setting up, and calibrating maintenance aids; warming up equipment; etc.

PREVENTIVE MAINTENANCE (PM): All actions performed to retain an item in specified condition by providing systematic inspection, detection, and prevention of incipient failures.

SECTION 3: DEFINITIONS OF TERMS, ACRONYMS AND ABBREVIATIONS

-Q-

QUALIFICATION TEST: A test conducted under specified conditions, by or on behalf of the customer, using items representative of the production configuration, to determine if item design requirements have been satisfied. Serves as a basis for production approval. Also known as a Demonstration Test.

-R-

REACTION TIME: The time between the instant a product is required to perform a function or mission and the time it is ready to perform that function or mission. It is the time needed for a product to be transitioned from a non-operating state to an operating state.

REASSEMBLY: Assembling the items that were removed during disassembly and closing the reassembled items.

RECONDITIONING: See Burn-In.

REDUNDANCY: The existence of more than one means for accomplishing a given function. Each means of accomplishing the function need not necessarily be identical. The two basic types of redundancy are active and standby.

> Active Redundancy - Redundancy in which all redundant items operate simultaneously.
>
> Standby Redundancy - Redundancy in which some or all of the redundant items are not operating continuously but are activated only upon failure of the primary item performing the function(s).

RELEVANT: That which can occur or recur during the life of an item.

RELIABILITY: (1) The duration or probability of failure-free performance under stated conditions. (2) The probability that an item can perform its intended function for a specified interval under stated conditions. (For non-redundant items this is equivalent to definition (1). For redundant items this is equivalent to definition of mission reliability.)

RELIABILITY-CENTERED MAINTENANCE (RCM): A disciplined logic or methodology used to identify preventive and corrective maintenance tasks to realize the inherent reliability of equipment at a minimum expenditure of resources.

RELIABILITY GROWTH: The improvement in reliability that results when design, material, or part deficiencies are revealed by testing and eliminated or mitigated through corrective action.

REPAIR TIME: The time spent replacing, repairing, or adjusting all items suspected to have been the cause of the malfunction, except those subsequently shown by interim test of the system not to have been the cause.

REPAIRABILITY: The probability that a failed item will be restored to operable condition within a specified time of active repair.
REPAIRABLE ITEM: An item which, when failed, can be restored by corrective maintenance to an operable state in which it can perform all required functions

REPLACEABLE ITEM (RI) or REPLACEABLE UNIT (RU): An item, unit, subassembly, or part which is normally intended to be replaced during corrective maintenance after its failure.

REQUEST FOR PROPOSAL (RFP): A letter or document sent to suppliers asking to show how a problem or situation can be addressed. Normally the supplier's response proposes a solution and quotes a price. Similar to a Request for Quote (RFQ), although the RFQ is usually used for products already developed.

-S-

SCHEDULED MAINTENANCE: Periodic prescribed inspection and servicing of products or items accomplished on the basis of calendar, mileage or hours of operation. Included in Preventive Maintenance.

SCREENING: A process for inspecting items to remove those that are unsatisfactory or likely to exhibit early failure. Inspection methods includes visual examination, physical dimension measurement, and functional performance measurement under specified environmental conditions.

SERVICEABILITY: The relative ease with which an item can be serviced (i.e., kept in operating condition).

SERVICING: The performance of any act needed to keep an item in operating condition, (i.e. lubricating, fueling, oiling, cleaning, etc.), but not including preventive maintenance of parts or corrective maintenance tasks.

SINGLE-POINT FAILURE: A failure of an item that causes the system to fail and for which no redundancy or alternative operational procedure exists.

SNEAK CIRCUIT ANALYSIS: An analytical procedure for identifying latent paths that cause occurrence of unwanted functions or inhibit desired functions, assuming all components are operating properly.

STANDARD PERFORMANCE SPECIFICATION (SPS): A type of specification that establishes requirements for military-unique items used in multiple programs or applications.

STORAGE LIFE: The length of time an item can be stored under specified conditions and still meet specified operating requirements. Also called shelf life.

SUBSYSTEM: A combination of sets, groups, etc. which performs an operational function within a product (system) and is a major subdivision of the product. (Example: Data processing subsystem, guidance subsystem).

SUPPLY DELAY TIME: The time between the demand on the supply system for a part or item to repair a product, or for a new product to replace a failed product, and the time when it is available.

SYSTEM: A composite of equipment and skills, and techniques capable of performing or supporting an operational role, or both. A complete system includes all equipment, related facilities, material, software, services, and personnel required for its operation and support to the degree that it can be considered self-sufficient in its intended operational environment.

SYSTEM DOWNTIME: The time interval between the commencement of work on a system (product) malfunction and the time when the system has been repaired and/or checked by the maintenance person, and no further maintenance activity is executed.

SYSTEM EFFECTIVENESS: (a) For repairable systems and items: the probability that a system can successfully meet an operational demand within a given time when operated under specified conditions. (b) For "one-shot" devices and non-repairable items: the probability that the system will operate successfully when called upon to do so under specified conditions.

SYSTEM FINAL TEST TIME: The time spent confirming that a system is in satisfactory operating condition (as determined by the maintenance person) following maintenance. It is possible for a system final test to be performed after each correction of a malfunction.

SYSTEM R&M PARAMETER: A measure of reliability or maintainability in which the units of measurement are directly related to operational readiness, mission success, maintenance labor costs, or logistics support costs.

-T-

TESTABILITY: A design characteristic which allows an item's status (operable, inoperable, or degraded) be determined and faults within the item to be isolated in a timely manner.

SECTION 3: DEFINITIONS OF TERMS, ACRONYMS AND ABBREVIATIONS

-U-

UNIT UNDER TEST (UUT): A UUT is any product or item (system, set, subsystem, assembly or subassembly, etc.) undergoing testing or otherwise being evaluated by technical means.

UNSCHEDULED MAINTENANCE: Corrective maintenance performed in response to a suspected failure.

UPTIME: That element of ACTIVE TIME during which an item is in condition to perform its required functions. (Increases availability and dependability).

UPTIME RATIO: A composite measure of operational availability and dependability that includes the combined effects of item design, installation, quality, environment, operation, maintenance, repair and logistic support: The quotient of uptime divided by the sum of uptime and downtime.)

USEFUL LIFE: The number of life units from manufacture to when the item has an unrepairable failure or unacceptable failure rate. Also, the period of time before the failure rate increases due to wearout.

UTILIZATION RATE: The planned or actual number of life units expended, or missions

SECTION 3: DEFINITIONS OF TERMS, ACRONYMS AND ABBREVIATIONS

DT - Development Test
DTIC - Defense Technical Information Center
DMH/MA - Direct Manhours per Maintenance Action
DT&E - Development Test and Evaluation

-E-

ECP - Engineering Change Proposal
EDIF - Electronic Data Interchange Format
EHC - Explosive Hazard Classification
EMC - Electromagnetic Compatibility
EMD - Engineering and Manufacturing Development
EMI - Electromagnetic Interference
EMP - Electromagnetic Pulse
EOL - End of Life
ESD - Electrostatic Discharge
ESS - Environmental Stress Screening
ETE - External Test Equipment

-F-

FA - False Alarm
FAR - False Alarm Rate
FEA - Finite Element Analysis
FMEA - Failure Modes and Effects Analysis
FMECA - Failure Modes, Effects, and Criticality Analysis
FD - Fault Detection
FD&I - Fault Detection and Isolation
FEA - Finite Element Analysis

-G-

GaAs	-	Gallium Arsenide
GEM	-	Generalized Emulation of Microcircuits
GIDEP	-	Government-Industry Data Exchange Program
GPTE	-	General Purpose Test Equipment
GS	-	Guide Specification

-H-

HALT	-	Highly Accelerated Life Test
HAST	-	Highly Accelerated Stress Test
HCR	-	Human Cognitive Reliability
HE	-	Human Engineering

-I-

IC	-	Integrated Circuit
IEC	-	International Electrotechnical Commission
IEEE	-	Institute of Electrical and Electronic Engineers
ILS	-	Integrated Logistics Support
IOT&E	-	Initial Operational Test and Evaluation
IPD	-	Integrated Product Team
IPDT	-	Integrated Product Development Team

-L-

LCC	-	Life Cycle Cost
LRM	-	Line Replaceable Module
LRU	-	Line Replaceable Unit
LSA	-	Logistics Support Analysis

-M-

MA	-	Maintenance Action
MCM	-	Multichip Module
MDT	-	Mean Downtime
MIMIC	-	Monolithic Microwave Millimeter Wave Integrated Circuit
MOS	-	Metal Oxide Semiconductor
MOV	-	Metal Oxide Varistor
MPCAG	-	Military Parts Control Advisory Group
MR	-	Mission Reliability or Maintenance Rate

SECTION 3: DEFINITIONS OF TERMS, ACRONYMS AND ABBREVIATIONS

MTBF	-	Mean Time Between Failure
MTBCF	-	Mean Time Between Critical Failure
MTBD	-	Mean Time Between Demand
MTBDE	-	Mean Time Between Downing Events
MTBF	-	Mean Time Between Failure
MTBM	-	Mean Time Between Maintenance
MTTF	-	Mean Time To Failure
MTTR	-	Mean Time To Repair
MTTRS	-	Mean Time To Restore System
MTTS	-	Mean Time To Service
MVT	-	Majority Vote Comparator

-N-

NDI	-	Non-Developmental Item or Non-Destructive Inspection

-O-

O&M	-	Operation and Maintenance
O&SHA	-	Operating and Support Hazard Analysis
OHHA	-	Occupational Health Hazard Assessment
OT&E	-	Operational Test and Evaluation

-P-

PAT	-	Process Action Team
PCB	-	Printed Circuit Board
PDR	-	Preliminary Design Review
PEM	-	Plastic Encapsulated Microcircuit
PHA	-	Preliminary Hazard Analysis
PHL	-	Preliminary Hazard List
PLD	-	Programmable Logic Device
PM	-	Preventive Maintenance
PMP	-	Parts Management Program
PPL	-	Preferred Parts List
PPSL	-	Program Parts Selection List
PRDR	-	Preproduction Reliability Design Review
PSP	-	Performance Shaping Factor
P&V	-	Power and Voltage

-Q-

QFD	-	Quality Function Deployment
QML	-	Qualified Manufacturers List

-R-

SECTION 3: DEFINITIONS OF TERMS, ACRONYMS AND ABBREVIATIONS

-T-

TAAF	-	Test, Analyze, and Fix
TMDE	-	Test, Measurement, and Diagnostic Equipment
TQM	-	Total Quality Management
TRB	-	Technology Review Board
TTF	-	Time to Failure

-U-

UR	-	Uptime Ratio or Utilization Rate
UUT	-	Unit Under Test

-V-

VHDL	-	VHSIC Hardware Description Language
VHSIC	-	Very High Speed Integrated Circuit

-W-

WSEIAC	-	Weapon System Effectiveness Industry Advisory Committee
WCA	-	Worst Case Analysis
WCCA	-	Worst Case Circuit Analysis
WRA	-	Weapon Replaceable Assembly
WUC	-	Work Unit Code

4.0 GENERAL STATEMENTS

4.1 Introduction and Background

For all but the most recent years of human history, the performance expected from man's system effectiveness relates to that property of a system output which was the real reason for buying the system in the first place - namely, the carrying out of some intended function. If the system is effective, it carries out this function well. If it is not effective, attention must be focused on those system attributes which are deficient.

4.2 The System Engineering Process

In recent years, the word "system" has come to include:

(1) The prime mission equipment

(2) The facilities required for operation and maintenance

(3) The selection and training of personnel

(4) Operational and maintenance procedures

(5) Instrumentation and data reduction for test and evaluation

(6) Special activation and acceptance programs

(7) Logistic support programs

System engineering is the application of scientific, engineering, and management effort to:

(1) Transform an operational need into a description of system performance parameters and a system configuration through the use of an iterative process of definition, synthesis,

SECTION 4: GENERAL STATEMENTS

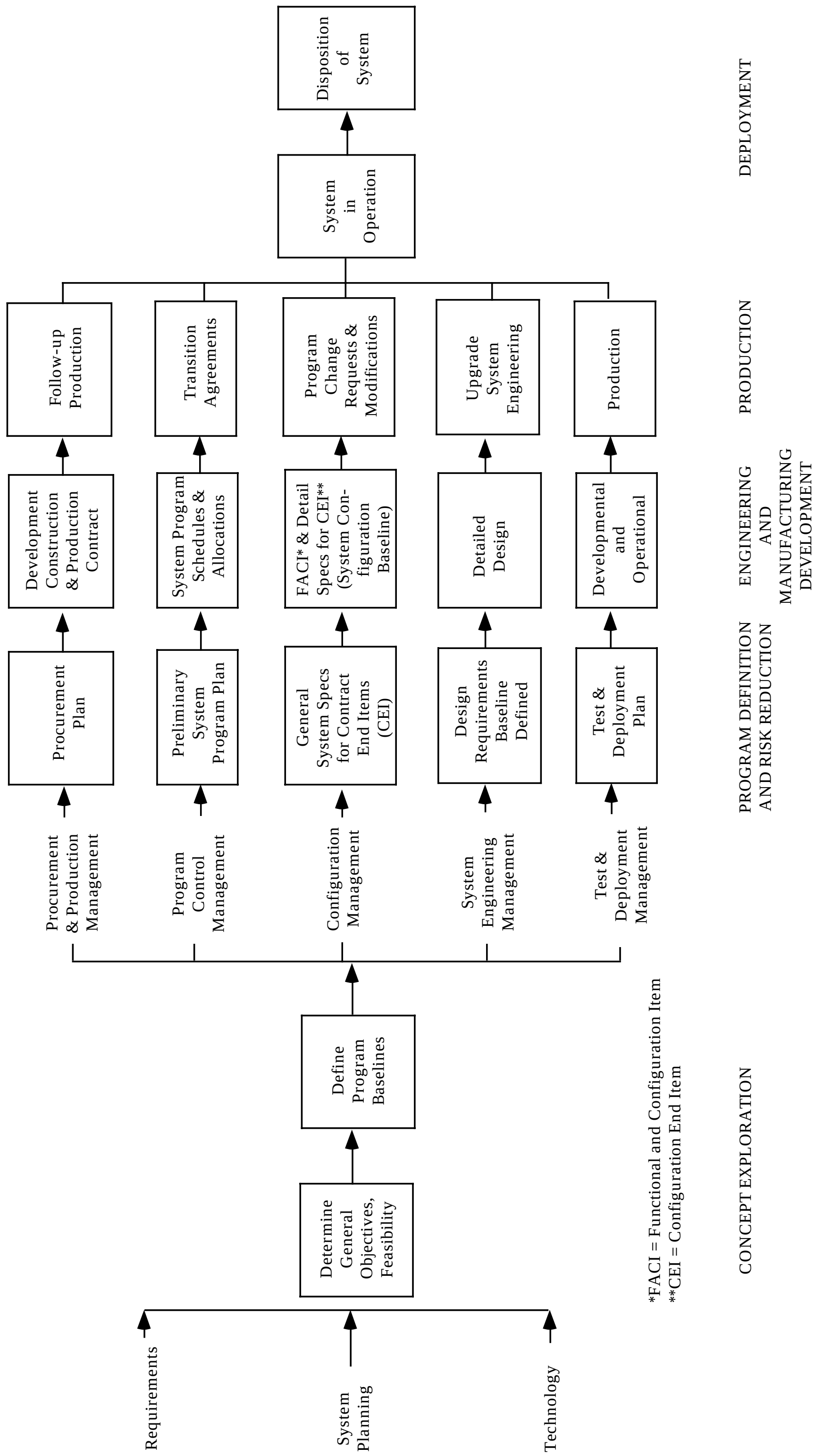


FIGURE 4.2-1: SYSTEM MANAGEMENT ACTIVITIES

help shape the basic system design, it is not specifically included in Step 1. Step 1 is intended to form a first hypothesis as a start toward the eventual solution.

In Step 2, the first hypothesis is evaluated against constraints such as design, cost, and time and against specific mission objectives to create criteria for designing equipment, defining intersystem interfaces, defining facilities, and determining requirements for personnel, training, training equipment and procedures.

Step 3 consists of system design studies that are performed concurrently with Steps 2 and 4 to:

(1) Determine alternate functions and functional sequences

(2) Establish design personnel, training and procedural data requirements imposed by the functions

(3) Find the best way to satisfy the mission requirements

(4) Select the best design approach for integrating mission requirements into the actual hardware and related support activities

Normally, the studies in Step 3 involve tradeoffs where data are in the form of schematic block diagrams, outline drawings, intersystem and intrasystem interface requirements, comparative matrices, and data supporting the selection of each approach. Some of the scientific tools used in the system design studies in Step 3 are: probability theory, statistical inference, simulation, computer analysis, information theory, queuing theory, servomechanism theory, cybernetics, mathematics, chemistry, and physics.

Step 4 uses the design approach selected in Step 3 to integrate the design requirements from Step 2 into the Contract End Items (CEI's). The result of Step 4 provides the criteria for detailed design, development, and test of the CEI based upon defined engineering information and associated tolerances. Outputs from Step 4 are used to:

(1) Determine intersystem interfaces

(2) Formulate additional requirements and functions that evolve from the selected devices or techniques

(3) Provide feedback to modify or verify the system requirements and functional flow diagrams prepared in Step 1

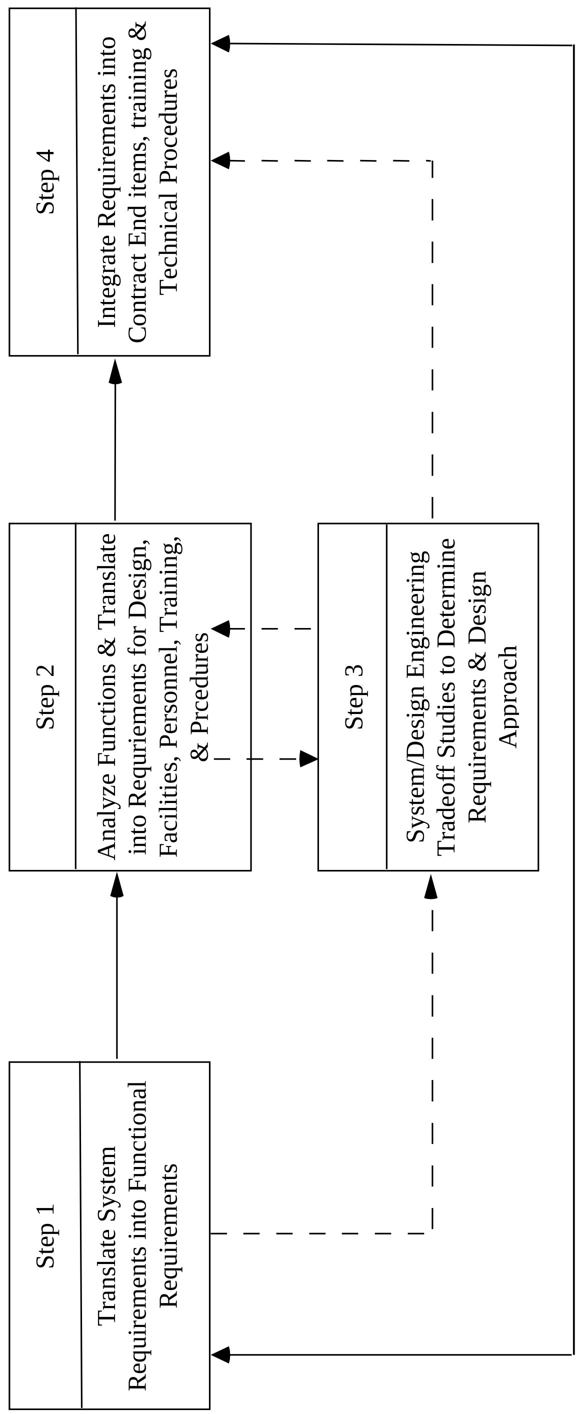


FIGURE 4.2-2: FUNDAMENTAL SYSTEM PROCESS CYCLE

When the first cycle of the system engineering process is completed, the modifications, alternatives, imposed constraints, additional requirements, and technological problems that have been identified are recycled through the process with the original hypothesis (initial design) to make the design more practical. This cycling is continued until a satisfactory design is produced, or until available resources (time, money, etc.) are expended and the existing design is accepted, or until the objectives are found to be unattainable.

Other factors that are part of the system engineering process - such as reliability, maintainability, safety, and human factors - exist as separate but interacting engineering disciplines and provide specific inputs to each other and to the overall system program. Pertinent questions at this point might be: "How do we know when the design is adequate?" or "How is the effectiveness of a system measured?" The answers to these questions lead to the concept of system effectiveness.

4.3 System Effectiveness

System effectiveness is a measure of the ability of a system to achieve a set of specific mission requirements. It is a function of readiness (or availability), and mission success (or dependability).

Cost and time are also critical in the evaluation of the merits of a system or its components, and must eventually be included in making administrative decisions regarding the purchase, use, maintenance, or discard of any equipment or system.

The operational effectiveness of a system obviously is influenced by the way the equipment was designed and built. It is, however, just as influenced by the way the equipment is used and maintained; i.e., system effectiveness is influenced by the designer, production engineer, maintenance man, and user/operator. The concepts of availability and dependability illustrate these influences and their relationships to system operational effectiveness. The following are the definitions of these concepts:

(1) **Availability** - A measure of the degree to which an item is in an operable and committable state at the start of a mission, when the mission is called for at an unknown (random) time.

(2) **Dependability** - A measure of the degree to which an item is operable and capable of performing its required function at any (random) time during a specified mission profile, given item availability at the start of the mission. (This definition is significantly different than the definition of dependability used by most other US and international organizations dealing with reliability e.g., the International Electrotechnical Commission (IEC) and the Society of Automotive Engineers (SAE). The IEC defines Dependability in publication IEC 50 Chapter 191 as: "The collective term used to describe the availability performance and its influencing factors: reliability

performance, maintaintability performance and maintenance support performance." As such, its use is restricted to general descriptions in non-quantitative terms).

Dependability is related to reliability; the intention was that dependability would be a more general concept then reliability.

4.3.1 R/M Considerations in System Effectiveness

From a system effectiveness viewpoint, reliability and maintainability jointly provide system availability and dependability. Increased reliability directly contributes to system uptime, while improved maintainability reduces downtime. If reliability and maintainability are not jointly considered and continually reviewed, serious consequences may result. With military equipment, failures or excessive downtime can jeopardize a mission and possibly cause a loss of lives. Excessive repair time and failures also impose burdens on logistic support and maintenance activities, causing high costs for repair parts and personnel training, expenditure of many man-hours for actual repair and service, obligation of facilities and equipment to test and service, and to movement and storage of repair parts.

From the cost viewpoint, reliability and maintainability must be evaluated over the system life cycle, rather than merely from the standpoint of initial acquisition. An effective design approach to reliability and maintainability can reduce the cost of upkeep.

Both reliability and maintainability are important considerations for the user of the system, although maintainability is probably more important from the point of view of most users. Although frequent system failures may be an annoyance, if each failure can be repaired in a very short time so that the system has a high availability, and the maintenance costs are reasonable, then the poor reliability may be acceptable. For example, if failures occur on the average of every fifteen minutes but can be repaired in a microsecond, at acceptable cost, the user will not be too concerned. On the other hand, if repair of a failure takes hours or days, the user has a non-available weapon system which may have a significant effect on the operational commander's readiness posture.

4.4 Factors Influencing System Effectiveness

4.4.1 Equipment of New Design

A typical history of the development of a new equipment would reveal a number of interesting steps in the progression from original concept to acceptable production model. These steps are particularly marked if the equipment represents a technical innovation, i.e., if it "pushes the state of the art" by introducing entirely new functions or by performing established functions in an entirely new way. Starting with a well- defined operational need, the research scientist, designer, reliability engineer, statistician, and production engineer all combine their talents to execute a multitude of operations leading to one ultimate objective: the production of an equipment that will perform as intended, with minimum breakdowns and maximum speed of repair. All this must be done at minimum cost and usually within an accelerated time schedule.

These program requirements are severe, to say the least. In order to meet them, many compromises are required. One of the first of these compromises is often a sharp curtailment in the basic research time allotted to the job of proving the feasibility of the new design. After only brief preliminary study, a pilot model of the equipment is built. With luck, it will work; but it is likely to be somewhat crude in appearance, too big and too heavy, not well-designed for mass production, subject to frequent failure, and difficult to repair. Indeed, at this early stage in the program, it is quite possible that the first model might be incapable of working if it were taken out of the laboratory and subjected to the more severe stresses of field operation, whether this be military or civilian. By the time this situation is corrected, the development program will have included many design changes, part substitutions, reliability tests, and field trials, eventually culminating in a successful operational acceptance test.

Usually, it is not until the equipment appears to have some chance of reaching this ultimate goal of acceptance that attention is focused on reduction of the frequency of failure, thus providing the impetus for a serious reliability effort. Experience has shown that this is unfortunate. Ideally, such an effort should begin immediately after the feasibility study, because some problems can be eliminated before they arise, and others can be solved at an early development stage, when design modifications can be effected most easily and economically. Even with this early start, reliability will continue to be a primary problem in new equipment, especially when it is of novel design. Early neglect of reliability must be compensated for by extraordinary efforts at a later period, because an equipment simply is not usable if it fails too frequently to permit suitable reliance on the likelihood of its operation when needed. Since such early neglect has been common in the past, reliability has received strong emphasis in the research designed to bring equipment performance characteristics up to satisfactory levels.

The description just given is generally applicable to the development of radically new equipment. However, when attention is directed to equipment in everyday use or to new equipment built predominantly on standard design principles and from well-tested parts, it becomes evident that effectiveness is dependent not only on performance capabilities and reliability but also on a number of other factors, including operational readiness, availability, maintainability, and repairability. Definitions for these concepts are given in Section 3. From the definitions it can be seen that they are all so interrelated that they must be viewed together and discussed, not as separate concepts but within the framework of the overall system to which they contribute.

4.4.2 Interrelationships Among Various System Properties

The discussion above implies that it is probably not practicable to maximize all of the desirable properties of a system simultaneously. Clearly, there are "tradeoff" relationships between reliability and system cost, between maintainability and system cost, between reliability and maintainability, and between many other properties. It would be most helpful to have a numerical scale of values for each of the several properties, and to have a multi-dimensional plot or chart showing the interrelationship among those values. Before such relationships can be obtained, it is first necessary to define in a precise and quantitative manner the properties with which we are concerned. The following outline is intended to show some of the factors which must be considered:

A. SYSTEM PERFORMANCE (DESIGN ADEQUACY)
- (1) Technical Capabilities
  - (a) Accuracy
  - (b) Range
  - (c) Invulnerability to countermeasures
  - (d) Operational simplicity
- (2) Possible Limitations on Performance
  - (a) Space and weight requirements
  - (b) Input power requirements
  - (c) Input information requirements
  - (d) Requirements for special protection against shock, vibration, low pressure, and other environmental influences

B. OPERATIONAL READINESS
- (1) Reliability
  - (a) Failure-free operation
  - (b) Redundancy or provision for alternative modes of operation
- (2) Maintainability
  - (a) Time to restore failed system to satisfactory operating status
  - (b) Technical manpower requirements for maintenance
  - (c) Effects of use-cycle on maintenance. (Can some maintenance be performed when operational use of the system is not required?)
- (3) Logistic Supportability
- (4) Availability

C. SYSTEM COST
- (1) Development cost, and particularly development time, from inception to operational capability
- (2) Production cost
- (3) Operating and operational support costs

4.5 Optimization of System Effectiveness

The optimization of system effectiveness is important throughout the system life cycle, from concept through the operation. Optimization is the balancing of available resources (time, money, personnel, etc.) against resulting effectiveness parameters (performance, operational readiness, etc.), until a combination is found that provides the most effectiveness for the desired expenditure of resources. Thus, the optimum system might be one that:

(1) Meets or exceeds a particular level of effectiveness for minimum cost, and/or

(2) Provides a maximum effectiveness for a given total cost

Optimization is illustrated by the flow diagram of Figure 4.5-1 which shows the optimization process as a feedback loop consisting of the following three steps:

(1) Designing many systems that satisfy the operational requirements and constraints

(2) Computing resultant values for effectiveness and resources used

(3) Evaluating these results and making generalizations concerning appropriate combinations of design and support factors, which are then fed back into the model through the feedback loops

Optimization also can be illustrated by the purchase of a new car or, more specifically, by putting into precise, quantifiable terms the rule, or criteria, that will be followed in the automobile selection process. Although automobiles do have quantifiable characteristics, such as horsepower, cost, and seating capacity, they are basically similar in most cars of a particular class (low-price sedans, sports models, etc.). Thus the selection criteria essentially reduces to esthetic appeal, prior experience with particular models, and similar intangibles. In the same sense, the choice of best design for the weapon system is greatly influenced by experience with good engineering practices, knowledge assimilated from similar systems, and economics. Despite this fuzziness, the selection criteria must be adjusted so that:

(1) The problem size can be reduced to ease the choice of approaches

(2) All possible alternatives can be examined more readily and objectively for adaptation to mathematical representation and analysis

(3) Ideas and experiences from other disciplines can be more easily incorporated into the solution

(4) The final choice of design approaches can be based on more precise, quantifiable terms, permitting more effective review and revision, and better inputs for future optimization problems

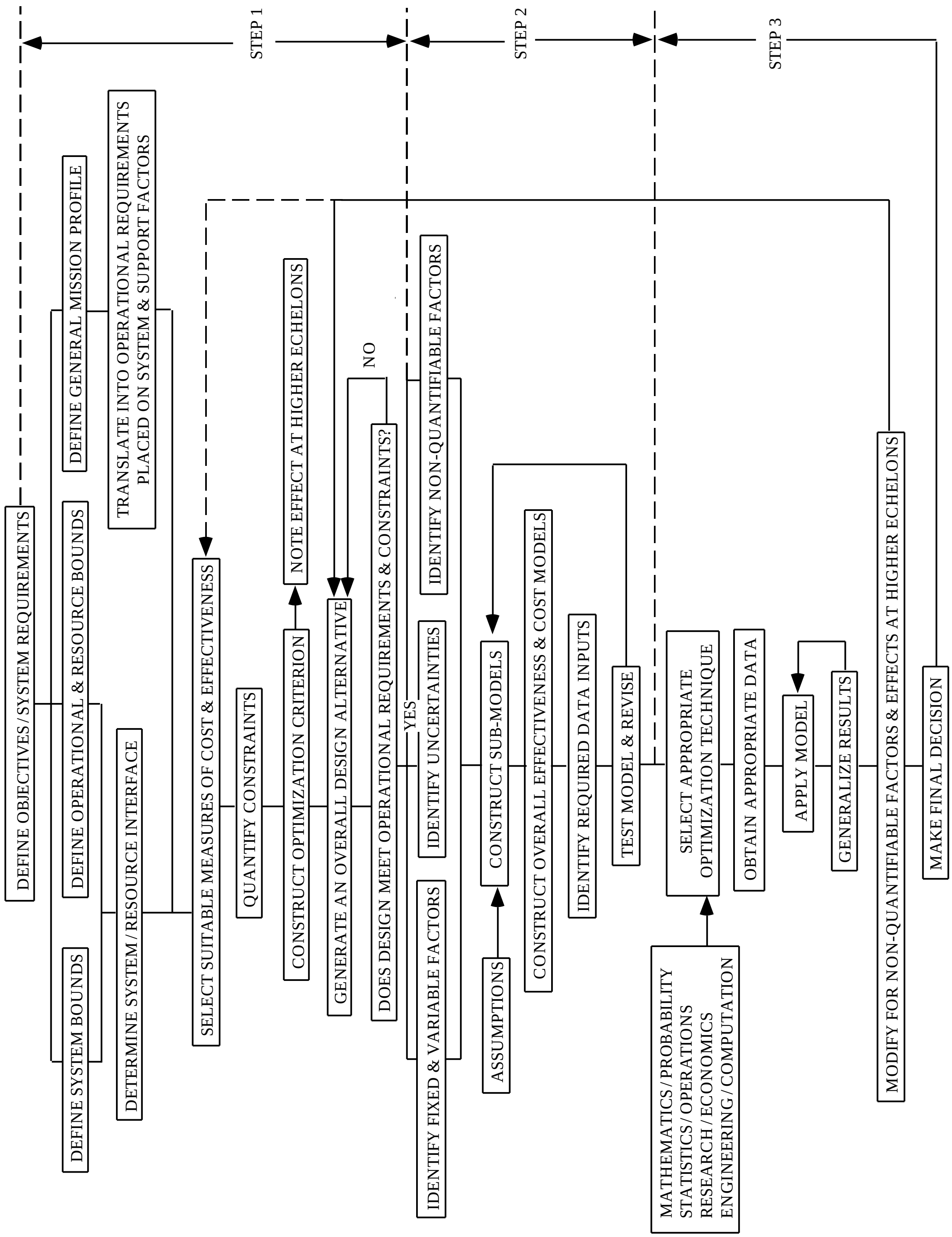


FIGURE 4.5-1: FLOW DIAGRAM FOR A GENERAL OPTIMIZATION PROCESS

The choice of parameters in the optimization model also is influenced by system definition. The automobile purchaser, for example, may not consider the manufacturer's and dealer's service policies. If these policies are considered, the system becomes the automobile plus the service policies. If service policies are not considered, the system consists only of the automobile.

The optimization of system effectiveness is a highly complex problem; there is a degree of interaction among the factors which enter into consideration of this problem. The actual techniques used to optimize system effectiveness will be described in greater detail in Section 10 of this handbook. Table 4.5-1, for example, lists only some of the more commonly-used techniques. These techniques are not peculiar to system effectiveness optimization, nor are they limited to system engineering.

This section is an introduction to the Handbook from a top level, or system, viewpoint. The remaining sections of this Handbook will expand upon the concepts introduced in this chapter. They will cover: (1) the basic reliability/maintainability/ availability theory, (2) practical application of the theory in terms of the design methodology and procedures of reliability engineering at the equipment and system level, (3) procedures for insuring that inherent reliability is not degraded during production and field deployment of systems, and (4) steps that management must take to insure the acquisition and deployment of reliable systems at minimum life cycle cost.

TABLE 4.5-1: PARTIAL LIST OF OPTIMIZATION TECHNIQUES

I. Mathematical Techniques	II. Statistical Techniques
Birth and death processes	Bayesian analysis
Calculus of finite differences	Decision theory
Calculus of variations	Experimental design
Gradient theory	Information theory
Numerical approximation	Method of steepest ascent
Symbolic logic	Stochastic processes
Theory of linear integrals	
Theory of maxima and minima	
III. Programming Techniques	IV. Other
Dynamic programming	Gaming theory
Linear programming	Monte Carlo techniques
Nonlinear programming	Queuing theory
	Renewal theory
	Search theory
	Signal flow graphs
	Value theory

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

SECTION 5: RELIABILITY/MAINTAINABILITY/AVAILABILITY THEORY

5.0 RELIABILITY/MAINTAINABILITY/AVAILABILITY THEORY

5.1 Introduction

The language of engineering is mathematics. The theories behind each engineering specialty are concisely stated in a set of mathematical procedures. For the engineering specialties of reliability, availability and maintainability (RAM), the theories are stated in the mathematics of probability and statistics.

The underlying reason for the use of these concepts is the inherent uncertainty in predicting a failure. Even given a failure model based on physical or chemical reactions, the results will not be the time a part will fail, but rather the time a given percentage of the parts will fail or the probability that a given part will fail in a specified time. Individual parts will fail according to their individual strengths, which will vary from part to part and are practically unknowable. Similarly, the time to repair a failure will also vary dependent on many factors whose values in individual cases are practically unknowable.

Since RAM parameters must be defined in probabilistic terms, probabilistic parameters such as random variables, density functions, and distribution functions are utilized in the development of RAM theory.

This section describes some of the basic concepts, formulas, and simple examples of application of RAM theory which are required for better understanding of the underlying principles and design techniques presented in later sections. Practicality rather than rigorous theoretical exposition is emphasized. Many excellent texts are available (see references) for the reader who is interested in delving into the rigorous theoretical foundations of these disciplines.

5.2 Reliability Theory

Because, as was mentioned previously, reliability is defined in terms of probability, probabilistic parameters such as random variables, density functions, and distribution functions are utilized in the development of reliability theory. Reliability studies are concerned with both discrete and continuous random variables. An example of a discrete variable is the number of failures in a given interval of time. Examples of continuous random variables are the time from part installation to failure and the time between successive equipment failures.

The distinction between discrete and continuous variables (or functions) depends upon how the problem is treated and not necessarily on the basic physical or chemical processes involved. For example, in analyzing "one shot" systems such as missiles, one usually utilizes discrete functions such as the number of successes in "n" launches. However, whether or not a missile is successfully launched could be a function of its age, including time in storage, and could, therefore, be treated as a continuous function.

5.2.1 Basic Concepts

The cumulative distribution function F(t) is defined as the probability in a random trial that the random variable is not greater than t (see note), or

$$F(t) = \int_{-\infty}^{t} f(t)\, dt \tag{5.1}$$

where f(t) is the probability density function of the random variable, time to failure. F(t) is termed the "unreliability function" when speaking of failure. It can be thought of as representing the probability of failure prior to some time t. If the random variable is discrete, the integral is replaced by a summation. Since F(t) is zero until t=0, the integration in Equation 5.1 can be from zero to t.

NOTE: Pure mathematicians object to the use of the same letter in the integral and also in the limits of the integral. This is done here, and in the rest of this section in spite of the objection in order to simplify the reference to time as the variable in such functions as F(t), R(t), M(t), f(t), etc.

The reliability function, R(t), or the probability of a device not failing prior to some time t, is given by

$$R(t) = 1 - F(t) = \int_{t}^{\infty} f(t)\, dt \tag{5.2}$$

By differentiating Equation (5.2) it can be shown that

$$\frac{-dR(t)}{dt} = f(t) \tag{5.3}$$

The probability of failure in a given time interval between t_1 and t_2 can be expressed by the reliability function

$$\int_{t_1}^{\infty} f(t)\, dt - \int_{t_2}^{\infty} f(t)\, dt = R(t_1) - R(t_2) \tag{5.4}$$

The rate at which failures occur in the interval t_1 to t_2, the failure rate, $\lambda(t)$, is defined as the ratio of probability that failure occurs in the interval, given that it has not occurred prior to t_1, the start of the interval, divided by the interval length. Thus,

SECTION 5: RELIABILITY/MAINTAINABILITY/AVAILABILITY THEORY

$$\lambda(t) = \frac{R(t_1) - R(t_2)}{(t_2 - t_1)\ R(t_1)} \tag{5.5}$$

or the alternative form

$$\lambda(t) = \frac{R(t) - R(t + \Delta t)}{\Delta t R(t)} \tag{5.6}$$

where $t = t_1$ and $t_2 = t + \Delta t$. The hazard rate, h(t), or instantaneous failure rate, is defined as the limit of the failure rate as the interval length approaches zero, or

$$h(t) = \lim_{\Delta t \to 0} \left[\frac{R(t) - R(t + \Delta t)}{\Delta t R(t)} \right]$$

$$= \frac{-1}{R(t)} \left[\frac{dR(t)}{dt} \right] = \frac{1}{R(t)} \left[\frac{-dR(t)}{dt} \right] \tag{5.7}$$

But it was previously shown, Eq. (5.3), that

$$f(t) = \frac{-dR(t)}{dt}$$

Substituting this into Eq. (5.7) we get $$h(t) \frac{f(t)}{R(t)} \tag{5.8}$$

This is one of the fundamental relationships in reliability analysis. For example, if one knows the density function of the time to failure, f(t), and the reliability function, R(t), the hazard rate function for any time, t, can be found. The relationship is fundamental and important because it is independent of the statistical distribution under consideration.

The differential equation of Eq. (5.7) tells us, then, that the hazard rate is nothing more than a measure of the change in survivor rate per unit change in time.

Perhaps some of these concepts can be seen more clearly by use of a more concrete example. Suppose that we start a test at time, t_0, with N_O devices. After some time t, N_f of the original devices will have failed, and N_S will have survived ($N_O = N_f + N_s$). The reliability, R(t), is given at any time t, by:

$$R(t) = \frac{N_S}{N_O} \tag{5.9}$$

$$= \frac{N_O - N_f}{N_O} = 1 - \frac{N_f}{N_O} \tag{5.10}$$

From Eq. (5.3)

$$f(t) = -\frac{dR(t)}{dt} = \frac{1}{N_O}\frac{dN_f}{dt} \tag{5.11}$$

Thus, the failure density function represents the proportion of the original population, (N_O), which fails in the interval (t, t + Δt).

On the other hand, from Eqs. (5.8), (5.9) and (5.11)

$$h(t) = \frac{f(t)}{R(t)} = \frac{\frac{1}{N_O}\frac{dN_f}{dt}}{N_S/N_O} = \frac{1}{N_S}\frac{dN_f}{dt} \tag{5.12}$$

Thus, h(t) is inversely proportional to the number of devices that survive to time t, (N_S), which fail in the interval (t, t + Δt).

Although, as can be seen by comparing Eqs. (5.6) and (5.7), failure rate, λ(t), and hazard rate, h(t), are mathematically somewhat different, they are usually used synonymously in conventional reliability engineering practice. It is not likely that this handbook will change firmly entrenched conventional practice, so the reader should be aware of this common deviation from exact mathematical accuracy.

Perhaps the simplest explanation of hazard and failure rate is made by analogy. Suppose a family takes an automobile trip of 200 miles and completes the trip in 4 hours. Their average rate was 50 mph, although they drove faster at some times and slower at other times. The rate at any given instant could have been determined by reading the speed indicated on the speedometer at that instant. The 50 mph is analogous to the failure rate and the speed at any point is analogous to the hazard rate.

In Eq. (5.8), a general expression was derived for hazard (failure) rate. This can also be done for the reliability function, R(t). From Eq. (5.7)

$$h(t) = -\frac{1}{R(t)}\left[\frac{dR(t)}{dt}\right] \tag{5.13}$$

$$\frac{dR(t)}{R(t)} = - h(t)\, dt$$

Integrating both sides of Eq. (5.13)

$$\int_0^t \frac{dR(t)}{R(t)} = -\int_0^t h(t)\, dt$$

$$ln\, R(t) - ln\, R(0) = -\int_0^t h(t)\, dt$$

but R(0) = 1, ln R(0) = 0, and

$$R(t) = \exp\left[-\int_0^t h(t)dt\right] \tag{5.14}$$

Eq. (5.14) is the general expression for the reliability function. If h(t) can be considered a constant failure rate (λ), which is true for many cases for electronic equipment, Eq. (5.14) becomes

$$R(t) = e^{-\lambda t} \tag{5.15}$$

Eq. (5.15) is used quite frequently in reliability analysis, particularly for electronic equipment. However, the reliability analyst should assure himself that the constant failure rate assumption is valid for the item being analyzed by performing goodness of fit tests on the data. These are discussed in Section 8.

In addition to the concepts of f(t), h(t), λ(t), and R(t), previously developed, several other basic, commonly-used reliability concepts require development. They are: mean-time-to-failure (MTTF), mean life (θ), and mean-time-between-failure (MTBF).

Mean-Time-To-Failure (MTTF)

MTTF is nothing more than the expected value of time to failure and is derived from basic statistical theory as follows:

$$\text{MTTF} = \int_0^\infty t\, f(t)\, dt$$

$$= \int_0^\infty t \left[- \frac{dR(t)}{dt} \right] dt \qquad (5.16)$$

Integrating by parts and applying "Hopital's rule," we arrive at the expression

$$\text{MTTF} = \int_0^\infty R(t)\, dt \qquad (5.17)$$

Eq. (5.17), in many cases, permits the simplification of MTTF calculations. If one knows (or can model from the data) the reliability function, R(t), the MTTF can be obtained by direct integration of R(t) (if mathematically tractable), by graphical approximation, or by Monte Carlo simulation. For repairable equipment MTTF is defined as the mean time to first failure.

Mean Life (θ)

The mean life (θ) refers to the total population of items being considered. For example, given an initial population of n items, if all are operated until they fail, the mean life (θ) is merely the arithmetic mean time to failure of the total population given by:

$$\theta = \frac{\sum_{i=1}^{n} t_i}{n} \qquad (5.18)$$

where:

t_i = time to failure of the i^{th} item in the population

n = total number of items in the population

Mean-Time-Between-Failure (MTBF)

This concept appears quite frequently in reliability literature; it applies to repairable items in which failed elements are replaced upon failure. The expression for MTBF is:

$$\text{MTBF} = \frac{T(t)}{r} \tag{5.19}$$

where:

T(t) = total operating time
r = number of failures

It is important to remember that MTBF only has meaning for repairable items, and, for that case, MTBF represents exactly the same parameter as mean life (θ). More important is the fact that a constant failure rate is assumed. Thus, given the two assumptions of replacement upon failure and constant failure rate, the reliability function is:

$$R(t) = e^{-\lambda t} = e^{-t/\theta} = e^{-t/\text{MTBF}} \tag{5.20}$$

and (for this case)

$$\lambda = \frac{1}{\text{MTBF}} \tag{5.21}$$

Figure 5.2-1 provides a convenient summary of the basic concepts developed in this section.

Failure Density Function (time to failure)	$f(t)$
Reliability Function	$R(t) = \int_t^{\infty} f(t)\,dt = \exp\left[-\int_0^t h(t)dt\right]$
Hazard Rate (Failure Rate)	$h(t) = f(t)/R(t)$ $\lambda(t) = \int_0^t h(t)\,dt$
Mean Time to Failure (MTTF) (no repair)	$\text{MTTF} = \int_0^{\infty} R(t)\,dt$
Mean Time Between Failure (constant failure rate, λ, with repair)	$\text{MTBF} = \frac{T(t)}{r} = 1/\lambda$

FIGURE 5.2-1: SUMMARY OF BASIC RELIABILITY CONCEPTS

5.3 Statistical Distributions Used in Reliability Models

There are many standard statistical distributions which may be used to model the various reliability parameters. It has been found that a relatively small number of statistical distributions satisfies most needs in reliability work. The particular distribution used depends upon the nature of the data, in each case. The following is a short summary of some of the distributions most commonly used in reliability analysis, criteria for their use, and examples of application. Figures 5.3-1 and 5.3-2 are summaries of the shape of common failure density, reliability, and hazard rate functions for the distributions described. Each distribution will be described in more detail, with reliability examples, in the following sections.

5.3.1 Continuous Distributions

5.3.1.1 Normal (or Gaussian) Distribution

There are two principal applications of the normal distribution to reliability. One application deals with the analysis of items which exhibit failure due to wear, such as mechanical devices. Frequently the wear-out failure distribution is sufficiently close to normal that the use of this distribution for predicting or assessing reliability is valid.

Another application is in the analysis of manufactured items and their ability to meet specifications. No two parts made to the same specification are exactly alike. The variability of parts leads to a variability in systems composed of those parts. The design must take this part variability into account, otherwise the system may not meet the specification requirement due to the combined effect of part variability. Another aspect of this application is in quality control procedures.

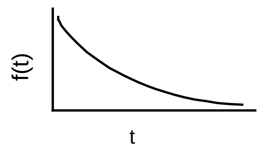
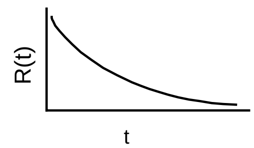
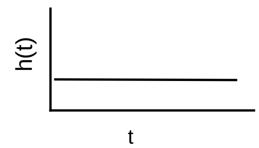
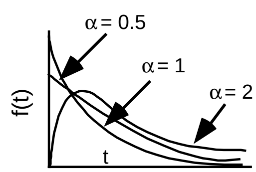
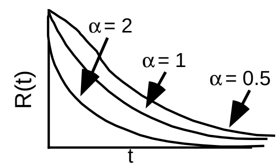
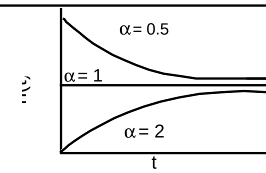
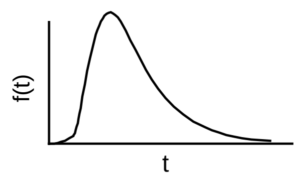
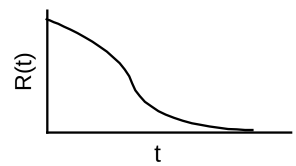
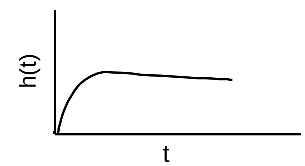
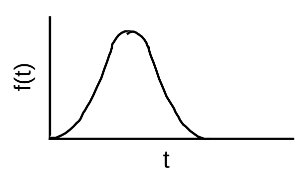
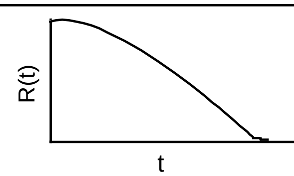
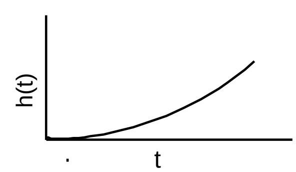
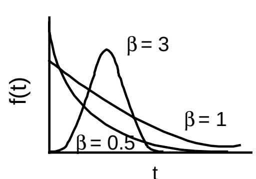
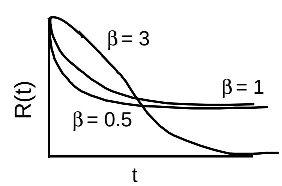
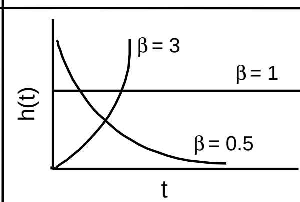
The basis for the use of normal distribution in this application is the central limit theorem which states that the sum of a large number of identically distributed random variables, each with finite mean and variance, is normally distributed.

Thus, the variations in value of electronic component parts, for example, due to manufacturing are considered normally distributed.

The failure density function for the normal distribution is

$$f(t) = \frac{1}{s\sqrt{2p}} \quad e^{\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^{2}\right]}, \text{ where } -\infty < t < \infty \qquad (5.22)$$

SECTION 5: RELIABILITY/MAINTAINABILITY/AVAILABILITY THEORY

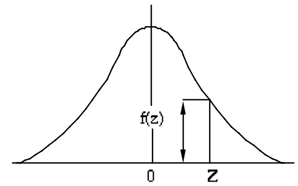
TYPE OF DISTRIBUTION	PROBABILITY DENSITY FUNCTION, f(t)	RELIABILITY FUNCTION R(t) = 1 - f(t)	HAZARD FUNCTION $h(t) = \frac{f(t)}{R(t)}$
EXPONENTIAL	 $f(t) = \lambda e^{-\lambda t}$	 $R(t) = e^{-\lambda t}$	 $h(t) = \lambda = \theta^{-1}$
GAMMA	 $f(t) = \frac{\lambda}{\Gamma(\alpha)} (\lambda t)^{\alpha-1} e^{\lambda t}$	 $R(t) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} \int_t^{\infty} t^{\alpha-1} e^{-\lambda t}\, dt$	 $h(t) = \frac{t^{\alpha-1} e^{-\lambda t}}{\int_t^{\infty} t^{\alpha-1} e^{-\lambda t} dt}$
LOGNORMAL	 $f(t) = \frac{1}{\sigma t (2\pi)} e^{-\frac{1}{2}\left(\frac{\ln t - \mu}{\sigma}\right)^2}$	 $R(t) = 1 - \phi\left(\frac{\ln t - \mu}{\sigma}\right)$ See Note	 $h(t) = \frac{f(t)}{1 - \phi\left(\frac{\ln t - \mu}{\sigma}\right)}$
NORMAL	 $f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$	 $R(t) = 1 - \phi\left(\frac{t-\mu}{\sigma}\right)$ See Note	 $h(t) = \frac{f(t)}{1 - \phi\left(\frac{t-\mu}{\sigma}\right)}$
WEIBULL	 $f(t) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-\left[\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right]}$	 $R(t) = e^{-\left[\frac{(t-\gamma)^{\beta}}{\eta}\right]}$	 $h(t) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}$

Note: $\phi\left(\frac{\ln t - \mu}{\sigma}\right)$ (lognormal) and $\phi\left(\frac{t-\mu}{\sigma}\right)$ (normal) is the standardized form of these distributions and is equal to the integral of the pdfs for those distributions (i.e., the cumulative distribution function).

FIGURE 5.3-1: SHAPES OF FAILURE DENSITY, RELIABILITY AND HAZARD RATE FUNCTIONS FOR COMMONLY USED CONTINUOUS DISTRIBUTIONS

TABLE 5.3-1: VALUES OF THE STANDARD NORMAL DISTRIBUTION FUNCTION

TABLE 5.3-2: ORDINATES F(z) OF THE STANDARD NORMAL CURVE AT z



z	0	1	2	3	4	5	6	7	8	9
0.0	.3989	.3989	.3989	.3988	.3986	.3984	.3982	.3980	.3977	.3973
0.1	.3970	.3965	.3961	.3956	.3951	.3945	.3939	.3932	.3925	.3918
0.2	.3910	.3902	.3894	.3885	.3876	.3867	.3857	.3847	.3836	.3825
0.3	.3814	.3802	.3790	.3778	.3765	.3752	.3739	.3725	.3712	.3697
0.4	.3683	.3668	.3653	.3637	.3621	.3605	.3589	.3572	.3555	.3538
0.5	.3521	.3503	.3485	.3467	.3448	.3429	.3410	.3391	.3372	.3352
0.6	.3332	.3312	.3292	.3271	.3251	.3230	.3209	.3187	.3166	.3144
0.7	.3123	.3101	.3079	.3056	.3034	.3011	.2989	.2966	.2943	.2920
0.8	.2897	.2874	.2850	.2827	.2803	.2780	.2756	.2732	.2709	.2685
0.9	.2661	.2637	.2613	.2589	.2565	.2541	.2516	.2492	.2468	.2444
1.0	.2420	.2396	.2371	.2347	.2323	.2299	.2275	.2251	.2227	.2203
1.1	.2179	.2155	.2131	.2107	.2083	.2059	.2036	.2012	.1989	.1965
1.2	.1942	.1919	.1895	.1872	.1849	.1826	.1804	.1781	.1758	.1736
1.3	.1714	.1691	.1669	.1647	.1626	.1604	.1582	.1561	.1539	.1518
1.4	.1497	.1476	.1456	.1435	.1415	.1394	.1374	.1354	.1334	.1315
1.5	.1295	.1276	.1257	.1238	.1219	.1200	.1182	.1163	.1145	.1127
1.6	.1109	.1092	.1074	.1057	.1040	.1023	.1006	.0989	.0973	.0957
1.7	.0940	.0925	.0909	.0893	.0878	.0863	.0848	.0833	.0818	.0804
1.8	.0790	.0775	.0761	.0748	.0734	.0721	.0707	.0694	.0681	.0669
1.9	.0656	.0644	.0632	.0620	.0608	.0596	.0584	.0573	.0562	.0551
2.0	.0540	.0529	.0519	.0508	.0498	.0488	.0478	.0468	.0459	.0449
2.1	.0440	.0431	.0422	.0413	.0404	.0396	.0387	.0379	.0371	.0363
2.2	.0355	.0347	.0339	.0332	.0325	.0317	.0310	.0303	.0297	.0290
2.3	.0283	.0277	.0270	.0264	.0258	.0252	.0246	.0241	.0235	.0229
2.4	.0224	.0219	.0213	.0208	.0203	.0198	.0194	.0189	.0184	.0180
2.5	.0175	.0171	.0167	.0163	.0158	.0154	.0151	.0147	.0143	.0139
2.6	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110	.0107
2.7	.0104	.0101	.0099	.0096	.0093	.0091	.0088	.0086	.0084	.0081
2.8	.0079	.0077	.0075	.0073	.0071	.0069	.0067	.0065	.0063	.0061
2.9	.0060	.0058	.0056	.0055	.0053	.0051	.0050	.0048	.0047	.0046
3.0	.0044	.0043	.0042	.0040	.0039	.0038	.0037	.0036	.0035	.0034
3.1	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026	.0025	.0025
3.2	.0024	.0023	.0022	.0022	.0021	.0020	.0020	.0019	.0018	.0018
3.3	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014	.0013	.0013
3.4	.0012	.0012	.0012	.0011	.0011	.0010	.0010	.0010	.0009	.0009
3.5	.0009	.0008	.0008	.0008	.0008	.0007	.0007	.0007	.0007	.0006
3.6	.0006	.0006	.0006	.0005	.0005	.0005	.0005	.0005	.0005	.0004
3.7	.0004	.0004	.0004	.0004	.0004	.0004	.0003	.0003	.0003	.0003
3.8	.0003	.0003	.0003	.0003	.0003	.0002	.0002	.0002	.0002	.0002
3.9	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001

The standardized cumulative distribution function is,

$$\phi(t) = \int_{-\infty}^{z} \frac{1}{2\pi} \exp(-\frac{z^2}{2})dz \tag{5.28}$$

then for a normally distributed variable t, with mean μ and standard deviation σ

$$P(\mathbf{t} \le t) = P\left(Z \le \frac{t-\mu}{\sigma}\right) = \Phi\left(\frac{t-\mu}{\sigma}\right) \tag{5.29}$$

The hazard function for a normal distribution is a monotonically increasing function of t. This can be shown by proving $h'(t) \ge 0$ for all t.

5.3.2 Examples of Reliability Calculations Using the Normal Distribution

5.3.2.1 Microwave Tube Example

A microwave transmitting tube has been observed to follow a normal distribution with μ = 5000 hours and σ = 1500 hours. Find the reliability of such a tube for a mission time of 4100 hours and the hazard rate of one of these tubes at age 4400 hours.

$$R(t) = P\left(z > \frac{t - u}{\sigma}\right)$$

$$R(4100) = P\left(z > \frac{4100 - 5000}{1500}\right)$$

$$= P(z > -0.6) = 1 - P(z < -0.6)$$

$$= 1 - 0.27 = 0.73$$

as found in Table 5.3-1. Remember $P(z > -z_i) = P(z < z_i)$ by symmetry of the normal distribution.

$$h(t) = \frac{f(t)}{R(t)} = \frac{f(z)/\sigma}{R(t)}$$

$$f(t = 4400) = \frac{f\left(z = \frac{4400 - 5000}{1500}\right)}{1500} = \frac{1}{1500} f(z = -0.4)$$

$$= \ (0.00067)(0.37) \ = \ 0.00025$$

where f(z = 0.4) was obtained from Table 5.3-2. Remember f(z)=f(-z) because of the symmetry of the normal distribution.

$$R(4400) \ = \ P\left(z > \frac{4400 - 5000}{1500}\right) \ = \ P(z > -0.4) \ = \ 1 - P(z < -0.4) = 0.65$$

$$h(4400) \ = \ \frac{f(4400)}{R(4400)} \ = \ \frac{0.00025}{0.65} \ = \ 0.00038 \text{ failures/hour}$$

5.3.2.2 Mechanical Equipment Example

A motor generator has been observed to follow a normal distribution with μ = 300 hours and σ = 40 hours. Find the reliability of the motor generator for a mission time (or time before maintenance) of 250 hours and the hazard rate at 200 hours.

$$R(250) \ = \ P\left(z > \frac{250 - 300}{40}\right) \ = \ P(z > -1.25)$$

$$= \ 1 - P(z < -1.25) \ = \ 1 - 0.11 \ = \ 0.89$$

where P(z < -1.25) was interpolated from Table 5.3-1.

$$h(t) \ = \frac{f(t)}{R(t)} \ = \ \frac{f(z)/\sigma}{R(t)}$$

$$f(t = 200) \ = \frac{f\left(z = \frac{200 - 300}{40}\right)}{40} \ = \frac{1}{40} \ f(z = -2.5)$$

$$f(z = -2.5) \ = \ (0.025)(0.0175) = 0.00044$$

where f (z = 2.5) was found in Table 5.3-2.

$$R(200) \ = \ P\left(z > \frac{200 - 300}{40}\right) \ = \ P(z > -2.5) = 1 - P(z < -2.5) \ = \ 0.994$$

$$h(200) \ = \frac{f(200)}{R(200)} \ = \ \frac{0.00044}{0.994} \ = \ 0.00044 \text{ failures/hour}$$

5.3.3 Lognormal Distribution

The lognormal distribution is the distribution of a random variable whose natural logarithm is distributed normally; in other words, it is the normal distribution with *ln* t as the variate. The density function is

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{ln(t)-\mu}{\sigma}\right)^2\right] \quad \text{for } t \geq 0 \qquad (5.30)$$

where the mean = $\exp\left(\mu + \frac{\sigma^2}{2}\right)$ (5.31)

and the standard deviation = $\left[\exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2)\right]^{1/2}$ (5.32)

where μ and σ are the mean and standard deviation (SD) of *ln* (t).

The lognormal distribution is used in reliability analysis of semiconductors and fatigue life of certain types of mechanical components. This distribution is also commonly used in maintainability analysis and will be further discussed in Section 5.6.2.1.

The cumulative distribution function for the lognormal is,

$$F(t) = \int_0^t \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{ln(t)-\mu}{\sigma}\right)^2\right] dt \qquad (5.33)$$

this can be related to the standard normal variant Z by

$$F(t) = P[t \leq t] = P\left[Z \leq \left(\frac{ln t - \mu}{\sigma}\right)\right] \qquad (5.34)$$

the reliability function is 1-F(t) or

$$R(t) = (1 - F(t)) = P\left[Z > \left(\frac{ln(t)-\mu}{\sigma}\right)\right] \qquad (5.35)$$

the hazard function, h(t), is given as follows

$$h(t) = \frac{f(t)}{R(t)} = \frac{\varnothing\left(\frac{ln(t)-\mu}{\sigma}\right)}{t\sigma R(t)} \tag{5.36}$$

where $\varnothing$ is the standard normal probability function and μ and σ are the mean and t standard deviation of the natural logarithm of the random variable t.

5.3.3.1 Fatigue Failure Example

Suppose it has been observed that gun tube failures occur according to the lognormal distribution with μ = 7 and σ = 2 (remember μ and σ are the mean and SD of the ln (t) data). Find the reliability for a 1000 round mission and the hazard rate at 800 rounds. For this case, the variable t is the number of rounds.

$$R(t) = P\left(z > \frac{ln(t)-\mu}{\sigma}\right)$$

$$R(1000) = P\left(z > \frac{ln(1000)-7.0}{2.0}\right) = P\ (z > -0.045\ ) = 0.52$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{f(z)/\sigma t}{R(t)}$$

The numerator represents the transformation in the lognormal case.

$$h(800) = \frac{f(800)}{\sigma t R(800)} = \frac{f\left(z = \frac{ln(800)-7}{2}\right)}{(2)(800)R(800)} = \frac{f\left(z = \frac{ln\ 800 - 7}{2}\right)}{(2)(800)\ P\left(z > \frac{ln\ 800 - 7}{2}\right)}$$

$$= \frac{f\ (z = -0.16)}{1600\ P\ (z > -0.16)} = \frac{0.3939}{(1600)(0.5636)} = 0.0004 \text{ failures/round}$$

where P (z > -0.16) was interpolated from Table 5.3-1 and f (z = -0.16) was obtained from Table 5.3-2.

5.3.4 Exponential Distribution

This is probably the most important distribution in reliability work and is used almost exclusively for reliability prediction of electronic equipment (Ref. MIL-HDBK-217). It describes the situation wherein the hazard rate is constant which can be shown to be generated by a Poisson process. This distribution is valuable if properly used. It has the advantages of

(1) A single, easily estimated parameter (λ)

(2) Is mathematically very tractable

(3) Has fairly wide applicability

(4) Is additive - that is, the sum of a number of independent exponentially distributed variables is exponentially distributed

Some particular applications of this model include

(1) Items whose failure rate does not change significantly with age

(2) Complex and repairable equipment without excessive amounts of redundancy

(3) Equipment for which the early failures or "infant mortalities" have been eliminated by "burning in" the equipment for some reasonable time period

The failure density function is

$$f(t) = \lambda e^{-\lambda t} \qquad \text{for } t > 0, \qquad (5.37)$$

where λ is the hazard (failure) rate, and the reliability function is

$$R(t) = e^{-\lambda t} \qquad (5.38)$$

the mean life (θ) = $1/\lambda$, and, for repairable equipment, the MTBF = $\theta = 1/\lambda$.

5.3.4.1 Airborne Fire Control System Example

The mean time to failure (MTTF = θ, for this case) of an airborne fire control system is 10 hours. What is the probability that it will not fail during a 3 hour mission?

$$R(3) = e^{-\lambda t} = e^{-t/\theta} = e^{-3/10} = e^{-0.3} = 0.74$$

5.3.4.2 Computer Example

A computer has a constant error rate of one error every 17 days of continuous operation. What is the reliability associated with the computer to correctly solve a problem that requires 5 hours time? Find the hazard rate after 5 hours of operation.

MTTF = θ = 408 hours

$$\lambda = \frac{1}{\theta} = \frac{1}{408} = 0.0024 \text{ failure/hour}$$

$$R(5) = e^{-\lambda t} = e^{-(0.0024)(5)} = e^{-0.012} = 0.99$$

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda = 0.0024 \text{ failures/hours}$$

5.3.5 Gamma Distribution

The gamma distribution is used in reliability analysis for cases where partial failures can exist, i.e., when a given number of partial failures must occur before an item fails (e.g., redundant systems) or the time to second failure when the time to failure is exponentially distributed. The failure density function is

$$f(t) = \frac{\lambda}{\Gamma(\alpha)}(\lambda t)^{\alpha-1}e^{-\lambda t} \qquad \text{for } t > 0, \; \alpha > 0, \; \lambda > 0 \qquad (5.39)$$

where:

$$\lambda = \frac{\mu}{\sigma^2} \text{ and } \alpha = \lambda\mu \qquad (5.40)$$

μ = mean of data
α = standard deviation

and λ is the failure rate (complete failure) and α is the number of partial failures for complete failure or events to generate a failure. $\Gamma(\alpha)$ is the gamma function:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx \qquad (5.41)$$

which can be evaluated by means of standard tables (See Table 5.3-3).

When $(\alpha$-1) is a positive integer, $\Gamma(\alpha) = (\alpha$-1)!, which is usually the case for most reliability analysis, e.g., partial failure situation. For this case the failure density function is

$$f(t) = \frac{\lambda}{(\alpha-1)!}(\lambda t)^{\alpha-1}e^{-\lambda t} \qquad (5.42)$$

which, for the case of α = 1 becomes the exponential density function, previously described.

TABLE 5.3-3: GAMMA FUNCTION $\Gamma(n)$

$$\Gamma(n) = \int_0^\infty e^{-x} X^{n-1} dx$$

n	$\Gamma(n)$	n	$\Gamma(n)$	n	$\Gamma(n)$	n	$\Gamma(n)$
1.00	1.00000	1.25	.90640	1.50	.88623	1.75	.9196
1.01	.99433	1.26	.90440	1.51	.88659	1.76	.92137
1.02	.98884	1.27	.90250	1.52	.88704	1.77	.92376
1.03	.98355	1.28	.99072	1.53	.88757	1.78	.92623
1.04	.97844	1.29	.89904	1.54	.88818	1.79	.92877
1.05	.97350	1.30	.89747	1.55	.88887	1.80	.93138
1.06	.96874	1.31	.89600	1.56	.88964	1.81	.93408
1.07	.96415	1.32	.89464	1.57	.89049	1.82	.93685
1.08	.95973	1.33	.89338	1.58	.89142	1.83	.93969
1.09	.95546	1.34	.89222	1.59	.89243	1.84	.94261
1.10	.95135	1.35	1.89115	1.60	.89352	1.85	.94561
1.11	.94739	1.36	.89018	1.61	.89468	1.86	.94869
1.12	.94359	1.37	.88931	1.62	.89592	1.87	.95184
1.13	.93993	1.38	.88854	1.63	.89724	1.88	.95507
1.14	.93642	1.39	.88785	1.64	.89864	1.89	.95838
1.15	.93304	1.40	.88726	1.65	.90012	1.90	.96177
1.16	.92980	1.41	.88676	1.66	.90167	1.91	.96523
1.17	.92670	1.42	.88636	1.67	.90330	1.92	.96878
1.18	.92373	1.43	.88604	1.68	.90500	1.93	.97240
1.19	.92088	1.44	.88580	1.69	.90678	1.94	.97610
1.20	.91817	1.45	.88565	1.70	.90864	1.95	.97988
1.21	.91558	1.46	.88560	1.71	.91057	1.96	.98374
1.22	.91311	1.47	.88563	1.72	.91258	1.97	.98768
1.23	.91075	1.48	.88575	1.73	.91466	1.98	.99171
1.24	.90852	1.49	.88595	1.74	.91683	1.99	.99527
						2.00	1.00000

Note: $\Gamma(n+x) = (n - 1+x)(n - 2+x) \ldots (1 + x)\Gamma(1 + x)$
e.g., $\Gamma(3.15) = (2.15)(1.15)\Gamma(1.15)$

$$F(t) = \int_0^t \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} dt \tag{5.43}$$

If α is an integer, it can be shown by integration by parts that

$$F(t) = \sum_{k=\alpha}^{\infty} \frac{(\lambda t)^k \exp[-\lambda t]}{K!} \tag{5.44}$$

$$\text{Then } R(t) = 1 - F(t) = \sum_{K=0}^{n-1} \frac{(\lambda t)^k \exp[-\lambda t]}{K!} \tag{5.45}$$

$$\text{and } h(t) = \frac{f(t)}{R(t)} = \frac{\frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}}{\sum_{K=0}^{n-1} \frac{(\lambda t)^K \exp[-\lambda t]}{K!}} \tag{5.46}$$

The gamma distribution can also be used to describe an increasing or decreasing hazard (failure) rate. When $\alpha > 1$, h(t) increases; when $\alpha < 1$, h(t) decreases. This is shown in Figure 5.3-1.

5.3.5.1 Missile System Example

An antiaircraft missile system has demonstrated a gamma failure distribution with $\alpha = 3$ and $\lambda = 0.05$ (failures/hour). Determine the reliability for a 24 hour mission time and the hazard rate at the end of 24 hours.

$$R(t) = \frac{\lambda\alpha}{\Gamma(\alpha)} \int_t^{\infty} t^{\alpha-1} e^{-\lambda t} dt$$

Ordinarily, special tables of the Incomplete Gamma Function are required to evaluate the above integral. However, it can be shown that if α is an integer

$$R(t) = \sum_{k=0}^{\alpha-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tag{5.47}$$

which later in the section will be shown to be a Poisson distribution. Using Eq. (5.47)

$$R(24) = \sum_{k=0}^{2} \frac{[(0.05)(24)]^k e^{-(0.05)(24)}}{k!} = \sum_{k=0}^{2} \frac{(1.2)^k (0.3)}{k!}$$

$$= (0.3) + (1.2)(0.3) + \frac{(1.2)^2 (0.3)}{2} \quad = 0.3 + 0.36 + 0.216 = 0.88$$

$$h(t) = \frac{f(t)}{R(t)}$$

$$f(t) = \frac{\lambda}{(\alpha - 1)!} (\lambda t)^{\alpha - 1} e^{-\lambda t}$$

$$f(24) = \frac{0.05}{2} (1.2)^2 e^{-1.2} = (0.025)(0.434) = 0.011$$

$$h(24) = \frac{f(24)}{R(24)} = \frac{0.011}{0.88} = 0.012 \text{ failures/hour}$$

5.3.6 Weibull Distribution

The Weibull distribution is particularly useful in reliability work since it is a general distribution which, by adjustment of the distribution parameters, can be made to model a wide range of life distribution characteristics of different classes of engineered items.

One of the versions of the failure density function is

$$f(t) = \frac{\beta}{\eta} \left(\frac{t - \gamma}{\eta} \right)^{\beta - 1} \exp\left[-\left(\frac{t - \gamma}{\eta} \right)^{\beta} \right] \qquad (5.48)$$

where:

- β is the shape parameter
- η is the scale parameter or characteristic life (life at which 63.2% of the population will have failed)
- γ is the minimum life

In most practical reliability situations, γ is often zero (failure assumed to start at t = 0) and the failure density function becomes

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta - 1} \exp\left[-\left(\frac{t}{\eta} \right)^{\beta} \right] \qquad (5.49)$$

and the reliability and hazard functions become

$$R(t) = \exp\left[-\left(\frac{t}{h}\right)^{b}\right] \tag{5.50}$$

$$h(t) = \left(\frac{b}{h}\right)\left(\frac{t}{h}\right)^{b-1} \tag{5.51}$$

Depending upon the value of β, the Weibull distribution function can take the form of the following distributions as follows,

$\beta < 1$	Gamma	$\beta = 1$	Exponential
$\beta = 2$	Lognormal	$\beta = 3.5$	Normal (approximately)

Thus, it may be used to help identify other distributions from life data (backed up by goodness of fit tests) as well as being a distribution in its own right. Graphical methods are used to analyze Weibull failure data and are described in Section 8.

5.3.6.1 Example of Use of Weibull Distribution

The failure times of a particular transmitting tube are found to be Weibull distributed with $\beta = 2$ and $\eta = 1000$ hours. Find the reliability of one of these tubes for a mission time of 100 hours, and the hazard rate after a tube has operated successfully for 100 hours.

$$R(t) = \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]$$

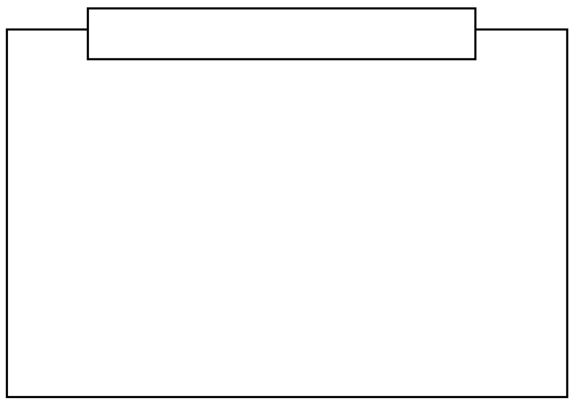
$$R(100) = \exp\left[-\left(\frac{100}{1000}\right)^{2}\right] = e^{-(0.1)^{2}} \approx 0.99$$

$$h(100) = \left(\frac{\beta}{\eta}\right)\left(\frac{t}{\eta}\right)^{\beta-1} = \left(\frac{2}{1000}\right)\left(\frac{100}{1000}\right)^{2-1} = 0.0002 \text{ failures/hour}$$

Note that in this example, the probability of success was the probability of obtaining a defective part.

5.3.7.1.2 Reliability Example

The binomial is useful for computing the probability of system success when the system employs partial redundancy. Assume a five channel VHF receiver as shown in Figure 5.3-3.



$$= \frac{5!}{3!2!}(0.9)^3(0.1)^2 + \frac{5!}{4!1!}(0.9)^4(0.1)^1 + \frac{5!}{5!0!}(0.9)^5(0.1)^0 = 0.99144$$

This is the probability that three or more of the five channels will survive the 24 hour operating period.

The problem can be solved another way, by subtracting the probability of three or more failures from one, e.g.:

$$P(S) = 1 - P(F)$$

$$= 1 - \sum_{x=(r+1)}^{n} \frac{n!}{n!(n-x)!} q^x p^{n-x}$$

$$= 1 - \left[\frac{5!}{3!2!}(0.1)^3(0.9)^2 + \frac{5!}{4!1!}(0.1)^4(0.9)^1 + \frac{5!}{5!0!}(0.1)^5(0.9)^0\right]$$

$$= 1 - 0.00856 = 0.99144 \text{ as before}$$

Note the change in notation (only) that x now represents the number of failures and q^x is the probability of x failures whereas before x represented the number of successes and p^x was the probability of x successes.

Computations involving the binomial distribution become rather unwieldy for even small sample sizes; however, complete tables of the binomial pdf and cdf are available in many statistics texts.

5.3.8 Poisson Distribution

This distribution is used quite frequently in reliability analysis. It can be considered an extension of the binomial distribution when n is infinite. In fact, it is used to approximate the binomial distribution when $n \geq 20$ and $p \leq 0.05$.

If events are Poisson distributed, they occur at a constant average rate and the number of events occurring in any time interval are independent of the number of events occurring in any other time interval. For example, the number of failures in a given time would be given by

$$f(x) = \frac{a^x e^{-a}}{x!} \tag{5.55}$$

where x is the number of failures and a is the expected number of failures.

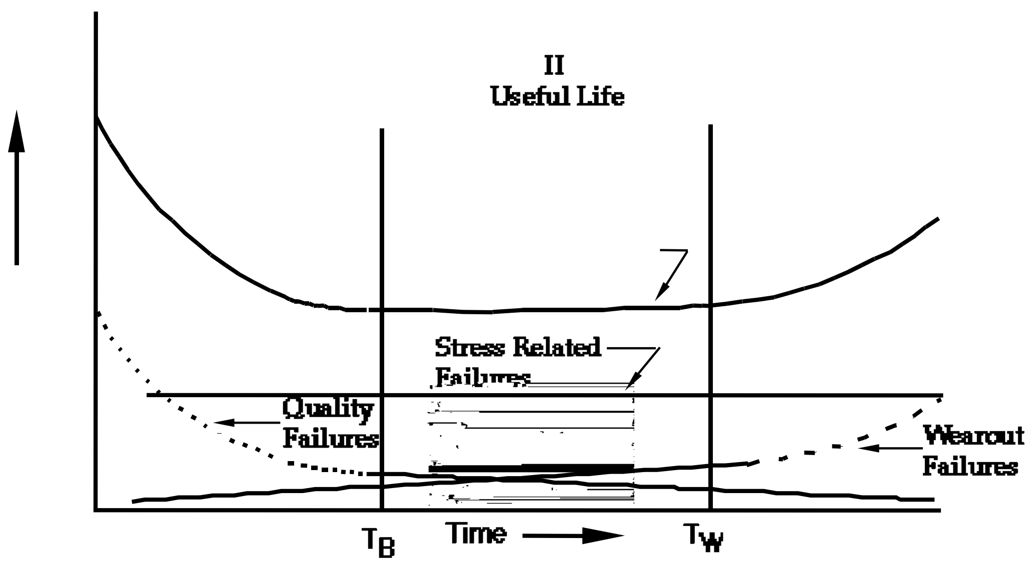
5.4 Failure Modeling

Failure modeling is a key to reliability engineering. Validated failure rate models are essential to the development of prediction techniques, allocation procedures, design and analysis methodologies, test and demonstration procedures/control procedures, etc. In other words, all of the elements needed as inputs for sound decisions to insure that an item can be designed and manufactured so that it will perform satisfactorily and economically over its useful life.

Inputs to failure rate models are operational field data, test data, engineering judgment, and physical failure information. These inputs are used by the reliability engineer to construct and validate statistical failure rate models (usually having one of the distributional forms described previously) and to estimate their parameters.

5.4.1 Typical Failure Rate Curve

Figure 5.4-1 shows a typical time versus failure rate curve for equipment. This is the "bathtub curve," which, over the years, has become widely accepted by the reliability community. It has proven to be particularly appropriate for electronic equipment and systems. The characteristic pattern is a period of decreasing failure rate (DFR) followed by a period of constant failure rate (CFR), followed by a period of increasing failure rate (IFR).



T_B = possible burn-in time T_W = wear begins

FIGURE 5.4-1: HAZARD RATE AS A FUNCTION OF AGE

Zone I is the infant mortality (DFR) period characterized by an initially high failure rate. This is normally the result of poor design, the use of substandard components, or lack of adequate controls in the manufacturing process. When these mistakes are not caught by quality control inspections, an early failure is likely to result. Early failures can be eliminated from the customer by “burn in” during which time the equipment is operated at stress levels equal to the intended actual operating conditions. The equipment is then released for actual use only when it has passed through the “burn-in” period.

Zone II, the useful life period, is characterized by an essentially constant failure rate (CFR). This is the period dominated by chance failures. Chance failures are those failures that result from strictly random or chance causes. They cannot be eliminated by either lengthy burn-in periods or good preventive maintenance practices. Equipment is designed to operate under certain conditions and up to certain stress levels. When these stress levels are exceeded due to random unforeseen or unknown events, a chance failure will occur. While reliability theory and practice is concerned with all three types of failures, its primary concern is with chance failures, since they occur during the useful life period of the equipment. Figure 5.4-1 is somewhat deceiving, since Zone II is usually of much greater length than Zones I or III. The time when a chance failure will occur cannot be predicted; however, the likelihood or probability that one will occur during a given period of time within the useful life can be determined by analyzing the equipment design. If the probability of chance failure is too great, either design changes must be introduced or the operating environment made less severe.

This CFR period is the basis for application of most reliability engineering design methods. Since it is constant, the exponential distribution of time to failure is applicable and is the basis for the design and prediction procedures spelled out in documents such as MIL- HDBK-217.

The simplicity of the approach utilizing the exponential distribution, as previously indicated, makes it extremely attractive. Fortunately, it is widely applicable for complex equipments and systems. If complex equipment consists of many components, each having a different mean life and variance which are randomly distributed, then the system malfunction rate becomes essentially constant as failed parts are replaced.

Thus, even though the failures might be wearout failures, the mixed population causes them to occur at random time intervals with a constant failure rate and exponential behavior. Figure 5.4-2 indicates this for a population of incandescent lamps in a factory. This has been verified for many equipments from electronic systems to bus motor overhaul rates.

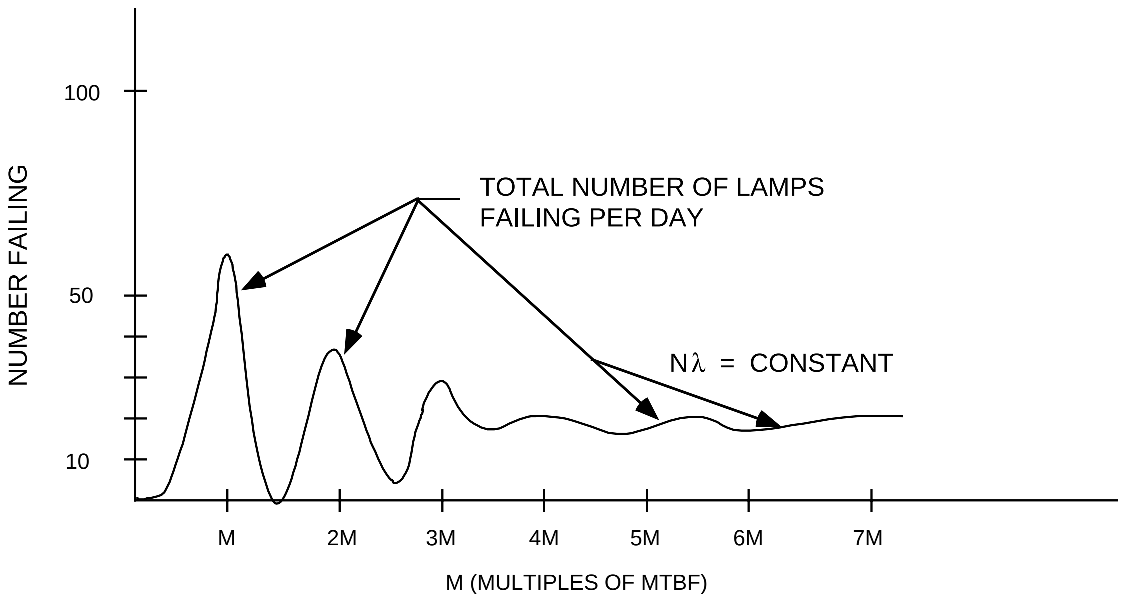


FIGURE 5.4-2: STABILIZATION OF FAILURE FREQUENCY

Zone III, the wearout period, is characterized by an IFR as a result of equipment deterioration due to age or use. For example, mechanical components such as transmission bearings will eventually wear out and fail, regardless of how well they are made. Early failures can be postponed and the useful life of equipment extended by good design and maintenance practices. The only way to prevent failure due to wearout is to replace or repair the deteriorating component before it fails.

Since modern electronic equipment is almost completely composed of semi-conductor devices which really have no short term wearout mechanism, except for perhaps electromigration, one might question whether predominantly electronic equipment will even reach Zone III of the bathtub curve.

From Figure 5.4-1, it can be seen that different statistical distributions might be used to characterize each zone. For example, the infant mortality period might be represented by gamma or Weibull, the useful life period by the exponential, and the wearout period by gamma or normal distributions.

The rest of this section will be devoted to models using the exponential distribution since it is applicable during the useful life period, which is the longest period of an equipment's life.

5.4.2 Reliability Modeling of Simple Structures

In this section, the reliability functions of some simple, structures will be derived. These functions are based upon the exponential distribution of time to failure.

SECTION 5: RELIABILITY/MAINTAINABILITY/AVAILABILITY THEORY

This is nothing more than the binomial expansion of $(R + Q)^n$

Thus,

$$P\text{ [at least (n-1) surviving]} = R^n + nR^{n-1}Q$$

$$P\text{ [at least (n-2) surviving]} = R^n + nR^{n-1}Q + \frac{n(n-1)R^{n-2}Q^2}{2!}$$

$$P\text{ [at least 1 surviving]} = 1 - Q^n$$

Let us look at the specific case of four display equipments which meet the previously mentioned assumptions.

$$(R + Q)^4 = R^4 + 4R^3Q + 6R^2Q^2 + 4RQ^3 + Q^4 = 1$$

from which

R^4 = P(all four will survive)

$4R^3Q$ = P(exactly 3 will survive)

$6R^2Q^2$ = P(exactly 2 will survive)

$4RQ^3$ = P(exactly 1 will survive)

Q^4 = P(all will fail)

We are usually interested in k out of n surviving.

$$R^4 + 4R^3Q = 1 - 6R^2Q^2 - 4RQ^3 - Q^4 = P\text{(at least 3 survive)}$$

$$R^4 + 4R^3Q + 6R^2Q^2 = 1 - 4RQ^3 - Q^4 = P\text{(at least 2 survive)}$$

$$R^4 + 4R^3Q + 6R^2Q^2 + 4RQ^3 = 1 - Q^4 = P\text{(at least 1 survives)}$$

If the reliability of each display for some time t is 0.9, what is the system reliability for time t if 3 out of 4 displays must be working?

$$R_S = R^4 + 4R^3Q = (0.9)^4 + 4(0.9)^3(0.1) = 0.6561 + 0.2916 = 0.9477$$

A similar example would be the case of launching 4 missiles, each of which had a probability of 0.9 of successfully hitting its target. What is the probability that at least 3 missiles will be on target? The procedure and result would be the same as the previous example.

For the case where all units have different reliabilities (or probabilities of success) the analysis becomes more difficult for the same assumptions. Let us look at the case of three units with reliabilities of R_1, R_2, and R_3, respectively. Then,

$$(R_1 + Q_1)(R_2 + Q_2)(R_3 + Q_3) = 1 \tag{5.63}$$

The above equation can be expanded to permit analysis as was done for the previous case of equal reliabilities. An easy way of bookkeeping is to set up Boolean truth tables where, $R_i = 1$, $Q_i = 0$, as follows:

1	2	3			
0	0	0	$Q_1 Q_2 Q_3$	=	all three fail
0	0	1	$Q_1 Q_2 R_3$	=	1 & 2 fail, 3 survives
0	1	0	$Q_1 R_2 Q_3$	=	1 & 3 fail, 2 survives
0	1	1	$Q_1 R_2 R_3$	=	1 fails, 2 & 3 survive
1	0	0	$R_1 Q_2 Q_3$	=	2 & 3 fail, 1 survives
1	0	1	$R_1 Q_2 R_3$	=	2 fails, 1 & 3 survive
1	1	0	$R_1 R_2 Q_3$	=	3 fails, 1 & 2 survive
1	1	1	$R_1 R_2 R_3$	=	all three survive

For the previous example, if we are not interested in which particular unit fails, we can set up expressions for at least 1, 2 or 3 units surviving. For example,

$$P(\text{at least 2 units surviving}) = R_1 R_2 R_3 + R_1 R_2 Q_3 + R_1 Q_2 R_3 + Q_1 R_2 R_3$$

The simple combinational reliability models developed in this section were, primarily, for illustrative purposes to demonstrate the basic theory involved. More complex examples are addressed in the references at the end of this section and in Section 7.

5.5 Bayesian Statistics in Reliability Analysis

Bayesian statistics have been increasingly used in reliability analysis. The advantage to the use of Bayesian statistics is that it allows prior information (e.g., predictions, test results, engineering judgment) to be combined with more recent information, such as test or field data, in order to arrive at a prediction/assessment of reliability based upon a combination of all available data. It also permits the reliability prediction/assessment to be continually updated as more and more test data are accumulated. The Bayesian approach is intuitively appealing to design engineers because it permits them to use engineering judgment, based upon prior experience with similar equipment designs, to arrive at an initial estimate of the reliability of a new design. It is particularly useful for assessing the reliability of new systems where only limited field data exists. For example, it can be argued that the result of a reliability test is not only information available on a product, but that information which is available prior to the start of the test, from component and subassembly tests, previous tests on the product, and even intuition based upon experience. Why should this information not be used to supplement the formal test result? Bayes' Theorem can be used to combine these results.

Thus, the basic difference between Bayesian and non-Bayesian (classical) approaches is that the former uses both current and prior data, whereas the latter uses current data only.

One of the main disadvantages to the use of the Bayesian approach is that one must be extremely careful in choosing the prior probabilities based upon part experience or judgment. If these are capriciously or arbitrarily chosen for Bayesian analysis, the end results of Bayesian analysis may be inaccurate and misleading. Thus, the key to the successful use of the Bayesian method resides in the appropriate choice of prior probability distributions. An objective prior such as existing test data is much better than a subjective prior based on opinion.

Bayes' analysis begins by assigning an initial reliability on the basis of whatever evidence is currently available. The initial prediction may be based solely on engineering judgment or it may be based on data from other similar types of items. Then, when additional test data is subsequently obtained, the initial reliabilities are revised on the basis of this data by means of Bayes' Theorem. The initial reliabilities are known as prior reliabilities in that they are assigned before the acquisition of the additional data. The reliabilities which result from the revision process are known as posterior reliabilities.

5.5.1 Bayes' Theorem

From basic probability theory, Bayes' Theorem is given by

$$\Pr[A|B] = \Pr[A] \frac{\Pr[B|A]}{\Pr[B]} \tag{5.64}$$

In the specific framework and context of reliability, the various terms in the equation may be motivated and defined as follows:

A — An hypothesis or statement of belief. ("The reliability of this component is 0.90.")

B — A piece of evidence, such as a reliability test result that has bearing upon the truth or credibility of the hypothesis. ("The component failed on a single mission trial.”)

Pr [A] The prior probability: the probability we assign to the hypothesis A before evidence B becomes available. (“We believe, based on engineering experience, that there is a 50-50 chance that the reliability of this component is about 0.90, as opposed to some-thing drastically lower, e.g., Pr [A] = 0.5.”)

Pr [B|A] The likelihood: the probability of the evidence assuming the truth of the hypothesis. (“The probability of the observed failure, given that the true component reliability is indeed 0.90, is obviously 0.10.”)

Pr [B] The probability of the evidence B, evaluated over the entire weighted ensemble of hypotheses A_i

Pr [A|B] The posterior probability of A, given the evidence B

The posterior probability is the end result of the application of Bayes' Equation. The following examples illustrate the use of Bayesian statistics in reliability analysis.

5.5.1.1 Bayes' Example (Discrete Distribution)

To demonstrate the use of Bayes' Equation within the framework of the binomial estimation of reliability, consider the following simplistic (but illustrative) example.

We wish to estimate the reliability of a simple pyrotechnic device which, upon being tested, either fires (success) or doesn't fire (failure). We have in the warehouse two lots of this component, one of which we have been assured has a reliability of R = 0.9 (that is, in the long term, 9 of 10 randomly selected components will work). The other lot supposedly contains only 50% good items. Unfortunately, we have lost the identity of which lot is which.

After randomly selecting one of the lots (such that the probability for each lot is 0.50), we then randomly select a single item from it (each item has equal chance of being chosen), which fails in test. What can be said about all this in the context of Bayesian analysis?

First, terms must be defined (see Figure 5.5-1).

A_1 “Lot chosen has R = 0.50”

A_2 “Lot chosen has R = 0.90”

Then, from above,

$$\Pr[A1] = 0.5, \quad \Pr[A2] = 0.5.$$

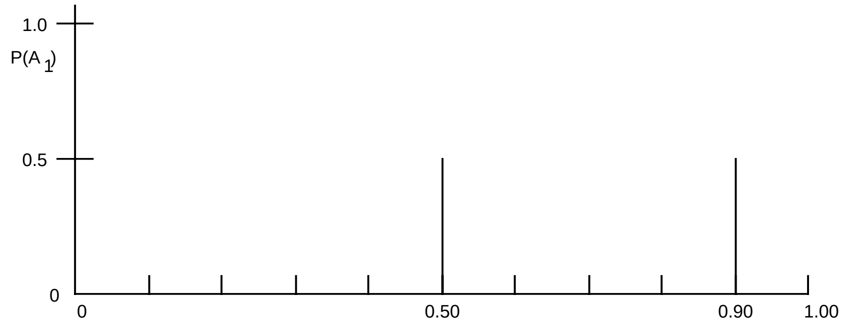


FIGURE 5.5-1: SIMPLE PRIOR DISTRIBUTION

Next, the test evidence must be considered. Therefore

B "One unit was tested and it failed."

The likelihoods required for Bayes' Equation are obviously

$$\Pr[B|A_1] = \Pr[\text{single test failure}|R = 0.5] = (1 - 0.5) = 0.5$$

$$\Pr[B|A_2] = \Pr[\text{single test failure}|R = 0.9] = (1 - 0.9) = 0.1$$

If A is partitioned into a set of states $[A_1, ..., A_n]$ and if $\Pr[A_i]$ and $\Pr[B|A_i]$ are known for each i; then Eq. (5.64) becomes

$$\Pr[A_i | B] = \Pr[A_i] \frac{\Pr[B | A_i]}{\Sigma \Pr[B | A_i] \bullet \Pr[A_i]} = \Pr[A_i] \frac{\Pr[B|A_i]}{\Pr[B]}$$

where the sum is over all n values of i. For this example, we have

$$\Pr[B] = \Pr[B|A_1] \Pr[A_1] + \Pr[B|A_2] \Pr[A_2] = 0.5(0.5) + 0.1(0.5) = 0.30.$$

Finally, all necessary inputs having been obtained, Bayes' Equation now yields

$$Pr[A_1|B] = \frac{Pr[A_1]\ Pr[B|A_1]}{Pr[B]} = \frac{0.5(0.5)}{0.30} = 0.833,$$

$$Pr[A_2|B] = \frac{Pr[A_2]\ Pr[B|A_2]}{Pr[B]} = \frac{0.5(0.1)}{0.30} = 0.167$$

The prior distribution in Figure 5.5-1 has been transformed, under the impact of a single trial resulting in failure, to the posterior distribution shown in Figure 5.5-2. The analyst may already be somewhat dubious that he has picked the lot with R = 0.9.

The process is usually a sequential one, i.e., as successive packets of new information (B_1, B_2, B_3, ...) become available, the posterior degree of belief in propositions A_i is successively modified by each new increment of information.

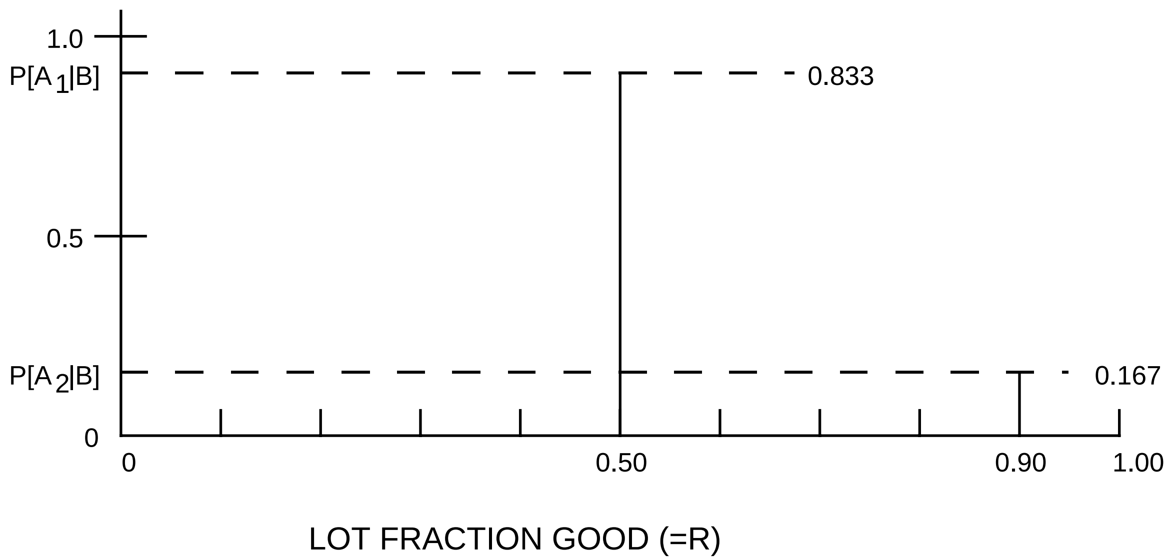


FIGURE 5.5-2: SIMPLE POSTERIOR DISTRIBUTION

Another way of visualizing this situation is by constructing a tree diagram like the one shown in Figure 5.5-3, where the probability of the final outcome "B" is given by the products of the probabilities corresponding to each individual branch.

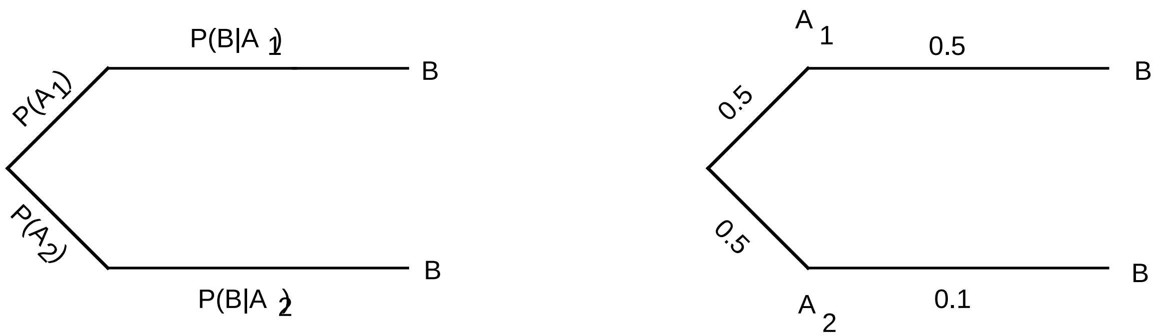


FIGURE 5.5-3: TREE DIAGRAM EXAMPLE

Eqs. (5.64 and 5.65) represent the prior failure rate and the prior variance. Let us assume that these are given by 0.02 and $(0.01)^2$, respectively. Assume that we then run a reliability test for 500 hours (t') and observe 14 failures (r'). What is the posterior estimate of failure rate?

The basic expression for the continuous posterior distribution is given by

$$f(\lambda|t) = \frac{f(l)\ f(t|l)}{f(t)} \tag{5.68}$$

where:

f(λ

Therefore,

$$t = \frac{\hat{\mu}_0}{\hat{\sigma}_0^2} = \frac{0.02}{(0.01)^2} = \frac{2 \times 10^{-2}}{1 \times 10^{-4}} = 200$$

$$r = \hat{\mu}_0 t = (2 \times 10^{-2})(200) = 4$$

Returning to the posterior gamma distribution, Eq. (5.69) we know that the posterior failure rate is

$$\hat{\mu}_1 = \frac{\text{shape parameter}}{\text{scale parameter}} = \frac{(r+r')}{(t+t')}$$

From the test data r' = 14, t' = 500, and we found that r = 4, and t = 200; thus

$$\hat{\mu}_1 = \frac{4 + 14}{200 + 500} = \frac{18}{700} = 0.0257$$

This compares with the traditional estimate of failure rate from the test result, 14/500 = 0.028. Thus, the use of prior information resulted in a failure rate estimate lower than that given by the test results.

5.6 Maintainability Theory

In reliability, one is concerned with designing an item to last as long as possible without failure; in maintainability, the emphasis is on designing an item so that a failure can be repaired as quickly as possible. The combination of high reliability and high maintainability results in high system availability; the theory of which is developed in Section 5.7.

Maintainability, then, is a measure of the ease and rapidity with which a system or equipment can be restored to operational status following a failure. It is a function of the equipment design and installation, personnel availability in the required skill levels, adequacy of maintenance procedures and test equipment, and the physical environment under which maintenance is performed.

As with reliability, maintainability parameters are also probabilistic and are analyzed by the use of continuous and discrete random variables, probabilistic parameters, and statistical distributions. An example of a discrete maintainability parameter is the number of maintenance actions completed in some time t, whereas an example of a continuous maintainability parameter is the time to complete a maintenance action.

5.6.1 Basic Concepts

A good way to look at basic maintainability concepts is in terms of functions which are analogous to those in reliability. They may be derived in a way identical to that done for reliability in the previous section by merely substituting t (time-to-restore) for t (time-to-failure), μ (repair rate) for λ (failure rate), and M(t) (probability of successfully completing a repair action in time t, or $P(T \leq t)$) for F(t) (probability of failing by age t, or $P(T \leq t)$). In other words, the following correspondences prevail in maintainability and reliability engineering functions.

(1) The time-to-failure probability density function (pdf) in reliability corresponds to the time-to-maintain pdf in maintainability.

(2) The failure rate function in reliability corresponds to the repair rate function in maintainability. Repair rate is the rate with which a repair action is performed and is expressed in terms of the number of repair actions performed and successfully completed per hour.

(3) The probability of system failure, or system unreliability, corresponds to the probability of successful system maintenance, or system maintainability. These and other analogous functions are summarized in Table 5.6-1.

Thus, as illustrated in Figure 5.6-1, maintainability can be expressed either as a measure of the time (T) required to repair a given percentage (P%) of all system failures, or as a probability (P) of restoring the system to operational status within a period of time (T) following a failure.

Some of the commonly used maintainability engineering terms are portrayed graphically in Figure 5.6-2 as a maintainability "function" derived as illustrated for the case where the pdf has a lognormal distribution. Points (1), (2), and (3) shown in the figure identify the mean, median, and maximum corrective time-to-repair, respectively.

TABLE 5.6-1: COMPARISON OF BASIC RELIABILITY AND MAINTAINABILITY FUNCTIONS

RELIABILITY	MAINTAINABILITY	
Time to Failure (pdf) $f(t)$	Time to Repair (pdf) $g(t)$	(5.70)
Reliability $R(t) = \int_t^\infty f(t)\,dt$	Maintainability $M(t) = \int_0^t g(t)\,dt$	(5.71)
Failure Rate $\lambda(t) = \frac{f(t)}{R(t)}$	Repair Rate $\mu(t) = \frac{g(t)}{1 - M(t)}$	(5.72)
Mean-Time-to-Failure $MTTF = \int_{-\infty}^{\infty} tf(t)\,dt$ $= \int_0^\infty R(t)\,dt$	Mean Time to Repair $MTTR = \int_{-\infty}^{\infty} t\,g(t)\,dt$	(5.73)
Pdf of Time to Failure $f(t) = \lambda(t) \bullet R(t)$ $= \lambda(t) \exp\left[-\int_0^t \lambda(t)dt\right]$	Pdf of Time to Repair $g(t) = \mu(t)\,(1 - M(t))$ $= \mu(t) \exp\left[-\int_0^t \mu(t)dt\right]$	(5.74)

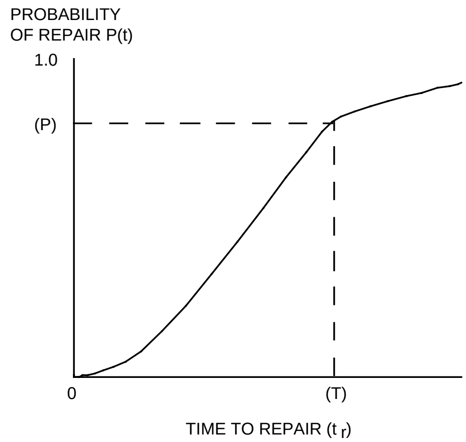


FIGURE 5.6-1: BASIC METHODS OF MAINTAINABILITY MEASUREMENT

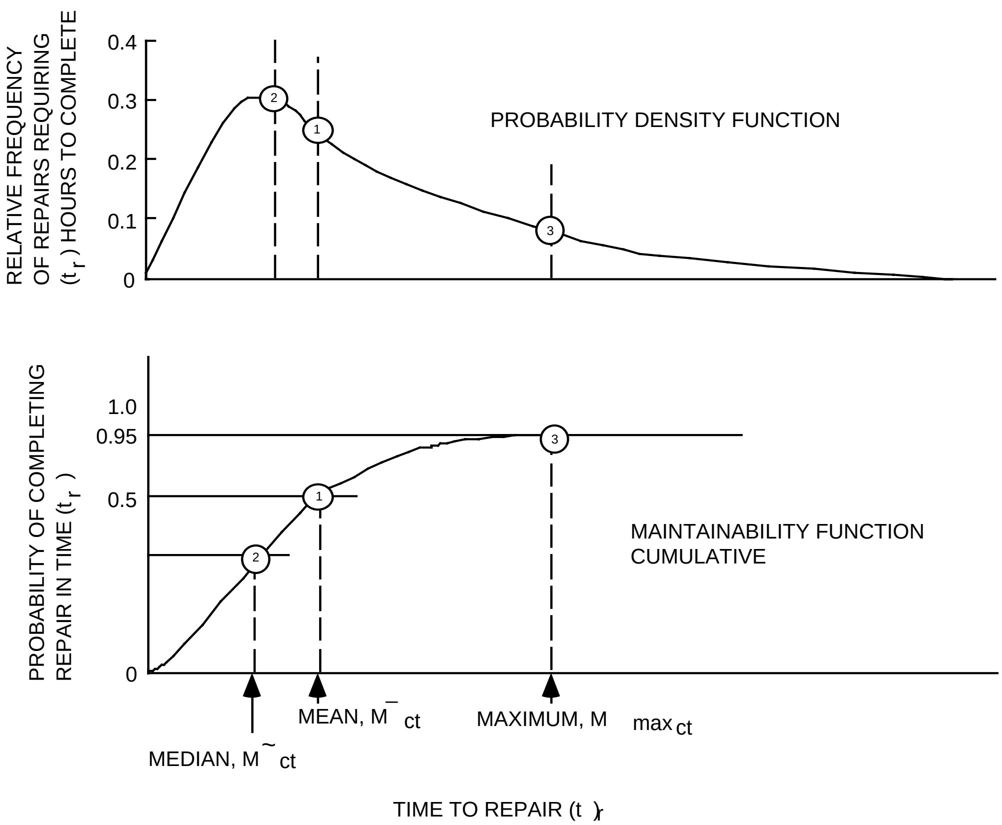


FIGURE 5.6-2: EXAMPLE MAINTAINABILITY FUNCTION DERIVED
FROM TIME-TO-REPAIR DISTRIBUTION

Points (1), (2), and (3) are defined as follows:

(1) Mean Time to Repair, $\overline{M}_{ct}$: The mean time required to complete a maintenance action, i.e., total maintenance downtime divided by total maintenance actions for a given period of time, given as

$$\overline{M}_{ct} = \frac{\Sigma(\lambda_i \overline{M}_{ct_i})}{\Sigma\lambda_i} \tag{5.75}$$

where: λ_i = failure rate for the i^{th} repairable element of the item for which maintainability is to be determined, adjusted for duty cycle, catastrophic failures, tolerance and inter-action failures, etc., which will result in deterioration of item performance to the point that a maintenance action will be initiated.

$\overline{M}_{ct_i}$ = average corrective time required to repair the i^{th} repairable element in the event of its failure.

(2) Median Time to Repair, $\tilde{M}_{ct}$: The downtime within which 50% of all maintenance actions can be completed.

(3) Maximum Time to Repair: The maximum time required to complete a specified, e.g., 95%, percentage of all maintenance actions.

These terms will be described in more detail in the following sections, in terms of the form that they take, given the statistical distribution of time-to-repair.

5.6.2 Statistical Distributions Used in Maintainability Models

A smaller number of statistical distributions is used for maintainability analysis than for reliability analysis. This may be due to the fact that maintainability has traditionally lagged reliability theory in development.

The most commonly used distributions for maintainability analysis have been the normal, lognormal, and exponential. Just as the exponential distribution has been the one most widely used in reliability analysis of equipment/systems, the lognormal distribution is the most commonly used for equipment/system maintainability analysis. A number of studies have validated the lognormal as being the most appropriate for maintainability analysis (Ref. [25]).

However, use of other distributions such as the Weibull and gamma is also possible, depending upon the analysis of the data and the use of “goodness of fit” tests.

Since the form and expressions for the more commonly used distributions were previously given in Section 5.2.2, this section will concentrate on the use of the normal, exponential, and lognormal distribution, and give examples of their use in maintainability analysis.

5.6.2.1 Lognormal Distribution

This is the most commonly used distribution in maintainability analysis. It applies to most maintenance tasks and repair actions comprised of several subsidiary tasks of unequal frequency and time duration.

The probability density function is given by

$$g(t = M_{ct_i}) = \frac{1}{M_{ct_i} S_{ln\,M_{ct}} \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{ln\,M_{ct_i} - \overline{ln\,M_{ct}}}{S_{ln\,M_{ct}}}\right)^2\right] \tag{5.76}$$

$$= \frac{1}{t\sigma_{t'}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t' - \overline{t'}}{\sigma_{t'}}\right)^2\right] \tag{5.77}$$

where:

$t = M_{ct_i}$ = repair time from each failure

$$\overline{ln\,M_{ct}} = \frac{\Sigma\, ln\,M_{ct_i}}{N}$$

$$S_{ln\,M_{ct}} = \sigma_{t'} = \sqrt{\frac{\Sigma\left(ln\,M_{ct_i}\right)^2 - (\Sigma ln\,M_{ct_i})^2/N}{N-1}} \tag{5.78}$$

$S_{ln\,M_{ct}} = \sqrt{\dfrac{\Sigma t_i'^2 - (\Sigma t_i')^2/N}{N-1}}$ = standard deviation of *ln* of repair times.

$t' = ln\,M_{ct_i} = ln\,t$

$$\overline{t}\,' = \overline{\ln M_{ct}} = \frac{\Sigma t_i'}{N}$$

N = number of repair actions (5.79)

The mean time to repair is given by

$$\text{MTTR} = \overline{M_{ct}} = \overline{t} = \int_0^\infty t\, g(t = M_{ct_i})\, dt \tag{5.80}$$

(also see Eq. (5.76))

$$= \exp\left[\overline{\ln M_{ct}} + \frac{1}{2}(S_{ln} M_{ct})^2\right] \tag{5.81}$$

$$= \exp\left[\overline{t}' + \frac{1}{2}(\sigma_{t'})^2\right] \tag{5.82}$$

The median time to repair is given by

$$\tilde{M}_{ct} = \breve{t} = antiln\ \frac{\Sigma\lambda_i\ \overline{\ln M_{ct}}}{\Sigma\lambda_i} \tag{5.83}$$

$$= \exp\left(\overline{\ln M_{ct_i}}\right) \tag{5.84}$$

$$= \exp\left(\overline{t'}\right) \tag{5.85}$$

The maximum time to repair is given by

$$M_{max_{ct}} = t_{max} = antiln\left(\overline{\ln M_{ct}} + \phi S_{ln} M_{ct}\right) \tag{5.86}$$

$$= antiln\left[\overline{t}' + z(t'_{1-\alpha})\sigma_{t'}\right] \tag{5.87}$$

where $\phi = z(t'_{1-\alpha})$ = value from normal distribution function corresponding to the percentage point $(1-\alpha)$ on the maintainability function for which $M_{max_{ct}}$ is defined.

Most commonly used values of ϕ or $z(t'_{1-\alpha})$ are shown in Table 5.6-2.

TABLE 5.6-2: VALUES OF ϕ OR $Z(T'_{(1-\alpha)})$ MOST COMMONLY USED IN MAINTAINABILITY ANALYSIS

$1-\alpha$	ϕ or $Z(t'_{(1-\alpha)})$
0.80	0.8416
0.85	1.036
0.90	1.282
0.95	1.645
0.99	2.326

Following is an example of maintainability analysis of a system which has a lognormal distribution of repair times.

5.6.2.1.1 Ground Electronic System Maintainability Analysis Example

Given the active repair times data of Table 5.6-3 on a ground electronic system find the following:

(1) The probability density function, g(t)

(2) The MTTR of the system

(3) The median time to repair the system

(4) The maintainability function

(5) The maintainability for a 20 hour mission

(6) The time within which 90% and 95% of the maintenance actions are completed

(7) The repair rate, u(t), at 20 hours

TABLE 5.6-3:   TIME-TO-REPAIR DATA ON A GROUND ELECTRONIC SYSTEM

Group No. j	Times to Repair t_j (hr.)	Frequency of Observation n_j
1	0.2	1
2	0.3	1
3	0.5	4
4	0.6	2
5	0.7	3
6	0.8	2
7	1.0	4
8	1.1	1
9	1.3	1
10	1.5	4
11	2.0	2
12	2.2	1
13	2.5	1
14	2.7	1
15	3.0	2
16	3.3	2
17	4.0	2
18	4.5	1
19	4.7	1
20	5.0	1
21	5.4	1
22	5.5	1
23	7.0	1
24	7.5	1
25	8.8	1
26	9.0	1
27	10.3	1
28	22.0	1
N' = 29	24.5	1

1. Probability Density Function of g(t)

To determine the lognormal pdf of the times-to-repair given in Table 5.6-3, the values of $\bar{t}'$ and $\sigma_{t'}$ should be calculated from

$$\bar{t}' = \frac{\sum_{j=1}^{N'} n_j t'_j}{\sum_{j=1}^{N'} n_j} \tag{5.88}$$

where n_j is the number of identical observations given in the third column of Table 5.6-3, N' is the number of different-in-value observed times-to-repair, or number of data groups, which for this problem is N' = 29, given in the second column of Table 5.6-3, and N is the total number of observed times-to-repair,

$$N = \sum_{i=1}^{N'} n_j$$

which, for this example, is 46,

and

$$\sigma_{t'} = \left[\frac{\sum_{i=1}^{N}(t'_i)^2 - N(\bar{t}')^2}{N-1}\right]^{\frac{1}{2}} = \left[\frac{\sum_{j=1}^{N'} n_j(t'_j)^2 - N(\bar{t}')^2}{N-1}\right]^{\frac{1}{2}} \tag{5.89}$$

To facilitate the calculations, Table 5.6-4 was prepared. From Table 5.6-4, $\bar{t}'$ and $\sigma_{t'}$, are obtained as follows:

$$\bar{t}' = \frac{\sum_{j=1}^{N'} n_j t'_j}{\sum_{j=1}^{N'} n_j} = \frac{30.330439}{46} = 0.65879$$

TABLE 5.6-4: CALCULATIONS TO DETERMINE $\bar{t'}$ AND σ_T
FOR THE DATA IN TABLE 5.6-3

j	t_j	$\ln t_j = t'_j$	$(t'_j)^2$	n_j	$n_j t'_j$	$n_j (t'_j)^2$
1	0.2	-1.60944	2.59029	1	-1.60944	2.59029
2	0.3	-1.20397	1.44955	1	-1.20397	1.44955
3	0.5	-0.69315	0.48045	4	-2.77260	1.92180
4	0.6	-0.51083	0.26094	2	-1.02166	0.52188
5	0.7	-0.35667	0.12721	3	-1.07001	0.38163
6	0.8	-0.22314	0.04979	2	-0.44628	0.09958
7	1.0	0.00000	0.00000	4	0.00000	0.00000
8	1.1	0.09531	0.00908	1	0.09531	0.00908
9	1.3	0.26236	0.06884	1	0.26236	0.06884
10	1.5	0.40547	0.16440	4	1.62188	0.65760
11	2.0	0.69315	0.48045	2	1.38630	0.96090
12	2.2	0.78846	0.62167	1	0.78846	0.62167
13	2.5	0.91629	0.83959	1	0.91629	0.83959
14	2.7	0.99325	0.98655	1	0.99325	0.98655
15	3.0	1.09861	1.20695	2	2.19722	2.41390
16	3.3	1.19392	1.42545	2	2.38784	2.85090
17	4.0	1.38629	1.92181	2	2.77258	3.84362
18	4.5	1.50408	2.26225	1	1.50408	2.26225
19	4.7	1.54756	2.39495	1	1.54756	2.39495
20	5.0	1.60994	2.59029	1	1.60994	2.59029
21	5.4	1.68640	2.84394	1	1.68640	2.84394
22	5.5	1.70475	2.90617	1	1.70475	2.90617
23	7.0	1.94591	3.78657	1	1.94591	3.78657
24	7.5	2.01490	4.05983	1	2.01490	4.05983
25	8.8	2.17475	4.72955	1	2.17475	4.72955
26	9.0	2.19722	4.82780	1	2.19722	4.82780
27	10.3	2.33214	5.43890	1	2.33214	5.43890
28	22.0	3.09104	9.55454	1	3.09104	9.55454
29	24.5	3.19867	10.23151	1	3.19867	10.23151
Sum				46	30.30439	75.84371

$$\sum_{j=1}^{N'=29} n_j = 46 = N \qquad \sum_{j=1}^{N'} n_j t'_j = 30.30439 \qquad \sum_{j=1}^{N'} n_j \left(t'^{\,2}_j\right) = 75.84371$$

and from Eq. (5.89)

$$\sigma_{t'} = \frac{46\ 0\ 65879^{2}}{46 \quad 1}^{\frac{1}{2}}$$

2. MTTR (Mean Time to Repair) of the System

The mean time to repair of the system, $\bar{t}$, is obtained from Eq. (5.83).

$$\bar{t} = \exp\left[\bar{t'} + \frac{1}{2}(\sigma_{t'})^2\right] = \exp\left[0.65879 + \frac{1}{2}(1.11435)^2\right] = 3.595 \text{ hr.}$$

3. Median Time to Repair

The median of the times-to-repair the system, $\hat{t}$, is obtained from Eq. (5.85)

$$\hat{t} = \exp\left(\bar{t'}\right) = e^{0.65879} = 1.932 \text{ hr.}$$

This means that in a large sample of t's, half of the t's will have values smaller than $\hat{t}$, and the other half will have values greater than $\hat{t}$. In other words, 50% of the repair times will be $\leq \hat{t}$.

4. Maintainability Function M(t)

The maintainability of a unit can be evaluated as follows, using Eq. (5.71):

$$M(t_1) = \int_0^{t_1} g(t)\,dt = \int_{-\infty}^{t'_1} g(t')\,dt' = \int_{-\infty}^{z(t'_1)} f(z)\,dz \tag{5.90}$$

where $t' = \ln$

The maintainability function for the system, M(t), from (5.90) is:

$$M(t) = \int_{-\infty}^{z(t')} \phi(z)\, dz$$

where:

$$z(t') = \frac{t' - \overline{t}\,'}{\sigma_{t'}}$$

$$t' = \ln t$$

From the data in Table 5.6-3 we previously calculated

$$\overline{t}\,' = 0.65879$$

$$\sigma_{t'} = 1.11435$$

The quantified M(t) is shown in Figure 5.6-5. The values were obtained by inserting values for t' = *ln* t into the expression,

$$z(t') = \frac{t' - 0.65879}{1.11435}$$

solving for z(t'), and reading the value of M(t) directly from the standard normal tables in Table 5.3-3.

5. <u>Maintainability for a 20 Hour Mission</u>

$$M(20) = \int_{-\infty}^{z(\ln 20)} \phi(z)\, dz$$

where *ln* 20 = 2.9957

and

$$z(\ln 20) = \frac{2.9957 - 0.65879}{1.111435} = 2.0972$$

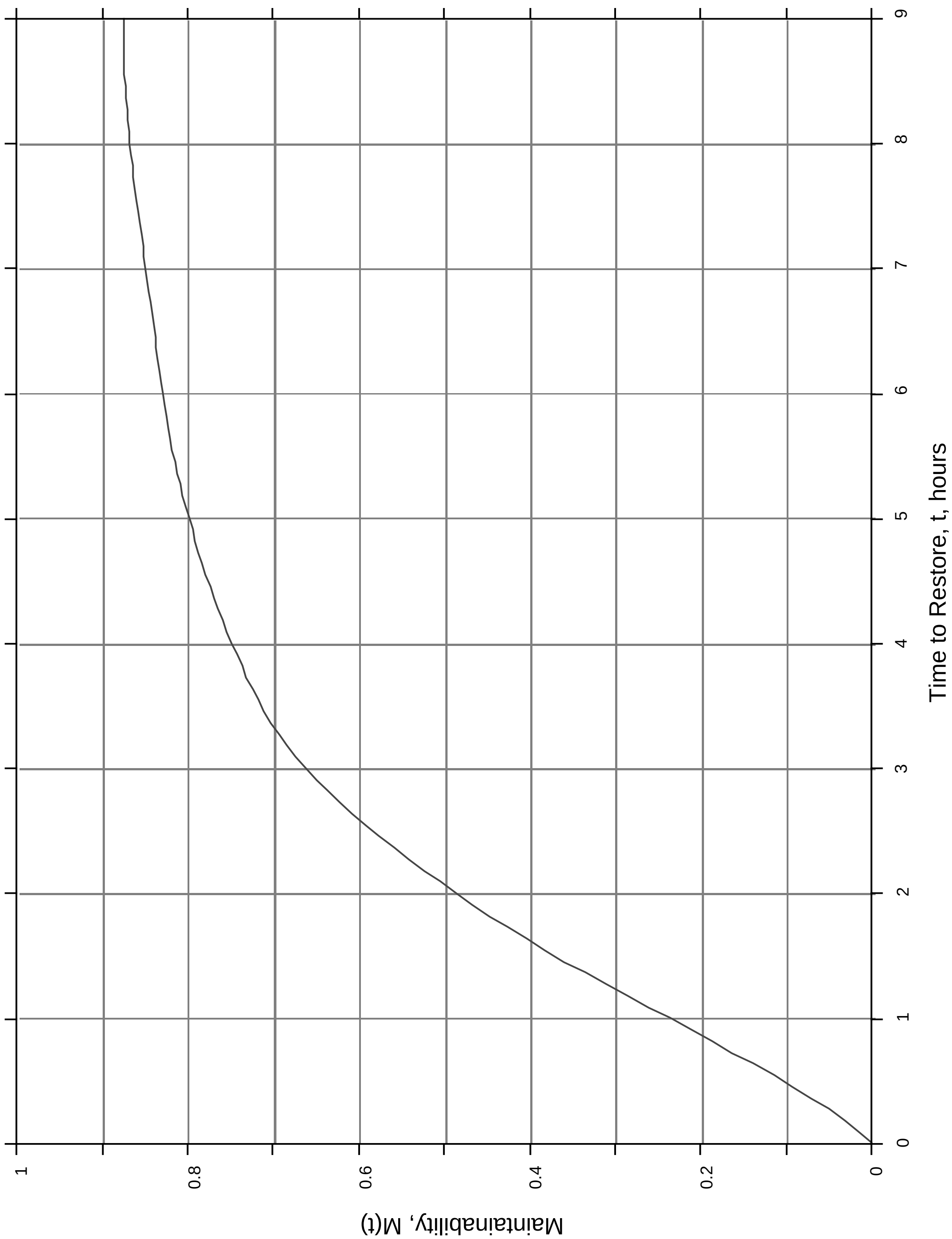


FIGURE 5.6-5: PLOT OF THE MAINTAINABILITY FUNCTION FOR THE TIMES-TO-REPAIR DATA OF EXAMPLE 2

From Table 5.3-3 we find that for z = 2.0972

$$M(20) = \int_{-\infty}^{2.0972} \phi(z)\, dz = 1 - 0.018 = 0.982 \text{ or } 98.2\%$$

6. The time within which 90% and 95% of the Maintenance Actions are Completed ($M_{max_{ct}}$)

This is the time $t_{1-\alpha}$ for which the maintainability is $1 - \alpha$, or

$$M(t_{1-\alpha}) = P(t \leq t_{1-\alpha}) = \int_0^{t_{1-\alpha}} g(t)\, dt = \int_{-\infty}^{t'_{1-\alpha}} g(t')\, dt' = \int_{-\infty}^{z(t'_{1-\alpha})} \phi(z)\, dz \qquad (5.91)$$

and

$$z(t'_{1-\alpha}) = \frac{t'_{1-\alpha} - \bar{t}'}{\sigma_{t'}} \qquad (5.92)$$

The commonly used maintainability, or $(1-\alpha)$, values are 0.80, 0.85, 0.90, 0.95, and 0.99. Consequently, the $z(t'_{1-\alpha})$ values which would be used most commonly would be those previously given in Table 5.6-2. Using Eq. (5.92) the time $t_{1-\alpha}$ would then be calculated from

$$t'_{1-\alpha} = \bar{t}' + z(t'_{1-\alpha}) \bullet \sigma_{t'}$$

or

$$t'_{1-\alpha} = antiln\, (t'_{1-\alpha}) = antiln\left[\bar{t}' + z(t'_{1-\alpha})\sigma_{t'}\right] \qquad (5.93)$$

Thus, for 90% $M_{max_{ct}}$, from the previously obtained value of $\bar{t}'$ and $\sigma_{t'}$

$$t_{0.90} = antiln\left[\bar{t}' + z(t'_{0.90})\sigma_{t'}\right] = antiln\left[0.65879 + 1.282(1.11435)\right]$$

$$= antiln\ (2.08737) = 8.06 \text{ hrs.}$$

For 95% $M_{max_{ct}}$

$$t_{0.95} = antiln\left[0.65879 + 1.645(1.11435)\right] = antiln\,(2.491896) = 12.08 \text{ hrs.}$$

7. Repair Rate at t = 20 hours

Using Eq. (5.60) and substituting the values for g(20) from Table 5.6-5 and the previously calculated value for M(20)

$$\mu(20) = \frac{g(20)}{1 - M(20)} = \frac{0.00199}{1 - 0.982} = \frac{0.00199}{0.018} = 0.11 \text{ repairs/hr.}$$

5.6.2.2 Normal Distribution

The normal distribution has been adequately treated in Section 5.3.2.1 in the discussion on reliability theory. The same procedures and methodology apply for maintainability if one merely uses repair time for t, mean repair time for μ, and standard deviation of repair times for σ.

In maintainability, the normal distribution applies to relatively straightforward maintenance tasks and repair actions (e.g., simple removal and replacement tasks) which consistently require a fixed amount of time to complete. Maintenance task times of this nature are usually normally distributed, producing a probability density function given by

$$g(t = M_{ct}) = \frac{1}{S_{M_{ct}}\sqrt{2\pi}} \exp\left[\frac{-(M_{ct_i} - \overline{M}_{ct})^2}{2(S_{M_{ct}})^2}\right] \tag{5.94}$$

where:

M_{ct_i} = repair time for an individual maintenance action

$\overline{M}_{ct} = \frac{\Sigma(M_{ct_i})}{N}$ = average repair time for N observations

$S_{M_{ct}} = \sqrt{\frac{\Sigma(M_{ct_i} - \overline{M}_{ct})^2}{N-1}}$ = standard deviation of the distribution of repair times, based on N observations

N = number of observations

The mean time to repair ($\overline{M}_{ct}$) is given by

$$\overline{M}_{ct} = \frac{\Sigma M_{ct_i}}{N} \tag{5.95}$$

The median time to repair ($\tilde{M}_{ct}$) is given by

$$\tilde{M}_{ct} = \frac{\Sigma M_{ct_i}}{N} \tag{5.96}$$

which is equal to the mean time to repair because of the symmetry of the normal distribution (see Fig. 5.3-1).

The maximum time to repair is given by

$$M_{max_{ct}} = \overline{M}_{ct} + \phi S_{M_{ct}} \tag{5.97}$$

where:

ϕ = $z(t_{1-\alpha})$

= value from normal distribution function corresponding to the percentage point ($1-\alpha$) on the maintainability function for which $M_{max_{ct}}$ is defined. Values of ϕ as a function of ($1-\alpha$) are shown in Table 5.6-6. Note that this is the same as Table 5.6-2 with rounded-off values.

TABLE 5.6-6: VALUES OF ϕ FOR SPECIFIED α

$1-\alpha$	ϕ or $z(t_{1-\alpha})$
95%	1.65
90%	1.28
85%	1.04
80%	0.84

5.6.2.2.1 Equipment Example

An equipment whose repair times are assumed to be normally distributed was monitored and the following repair times were observed (in minutes):

6.5, 13.25, 17.25, 17.25, 19.75, 23, 23, 24.75, 27.5, 27.5, 27.5, 32, 34.75, 34.75, 37.5, 37.5, 40.25, 42.5, 44.75, 52

Find the following parameters.

(1) The pdf of g(t) and its value at 30 minutes

(2) The MTTR and median times to repair

(3) The maintainability for 30 minutes

(4) The time within which 90% of the maintenance actions are completed

(5) The repair rate, u(t), at 30 minutes

(1) Pdf of g(t)

$$\overline{M}_{ct} = \frac{\Sigma M_{ct_i}}{N} = \frac{583.25}{20} = 29.16 \text{ minutes}$$

$$S_{M_{ct}} = \sqrt{\frac{\Sigma(M_{ct_i} - \overline{M}_{ct})^2}{N-1}}$$

6.4.5.3 Reliability Prediction Methods

Four prediction methods are described as follows:

6.4.5.3.1 Similar Item Prediction Method

Several techniques have been developed and used in performing very early predictions of item reliability before any characteristics of the system design have been established. The most basic of these techniques involves a simple estimate of item reliability in terms of MTBF, failure rate, or similar parameters, based on experience gained from operational items of similar function.

In general, these similar item prediction techniques involve the following steps:

(1) Defining the new item in terms such as general equipment type (e.g., radar), operational use (e.g., ground based) and other known characteristics.

(2) Identifying an existing item or class of equipment that most nearly compares with the new item.

(3) Obtaining and analyzing historical data generated during operation of the existing equipment to determine as nearly as possible the reliability of the items under the stated operating environment.

(4) Drawing conclusions concerning the level of reliability that will be demonstrated by the new items. Such conclusions assume that similar equipment will exhibit similar reliability and that reliability achievement evolves in an orderly manner from one generation of equipments to the next. These reliability prediction techniques permit very early estimation of the failure rate of a new item based on experience gained from operational items of similar function. The accuracy of the estimates, however, depends on the quality of historical data and the similarity between the existing and new equipments. If the technology of the new items is too advanced, then the operational data for the old items will not be relevant and another technique will have to be considered.

The similar item prediction method utilizes specific experience on similar items. The more rapid way of estimating reliability is to compare the item under consideration with a similar item whose reliability has previously been determined by some means and has undergone field evaluation. The method has a continuing and meaningful application for items undergoing orderly evolution. Not only is the contemplated new design similar to the old design, but small differences can be easily isolated and evaluated. In addition, difficulties encountered in the old design are signposts to improvements in the new design.

Major factors for a direct comparison of similar items should include:

(1) Item physical and performance comparison

(2) Design similarity

(3) Manufacturing similarity

(4) Similarity of the service use profile (logistic, operational, and environmental)

(5) Program and project similarity

(6) Proof of reliability achievement

The validity of the similar item method is dependent upon the degree of equivalence between the items and not simply the generic term used to describe the items. For example, although both are power supplies (generic type), the achieved reliability of a ten watt power supply should not normally be used as a prediction method for a proposed one kilowatt power supply as the much higher power level of the proposed power supply may result in much lower reliability achievement due to significant design differences and stresses. A comparison may be made if there are scale factors to realistically relate reliability with item parameters such as power levels.

An example of this technique is: a new computer product which is composed of a processor, a display, a modem and a keyboard is expected to operate in a 20°C environment. Data on similar items indicates mean-time-between-failure (MTBF) values as shown in the second column of Table 6.4-7. The similar item data is for a computer operating in a 30°C environment. If a 30% reliability improvement factor (as a result of improved technology) is expected, what MTBF can we expect?

TABLE 6.4-7: RELIABILITY ANALYSIS SIMILAR ITEM

Item	Similar Data MTBF (hrs.)	Temperature Factor*	Improvement Factor	New Product MTBF (Hrs.)
Processor	5,000	1.1	1.3	7,150
Display	15,000	1.1	1.3	21,450
Modem	30,000	1.1	1.3	42,900
Keyboard	60,000	1.1	1.3	85,800
System	3,158			4,516

*Each item MTBF is corrected using temperature conversion factors from the "Reliability Toolkit: Commercial Practices Edition," page 176 (Ref. [8]).

Each item MTBF is corrected for the change in temperature of 30°C to 20°C. Technology improvement factors are also included and the system MTBF is calculated using the expression:

$$MTBF_s = \sum_{i}^{n} \frac{1}{\lambda_i}$$

where:

$MTBF_s$ = mean-time-between-failure of the system

λ_i = failure rate of the i component which equals $1/MTBF_i$

6.4.5.3.2 <u>Parts Count Prediction Method</u>

This technique is used when one has a "feel" for the number of component parts (actual or estimated) by class or type that will be used in an equipment/system but does not have enough data as to the stresses to which each part will be subjected in the final design. It involves counting the number of parts of each class or type, multiplying this number by the generic failure rate for each part class or type, and summing these products to obtain the failure rate for the equipment. The procedure distinguishes a part class as being all parts of a given function (e.g., resistors, capacitors, transformers). Part types are used to further define parts within a class (e.g., fixed composition resistors, fixed wire wound resistors).

This method is used in the preliminary design stage when the number of parts in each generic type class such as capacitors, resistors, etc., are reasonably fixed and the overall design complexity is not expected to change appreciably during later stages of development and production. The parts count method assumes the time to failure of the parts is exponentially distributed (i.e., a constant failure rate).

The item failure rate can be determined directly by the summation of part failure rates if all elements of the item reliability model are in series or can be assumed in series for purposes of an approximation. In the event the item reliability model consists of non-series elements (e.g., redundancies, alternate modes of operation), item reliability can be determined either by considering only the series elements of the model as an approximation or by summing part failure rates for the individual elements and calculating an equivalent series failure rate for the non-series elements of the model.

The information needed to support the parts count method includes:

(1) Generic part types (including complexity for microelectronics),

(2) Part quantity,

(3) Part quality levels (when known or can be assumed), and

(4) Item environment.

The general expression for item failure rate with this method is:

$$\lambda_{ITEM} = \sum_{i=1}^{n} N_i \left(\lambda_{Gi} \pi_{Qi}\right) \quad (6.42)$$

where:

λ_{ITEM}	=	total failure rate
λ_{Gi}	=	generic failure rate for the i^{th} generic part
π_{Qi}	=	quality factor for the i^{th} generic part
N_i	=	quantity of i^{th} generic part
n	=	number of different generic part categories

Equation 6.42 applies to an entire item being used in one environment. If the item comprises several units operating in different environments (such as avionics with units in airborne, inhabited, fighter (A_{IF}) and uninhabited, fighter (A_{UF}) environment), then equation 6.42 should be applied to the portions of the item in each environment. These “environment-item” failure rates should be added to determine total item failure rate.

Quality factors are to be applied to each part type where quality level data exists or can be reasonably assumed. Multi-quality levels and data exist for parts, such as microelectronics, discrete semiconductors, and for established reliability (ER) resistors and capacitors. For other parts such as nonelectronics, $\pi_Q = 1$ providing that parts are procured in accordance with applicable parts specifications.

The generic (average) failure rate (λ_{Gi}) and the quality factor (π_{Qi}) can be obtained from the latest version of MIL-HDBK-217 (Ref. [11]) or manufacturer's data. MIL-HDBK-217 contains a number of tables of generic failure rates for various classes and types of parts, as well as the associated quality factors. Tables 6.4-8 and 6.4-9 (taken from MIL-HDBK-217F, Notice 2), are specific examples of generic failure rates and quality factors for diodes and transistors.

An example of how this technique might be applied to predict the MTBF and reliability of a mobile electronic receiver is shown in Figure 6.4-4. The part failure rates for a ground mobile environmental condition are presented from MIL-HDBK-217 for the various part types.

6.4.5.3.3 Parts Stress Analysis Prediction Method

The previous method described was based upon average failure rates for each component part type. It is well known that part failure rates vary significantly with applied stresses, sometimes by several orders of magnitude. For example, a 110 volt light bulb does not operate very long when subjected to 220 volts. It is this interaction between strength of the component and the stress level at which the component operates which determines the failure rate of a component in a given situation. Thus, at different stress levels component parts assume different failure rates. This is the rationale for the stress analysis prediction technique. This technique is based upon a knowledge of the stress to which the part will be subjected, e.g., temperature, humidity, vibration, etc., and the effect of those stresses on the part's failure rate. Some of the factors that influence part reliability, for a sample of part types, are shown in Table 6.4-10.

TABLE 6.4-12: BIPOLAR COMPLEXITY FAILURE RATE C1

Digital		Linear	
No. Gates	C_1	No Transistors	C_1
1 to 100	.0025	1 to 100	.01
101 to 1,000	.005	101 to 300	.02
1,001 to 3,000	.010	301 to 1,000	.04
3,001 to 10,000	.020	1,001 to 10,000	.06
10,001 to 30,000	.040		
30,001 to 60,000	.080		

STEP 3: Determine junction temperature: The standard junction temperature is calculated using the following relationship:

$$T_J = T_C + \theta_{JC} P$$

where:

T_J = junction temperature in degrees centigrade

T_C = case temperature in degrees centigrade

θ_{JC} = junction to case thermal resistance in degrees centigrade per watt

P = power dissipated in watts

Values for the factors are given, so

$$T_J = 45 + 11(.20)$$
$$= 47.2°C$$

STEP 4: Determine the temperature acceleration factor, π_T from the temperature equation as stated in MIL-HDBK-217. The equation is:

$$\pi_T = 0.1 \exp\left[-A\left(\frac{1}{T_J + 273} - \frac{1}{298}\right)\right]$$

where:

A = temperature coefficient, 4642

T_J = junction temperature (°C)

π_T = .29

STEP 5: Determine the packaging factor C_2 given a 64 pin non-hermetic dual-in-line package from the equation in MIL-HDBK-217. The equation is:

$$C_2 = 3.6 \times 10^{4} (N_p)^{1.08}$$

where:

$$C_2 = 3.6 \times 10^{4} (64)^{1.08}$$

$$= .032$$

STEP 6: Find the environmental factor from MIL-HDBK-217 which is shown in Table 6.4-13. For ground fixed conditions, the value is 2.0.

TABLE 6.4-13: ENVIRONMENTAL FACTOR - π_E

Environment	π_E
G_B (Ground Benign)	0.5
G_F (Ground Fixed)	2.0
G_M (Ground Mobile)	4.0

STEP 7: Select the quality value from MIL-HDBK-217. Since the product is a commercial device with an unknown screening level, the quality factor has a value of 10.0 as shown in Table 6.4-14. When the screening level is known, MIL-HDBK-217 has a table that relates π_Q values (lower than 10.0) to the specific screening level.

TABLE 6.4-14: QUALITY FACTORS - π_Q

Description	π_Q	Description	π_Q
Class S	0.25	Class B-1	2.00
Class B	1.00	Commercial	10.00

STEP 8: Using the equation for manufacturing learning from MIL-HDBK-217 which is:

$$\pi_L = 0.1 \exp (5.35 - .354/Y)$$

$$\pi_L = 1 \text{ for production lines in operation longer than 2 years}$$

where:

Y = years, which is 2

π_L = .01 exp (5.35 - .35(2))

= 1.05, which is rounded to 1.

STEP 9: Perform the calculation.

$$\lambda_p = [C_1\pi_T + C_2\pi_E]\,\pi_Q\pi_L$$

$$= [\,(0.08)(.29) + (.032)(2.0)]\,(10)\,(1.0)$$

$$= 0.87 \text{ failures per } 10^6 \text{ hours}$$

After one has calculated the failure rate for each component, the equipment failure rate is determined by summing the failure rates of the individual components as shown in equation 6.43.

$$\lambda_{EQUIP} = \sum_{i=1}^{n} \lambda_i \qquad (6.43)$$

and the MTBF is

$$MTBF = \frac{1}{\lambda_{EQUIP}} \qquad (6.44)$$

Stress analysis failure rate predictions such as this permit extremely detailed analyses of equipment or system reliability. However, since details of the system design are required in determining stress ratios, temperature and other application and environmental data, these techniques are only applicable during the later stages of design. Because of the high level of complexity of modern systems, the application of the procedure is time consuming.

6.4.5.3.3.3 Modification for Non-Exponential Failure Densities (General Case)

Although the exponential technique indicated in the previous sections can be used in most applications with little error, it must be modified (1) if the system contains parts for which the density function of failure times cannot be approximated by an exponential distribution over the time period of interest; or (2) if the parts which are the dominant factor in overall system unreliability do not follow an exponential density function of times to failure. Mechanical parts such as gears, motors, and bearings usually fall into this category.

In these cases, one cannot add the failure rates of all parts because there are some parts whose failure rates vary significantly with time. The method used is to consider separately within each block diagram the portion of the block containing parts with constant failure rates, and the portion containing parts with time varying failure rates. If the former portion contains n parts, then the reliability of this portion is

$$R_1(t) = \exp^{-\left(\sum_{i=1}^{n} \lambda_1\right)t} \tag{6.45}$$

The reliability of the second portion at time t is formed by using the appropriate failure density function for each part whose parameters have been determined through field experience or testing. If this portion contains B parts, then

$$R_2(t) = \prod_{i=1}^{B} R_i(t) \tag{6.46}$$

where:

$$R_i(t) = \int_t^{\infty} f_i(t)dt \tag{6.47}$$

and f(t) is the probability density function, general expression, of each of the B parts.

As discussed in 5.3.6, the Weibull distribution can be used to describe the distribution of times to failure in a wide variety of cases. If we use the Weibull to describe the portion of the block diagram containing parts with varying failure rate, equation 6.47 becomes:

$$R_2(t) = \prod_{i=1}^{B}\left(\int_t^{\infty} \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} e^{\left(\frac{t}{\theta}\right)^{\beta}}\right) = \prod_{i=1}^{B}\left(e^{\left(\frac{t}{\theta}\right)^{\beta}}\right) \tag{6.48}$$

where:

B	=	numbered parts
t	=	time
θ_i	=	Weibull scale parameter for part i
β_i	=	Weibull shape parameter for part i

The reliability for the block diagram, under the assumption of independence between the two portions, is

$$R(t) = R_1(t)\, R_2(t) \qquad (6.49)$$

For example, consider the failure rates of two elements, x and y, that make up a system. Let x be a microprocessor controller with a constant failure of 2 failures per million hours. Let y be a roller bearing operating at 1000 revolutions per minute for which 90% of the population will operate without failure for 3.6 x 10^9 revolutions. Bearing life test results have been fitted to the Weibull distribution with a shape parameter, β, of 1.5.

STEP 1: The microcircuit reliability is found by using equation 6.38.

$$R_1(t) = \exp(-\lambda t)$$

$$= \exp\left[\,-(2 \times 10^{-6})(50{,}000)\,\right]$$

$$R_1(t) = 0.905$$

STEP 2: The bearing reliability is determined by converting the revolutions into hours given that the speed is 60,000 revolutions per hour. This is 3.6 x 10^9 revolutions divided by 60,000 revolutions per hours which equals 60,000 hours.

Then scale parameter, θ, is determined from the standard Weibull equation shown as 6.48.

$$R(t) = \exp - \left(\frac{t}{\theta}\right)^{\beta}$$

where:

R(t)	=	0.9 at 60,000 hours (given)
t	=	60,000 hours
β	=	Weibull shape of 1.5 for product characteristic of early wearout
θ	=	mean-time-to-failure

$$R(t) \quad = \quad 0.9 = \exp\left(-\frac{60{,}000}{\theta}\right)^{1.5}$$

$$\theta \quad = \quad 60{,}000/(-\ln 0.9)^{1/1.5}$$

$$\theta \quad = \quad 268{,}967 \text{ hours}$$

This scale parameter is used to determine the reliability at the 50,000-hour point using equation 6.48.

$$R(t) \quad = \quad \exp -\left(\frac{t}{\theta}\right)^{B}$$

$$R(t) \quad = \quad 0.9 = \exp\left(-\frac{50{,}000}{268{,}976}\right)^{1.5}$$

$$= \quad 0.923$$

STEP 3: The system reliability is found using equation 6.49 where

$$R(t) \quad = \quad R_1(t)\, R_2(t)$$

$$= \quad (0.905)\,(0.923)$$

$$= \quad 0.835$$

STEP 4: Calculate the system MTBF as follows:

$$\text{MTBF} = \frac{\int_0^T R(t)dt}{1-R(T)} = \frac{\int_0^T \left(e^{-\lambda t}\right)\left(e^{-\left(\frac{t}{\theta}\right)}\right)dt}{1-\left[\left(e^{-\lambda t}\right)\left(e^{-\left(\frac{T}{\theta}\right)^{\beta}}\right)\right]}$$

where T is the time period of interest (T = 50,000 hours in this case).

$$\text{MTBF} = \frac{\int_0^{50,000} R(t)dt}{1 - R(50,000)} = 279,795 \text{ hours}$$

6.4.5.3.3.4 Nonoperating Failure Rates

The component failure rates in MIL-HDBK-217 (Ref. [11]) and in the Nonelectronic Parts Reliability Data (Ref. [12]) are based upon operating time. There are, however, equipment and systems in which nonoperating time represents a significant portion of the useful life, e.g., missiles, fuses, projectiles, etc.

Nonoperating component failure rate prediction models have been developed in the technical report, RADC-TR-85-91, *Impact of Nonoperating Periods on Equipment Reliability* (Ref. [15]). These models are patterned after those found in MIL-HDBK-217 and are applicable to equipment/systems subjected to nonoperating conditions.

Nonoperating failure rates are computed in a manner similar to operating failure rates only using somewhat different models and different multiplying factors. A typical nonoperating failure rate model is as shown in the following equation for discrete semiconductors.

$$\lambda_p = \lambda_{nb} \pi_{NT} \pi_{NQ} \pi_{NE} \pi_{cyc} \text{ failures}/10^6 \text{ nonoperating hours} \tag{6.50}$$

where:

λ_p = predicted transistor or diode nonoperating failure rate

λ_{nb} = nonoperating base failure rate

π_{NT} = nonoperating temperature factor, based on device style

π_{NE} = nonoperating environmental factor

π_{NQ} = nonoperating quality factor

π_{cyc} = equipment power on-off cycling factor

The nonoperating failure rate prediction models can be used separately to predict nonoperating failure rate and reliability, or they can be used to complement the operating failure rate prediction models in the other sections of the Handbook. The following equations illustrate the methods for predicting equipment (or system) level nonoperating reliability (R_n), service life failure rate (λ_{sl}) and combined operating/nonoperating reliability ($R_{o/n}$).

$$R_{n_i} = \exp(-\lambda_{ni} t_{ni}) \qquad R_n = \prod_{i=1}^{n} R_{n_i}$$

$$\lambda_{(sl)_i} = D_{o_i}\lambda_{o_i} + D_{n_i}\lambda_{n_i} \qquad \lambda_{sl} = \sum_{i=1}^{n} \lambda_{(sl)_i}$$

$$R_{(o/n)_i} = \exp(-(\lambda_{n_i} t_{n_i} + \lambda_{o_i} t_{o_i})) \qquad R_{(o/n)_i} = \prod_{i=1}^{n} R_{(o/n)_i}$$

where:

R_{n_i} = nonoperating reliability of the i^{th} item

λ_{ni} = nonoperating failure rate in the i^{th} nonoperating environment

t_{ni} = nonoperating time in the i^{th} nonoperating environment

$\lambda_{(sl)_i}$ = service life failure rate of the i^{th} item, equal to the number of failures per unit time regardless of operational mode

D_{o_i} = duty cycle in the i^{th} operating environment, equal to the time in the i^{th} operating environment divided by total operating time plus total nonoperating time

λ_{oi} = operating failure rate in the i^{th} operating environment

D_{ni} = duty cycle in the nonoperating environment, equal to the time in the i^{th} nonoperating environment divided by total operating time plus total nonoperating time

λ_{ni} = nonoperating failure rate in the i^{th} nonoperating environment

$R_{(o/n)_i}$ = reliability of the i^{th} item for the mission duration plus nonoperating time between missions

The following cautions are offered to prevent the misuse of the nonoperating failure rate models:

(1) Temperature in the models for discrete semiconductors and microelectronic devices is the ambient nonoperating temperature, not operating case or junction temperatures.

(2) Nonoperating environment is the actual environment to which the component is exposed. For example, an airborne radar between missions is most likely exposed to a ground fixed environment.

(3) Equipment power on-off cycling is determined at the equipment level. The parameter does not refer to actuations of switches or relays, nor specific circuit applications within the operating state.

6.4.5.3.4 Reliability Physics Analysis (Ref. [17] and [18])

Reliability physics is a technique for identifying and understanding the physical processes and mechanisms of failure. The concept has been around for decades and has resulted in great strides in component reliability design, even as component complexity has increased. The purpose of a reliability physic analysis is to identify components and processes that exhibit wearout failure before the expected end of use and to isolate the root cause of the failure.

The basic approach to this analysis, which is applicable to new or old components or processes, is outlined in Table 6.4-15.

An example of reliability physics approach is determine the average failure rate of a pinion during the first 1,500 hours of operation given a speed of 90,000 revolutions per hour. The L_{10} life of the pinion is 450 x 10^6 revolutions with a Weibull slope of 3.0. L_{10} life is the length of time that 90% of the pinions will meet or exceed during use before they fail. Table 6.4-16 illustrates the steps involved.

TABLE 6.4-15: BASIC APPROACH TO RELIABILITY PHYSICS ANALYSIS

Step	Discussion
1. Define the operating and nonoperating life requirements	Length of time or number of cycles expected or needed for both operating and nonoperating periods should be determined.
2. Define the life environment	Temperature, humidity, vibration and other parameters should be determined so that the load environment can be quantified and the cycle rates determined. For example, a business computer might expect a temperature cycle once each day from 60°F to 75°F ambient. This would quantify the maximum and minimum temperatures and a rate of one cycle per day.
3. Identify the material properties	

Step (13) Evaluate feasibility of allocated requirements

The procedures for making pre-design or interim reliability predictions are basically the same as for final design predictions except that the difference lies in the degree of precision (and details) with which the basic steps are implemented.

For predictions made at any stage of development, each of the steps will be carried out to the maximum extent possible. The system failure and operating and maintenance conditions should be defined as explicitly as possible. Reliability block diagrams are constructed to the lowest identifiable function, and appropriate system reliability formulas are established.

Precise parts lists, of course, cannot be compiled prior to design of an equipment. It is necessary, however, to make the best possible estimate of the parts complements of the various item subdivisions (blocks on the reliability diagram).

Stress analyses obviously cannot be made prior to design. However, for portions of the equipment that have not been designed, gross stress analyses can be accomplished. Stress levels

8. "Reliability Toolkit: Commercial Practices Edition," Reliability Analysis Center and Rome Laboratory, 1995.

9. Boyd, M.A., "What Markov Modeling Can Do For You," Annual Reliability and Maintainability Symposium - Tutorial Notes, 1996.

10. Regulinski, T.L., "Availability Function for Communicating Computer Net," Proceedings Reliability and Maintainability Symposium, 1980.

11. "Reliability Prediction of Electronic Equipment," MIL-HDBK-217F, 1995.

12. "Nonelectronic Parts Reliability Data," (NPRD), Reliability Analysis Center, 1995.

13. Bellcore, TR-332, "Reliability Prediction Procedure," Issue 5, December 1995.

14. "Electronic Parts Reliability Data," (EPRD), Reliability Analysis Center, 1997.

15. "Impact of Nonoperating Periods on Equipment Reliability," RADC-TR-85-91, 1985.

16. "Nonoperating Reliability Data Book," (NONOP-1), Reliability Analysis Center, 1987.

17. "Reliability Assessment Using Finite Element Techniques," RADC-TR-89-281, Rome Laboratory, 1989.

18. "Computer-Aided Assessment of Reliability Using Finite Element Methods," RADC-TR-91-155, Rome Laboratory, 1991.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

7.0 RELIABILITY ENGINEERING DESIGN GUIDELINES

7.1 Introduction

Reliability engineering is the technical discipline of estimating, controlling, and managing the probability of failure in devices, equipment and systems. In a sense, it is engineering in its most practical form, since it consists of two fundamental aspects:

(1) Paying attention to detail

(2) Handling uncertainties

However, merely to specify, allocate, and predict reliability is not enough. One has to do something about it in terms of having available a family of design guidelines which the designer can use to achieve a desired reliability. These guidelines are provided in this section.

During a product development program, a design is developed to meet previously defined quantitative reliability requirements. The importance of designing in the required degree of reliability initially cannot be overemphasized, for once the design is approved, inherent reliability is fixed.

There are a host of design principles and tools of which the designer should be aware and should use as required to achieve a reliable electronic equipment/system design. They include:

(1) Parts Management
(2) Part derating
(3) Reliable circuit design
(4) Redundancy
(5) Environmental design
(6) Human factors design
(7) Failure modes and effects analysis (FMEA)
(8) Fault tree analysis (FTA)
(9) Sneak circuit analysis
(10) Design reviews
(11) Design for testability
(12) System safety program
(13) Finite element analysis

Each of these will be briefly discussed in this section in terms of its role in the design of reliable equipment/systems.

7.2 Parts Management

Many factors affect the ultimate levels of quality and reliability of parts. Possibly the most important factor is the degree to which the manufacturer is able to fabricate them in a defect-free manner. This factor is a strong function of the volume and continuity of part production. Additional factors affecting part reliability are the levels to which the part is screened, the application, and the manner in which the part is integrated into the system.

The volume of parts produced usually impacts field reliability, since manufacturers producing large numbers of parts on a continuous basis can easily benefit from Statistical Process Control (SPC). When used wisely, SPC has proven to be an effective tool for improving processes, thereby increasing the quality and reliability levels of manufactured parts. Manufacturing lines intermittently producing small numbers of parts on a line with non-standard manufacturing processes typically do not exhibit the reliability levels of fully loaded manufacturing lines using well-controlled manufacturing processes.

Critical parts are often highly reliable, simply due to the attention given to them by both the part manufacturers and by the users. As an example, consider integrated circuits. When first used extensively twenty years ago, they often were the predominant device type limiting system reliability. Since then, due to their critical nature, part manufacturers have improved their reliability by orders of magnitude and part users are learning how to apply them in a manner which results in a robust design. These efforts have resulted in integrated circuits that are much more reliable than many other part types used in systems.

Therefore, high usage, highly critical and high volume parts often show rapid technology maturation, whereas low usage, noncritical or low volume parts can exhibit slower reliability improvement and result in lower levels of field reliability. As an example, consider the items identified by field data as being high failure rate parts: fasteners, actuators, connectors, transducers and switches. These are ordinary and necessary parts which are not considered state-of-the-art, but yet can significantly impact field reliability.

The general elements of an effective Parts Management Plan (PMP) are (MIL-HDBK-965, "Acquisition Practices for Parts Management" provides guidance in selecting tasks to include in a PMP):

(1) Preferred Parts List
(2) Vendor and Device Selection
(3) Critical Devices/Technologies/Vendors
(4) Device Specifications
(5) Screening
(6) Part Obsolescence
(7) Failure Reporting, Analysis and Corrective Action (FRACAS)

Each of these elements can be tailored to meet the specific needs of each system. Reference [1] "Parts Selection, Application and Control" provides generic guidance in the development of this process.

Each of these elements are discussed in the following subsections.

A comprehensive PMP defines the manner in which each of the aforementioned elements will be addressed. It should identify the responsible personnel and include a milestone schedule. This plan can also be tailored in accordance with specific requirements of the system for each of the PMP elements. Tailoring should be accomplished considering:

(1) Development Cycle Time
(2) Warranty Period
(3) Maintainability
(4) Cost of Failure
(5) System Characteristics
  - (a) volume
  - (b) weight
  - (c) performance
  - (d) operating environment

(6) Budget
(7) Screenability
(8) Preventive Maintenance
(9) Customer Requirements
(10) Severity (or Criticality) of Failure

Understanding, defining and then implementing all the tasks involved in a PMP program is the key to its success. The representation and active participation of the following disciplines, as a minimum, are necessary to enable, in a concurrent engineering fashion, an effective PMP:

(1) Parts (components) engineering
(2) Reliability engineering
(3) Design engineering
(4) Manufacturing engineering

Successful implementation of a PMP requires a disciplined approach, and must have management participation and support to ensure cooperation among disciplines and resolve any differences based on the ultimate impacts on cost, schedule and performance.

7.2.1 Establishing a Preferred Parts List (PPL)

In the course of a design effort, equipment designers need to select the parts and materials to be used to meet specified equipment requirements for performance, reliability, quality, producibility and cost. This selection task is greatly enhanced if the designer has a list of preferred parts available to help in this selection process.

Preferred parts are those whose quality and reliability are well-known to the industry, and are probably parts that the company is already using in other equipments. Without a preferred parts list (PPL), designers may tend to choose parts in haphazardly. The result is the uncontrolled proliferation of parts throughout a manufacturer's product line, all varying in performance and reliability. All potential candidate parts should undergo an independent assessment before being placed on the preferred parts list. Information sheets or data sheets from part suppliers may paint an optimistic picture of the part's capabilities, but may omit information regarding the part's inherent characteristics that are critical to proper operation of the final product.

The absence of a PPL may have wide-ranging consequences for manufacturing, purchasing, and logistics. Manufacturing engineers may have to cope with parts that require a variety of assembly methods and unique tooling. More inventory may be needed and, as a result, inventory costs can mushroom out of control. Manufacturing automation may also be adversely affected. Purchasing representatives may have to deal with many different suppliers, making it hard for them to monitor quality and timely delivery, and to obtain volume cost discounts. Logistics specialists must now provide spares for many different parts, enter them into the supply system, and find storage space for all of them.

Some consequences of designing equipment without a PPL are:

(1) Proliferation of non-preferred parts and materials with identical functions
(2) Increased need for development and preparation of engineering justification for new parts and materials
(3) Increased need for monitoring suppliers and inspecting/screening parts and materials
(4) Selection of obsolete (or potentially obsolete) and sole-sourced parts and materials
(5) Possibility of diminishing sources
(6) Use of unproven or exotic technology ("beyond" state-of-the-art)
(7) Incompatibility with the manufacturing process
(8) Inventory volume expansion and cost increases
(9) Increasing supplier base and audit requirements
(10) Loss of "ship-to-stock" or "just-in-time" purchase opportunities
(11) Limited ability to benefit from volume buys
(12) Increased cost and schedule delays
(13) Nonavailability of reliability data
(14) Additional tooling and assembly methods may be required to account for the added variation in part characteristics
(15) Decreased part reliability due to the uncertainty and lack of experience with new parts
(16) Impeded automation efforts due to the added variability of part types
(17) Difficulty in monitoring vendor quality due to the added number of suppliers
(18) More difficult and expensive logistics support due to the increased number of part types that must be spared.

When a PPL is available at the beginning of the design process, designers avoid using non-approved parts and the laborious task of having to supply engineering justification for their use.

Preferred parts databases help to flag obsolete parts and also indicate a part's long term availability (i.e., how long a given part may actually be available in the market).

The PMP must provide some level of standardization to minimize the number of new parts entering the system/equipment, or the logistic support and supply system, and yet still be flexible enough to effectively capitalize on the advantages offered by alternative technologies. To be truly effective, the PMP must first ensure that the parts selected will provide the necessary level of performance and reliability over the projected life of the system/equipment. It must also be tailored to the expected life of the equipment to ensure, among other things, that replacement spares will continue to be available throughout the effective life of the system/equipment. The PPL should be updated periodically to ensure a proactive approach to minimizing the impact of part obsolescence.

7.2.2 Vendor and Device Selection

Major factors to consider when implementing a PMP is the evaluation of vendors and the selection of components. It is imperative that engineers select and use components from manufacturers in which they have confidence. This confidence can be attained either empirically through adequate past performance of the part manufacturer, or from verification that the manufacturer is indeed producing high quality parts. The latter can be achieved via evaluation of the part manufacturing processes through testing and subsequent data analysis.

To ensure the supply of adequate parts, both vendors and subcontractors must be effectively managed. A procedure is needed in which each vendor/technology is evaluated, certified and qualified in a cost-effective manner. Traditionally, this procedure was to test all devices and audit all vendors. Due to the increased emphasis on quality (especially in microcircuits), a more generic approach to vendor certification/qualification of processes is recommended. Then, existing data from technology families can be used for qualification by similarity. Ongoing vendor/customer testing programs on representative products may be used to determine acceptability. Procedures for performing and monitoring vendor/product testing and incoming inspection are still necessary, but should be tailored to allow each vendor to be handled on a case-by-case basis. For example, outgoing vendor quality and user incoming inspection and board level testing can be monitored to determine device quality and product design/manufacturing process compatibility. Data analysis can then determine the need for vendor testing, incoming inspection and increased vendor surveillance. These data can also form the basis for determining whether a "ship to stock" program (i.e., acceptance of a product without incoming inspection) is feasible.

Parts must be selected based on a knowledge of both the application environment in which the part is to operate and the conditions it is exposed to during part manufacturing, assembly, handling and shipping. It is equally important to understand how the failure rate during the part's useful life, and its wearout characteristics (lifetime), are impacted by the specific application conditions. Only with this understanding are robust designs possible.

One specific area of importance is the continuity of production. As mentioned earlier, facilities/production lines that manufacture parts on a continuous basis often produce higher quality parts than those manufactured on an intermittent basis. Intermittent production can be a characteristic of custom, low usage parts. High volume, continuous production is usually controlled in a statistical manner, whereas intermittent production may not be able to implement SPC. Additionally, intermittent lines often run into unanticipated problems associated with start-up which can adversely affect the quality, availability, and reliability of the part.

Many successful organizations have developed a qualified manufacturers list (QML) on which procurement decisions are based. A QML lists manufacturers who have proven that they can supply good parts with a high degree of confidence. The DoD is also using this methodology in the procurement of microcircuit devices, via the QML program (i.e. MIL-PRF-38535).

Part manufacturers can be evaluated in many ways. For suppliers of parts that have been manufactured for some time, analysis of historical reliability/quality data is usually the optimum method. In many cases, these data are readily available from the manufacturer and, in some cases, are published in their data catalogs. To be meaningful, historical data must be representative of the same, or a similar, part with few changes, and must be for a similar application under similar operational stresses.

Vendor evaluation can be accomplished by analyzing design, manufacturing, quality, and reliability practices. Figure 7.2-1 illustrates a methodology to evaluate potential vendors.

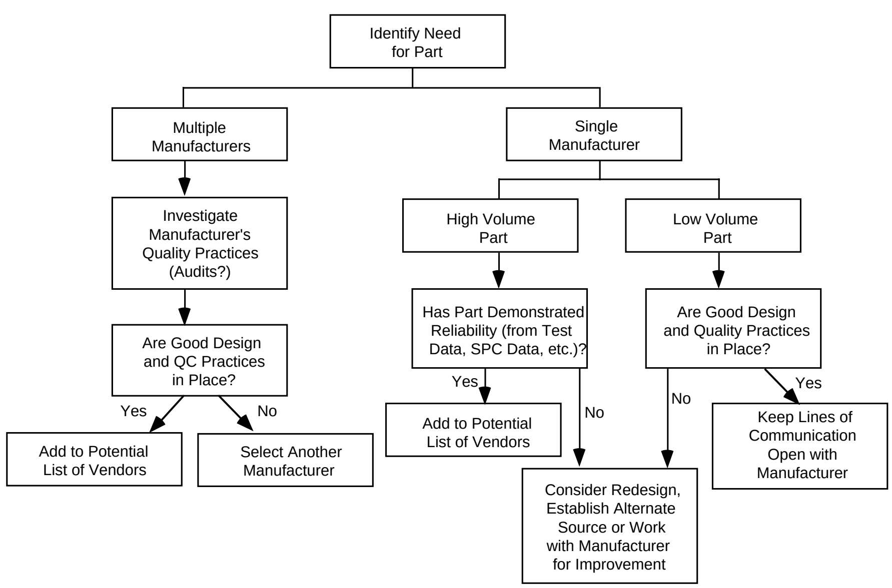


FIGURE 7.2-1: VENDOR SELECTION METHODOLOGIES

An audit/validation should focus on whether a documented baseline system exists and is being used. Additionally, required demonstration of generic product manufacturability, verified by reliability testing, is necessary. Representative questions such as those in Table 7.2-1 should be asked.

TABLE 7.2-1: QUESTIONS FOR PART SUPPLIERS

- Is a quality program defined and implemented?
- Have potential failure mechanisms been identified?
- Are the manufacturing materials and processes documented?
- Are there process controls in place?
- Are parts manufactured continuously or is there intermittent production?
- What defect levels are present?
- Is there a goal in place for continuous improvement?
- Have life limiting failure mechanisms been designed out?
- If it is not practical to design or screen out life limiting mechanisms, have they been modeled such that the user can quantify the part's lifetime in a specific application?
- Are efforts being taken to identify the causes of part failure and to improve the manufacturing process to alleviate their occurrence?
- Is the part screening and qualification process effective?
- Are design rules used and adhered to that result in high quality and reliability?
- Are design changes made only after analyzing and quantifying possible reliability impact?
- What is the on-time delivery success rate?

Recent improvements in customer/supplier relationships have resulted in alliances or partnerships where both parties work together to improve the quality and reliability of delivered products. However, to achieve these alliances, it is necessary to understand that:

(1) Effective preferred parts selection is a dynamic process which minimizes the number of different parts used, while providing designers an adequate choice of components to meet equipment performance requirements.
(2) Vendor selection certification and qualification criteria based on technical expertise are used to minimize the number of vendors.
(3) Good production suppliers are willing to support problem analysis and corrective actions.

A process based on these considerations should be formalized to assess and validate potential suppliers' quality, reliability and manufacturing systems, and delivery, cost and service performance. The resulting data, when reviewed by cognizant personnel (i.e., purchasing, design engineering and manufacturing), can be used to select the appropriate suppliers. Once this step is accomplished, alliances or partnerships can be established that can result in a win-win situation, where the procurement process changes from an adversarial to a cooperative relationship. To establish the continuous improvement process and manage the supplier, a reassessment and information exchange program can be put in place. The validation/audit plan results that were used to select a vendor can now be the reference from which progress is measured.

7.2.2.1 Critical Devices/Technology/Vendors

Critical part types are considered to be those that require additional attention due to potential reliability, safety or availability problems. Many parts programs focus too much attention on standard or non-controversial part types. It is imperative that special attention be given to critical parts, since they are often the parts driving the system reliability. The establishment of a listing of critical devices, technology and vendors, and a monitoring/action plan, should be part of every PMP, and should address components exhibiting the following characteristics:

(1) Performance Limitations: due to stringent environmental conditions or non-robust design practice.
(2) Reliability Limitations: component/materials with life limitations or use of unrealistic derating requirements.
(3) Vendors: those with a past history of delivery, cost performance or reliability problems
(4) Old Technology: those with availability problems
(5) New Technology: parts fabricated using immature design and manufacturing technology

The first three categories require historical data to track and define actions to minimize their occurrence or provide alternate solutions. In addition, sound component engineering judgment and the combined efforts of design, reliability, and manufacturing engineering, and vendors, are needed to ensure the identification and management of critical components.

The subject of old and new technology can involve the generation of different procurement procedures for tracking technology maturity, obsolescence and hidden hybrids (i.e., those devices that fall between generic device categories and, as a result, are incorrectly specified and tested, see 7.2.2.3).

The PMP should address the identification and control of limited or critical parts and off-the-shelf equipment. Also, the PMP must ensure that parts engineering, design, manufacturing and reliability personnel are notified of potential risks in using critical parts/technology. As stated previously, a PMP program must be tailored to account for the unique failure mechanisms associated with the parts being used. For example, if plastic packaged microcircuits are used, their expected lifetime must be determined as a function of the use environment (temperature, humidity, dissipated power, etc.) and then compared to the design life of the equipment in which the component operates. As another example, consider off-the-shelf equipment. In this case, the equipment must be analyzed to assess its suitability for use in its intended application. This is especially true for commercial equipment designed for benign environments that are to be used in more severe environments. A determination should be made of its reliability and performance, since it would be neither cost-effective nor practical for the vendor to change the design or production procedures. The task becomes one of evaluating subcontractor procedures, reliability design analyses and past performance.

7.2.2.1.1 ASIC Devices

The rapid technology changes in the field of microelectronics, both hybrid and monolithic, have to be monitored closely. Application Specific Integrated Circuits (ASICs) are one part type usually considered to be critical. The advent of ASICs requires a change in the device selection procedure. The term "ASIC" describes a wide variety of different types of devices which can include custom and semicustom standard cells, gate arrays, Programmable Logic Devices (PLD) and Field Programmable Gate Arrays (FPGA) typically designed for a specific application. Advantages include a relatively short development cycle and customized performance and functionality. Disadvantages are that the equipment schedule may be impacted because system designers are involved in the device design cycle, and unproven vendors and technologies may be used for the first time.

Typically, ASIC devices are designed for a very specific application and then produced and sold in very limited quantities. Thus, there is no market for marginal devices. Historically, generic ICs have been produced in a tiered market environment. Parts not meeting the highest level of performance could usually be sold to a less demanding customer at a reduced price. This is simply not the case with ASICs. Either they are 100% perfect or they are scrap. Given this fact, there is a very strong incentive to reduce or eliminate all possible variation in the part manufacturing processes to attain a very high yield of good parts. Thus, Total Quality Management (TQM) and SPC become imperative to the manufacture of these types of parts. To use ASICs, a supplier must select and certify a silicon foundry and design the device using foundry design rules. Performance would be demonstrated through simulation tools. The foundry would then fabricate wafers and packaged devices for test. The planning and management of ASIC design requires a very rigorous and controlled procedure to achieve desired device functionality, reliability, cost and delivery schedules.

7.2.2.1.2 GaAs and MMIC Devices

Gallium arsenide (GaAs) devices are now being used in military and commercial systems. GaAs offers some significant advantages over silicon that can result in improved device performance. It has unique qualities which allow the fabrication of devices that can operate at frequencies which outperform their silicon counterparts. In addition, GaAs offers inherent radiation hardness and improved power efficiency for high frequency digital and analog circuitry.

Monolithic Microwave Integrated Circuits (MMIC) are replacing hybrid microwave devices throughout the industry as a result of the Defense Advanced Research Project Agency (DARPA) sponsored Monolithic Microwave Millimeter Wave Integrated Circuit (MIMIC) program. Before the development of GaAs MMIC technology, discrete packaged devices and multifunction assemblies were commonly utilized in microwave applications. MMIC technology, however, offers several advantages including weight/size reduction, process tolerance and uniform performance with a reduced need of tuning circuits. These advantages, combined with GaAs's inherent performance advantages, have led to significant interest in the technology.

To date, information concerning the reliability of GaAs and MMIC components has shown varying results and inconsistent activation energies for a specific failure mechanism. Thus, the absolute reliability of GaAs devices is not easy to predict with accuracy, though an approximation can be made based on government/industry reliability studies.

7.2.2.2 Plastic Encapsulated Microcircuits (PEMs)

Plastic packaging is a leading factor in the growth of microelectronics companies and has had a significant positive effect in the telecommunications, computer and automotive industries. PEMs have demonstrated cost effectiveness while providing improved performance and reliability in these applications environments. Now, acquisition reform initiatives and continued improvements in plastic packaging and die protection (i.e., silicon nitride passivation) have led to their consideration and limited use in military environments. The RAC publication PEM2 (Ref. 2) provides additional information.

7.2.2.3 Hidden Hybrids

Quality and reliability efforts for microcircuits have been more intense than for any other commodity items, resulting in orders of magnitude improvement in their performance since the 1960's. However, the procurement of complementary devices/modules sometimes ignores the lessons learned in this area. We have chosen to call these devices "hidden hybrids," indicating a mix or composite of technologies.

Examples include the following:

(1) Crystal Oscillators
(2) Multichip and Microwave Modules
(3) Power Regulators (Supplies)
(4) Solid State Relays
(5) Transformers

In many cases, these items have escaped the traditional testing and technology/vendor evaluation that has led to the successes achieved for microelectronics devices. Crystal oscillators evolved from a combination of discrete components mounted on a printed wiring board to hybrid microcircuits made up of chip components (including the crystal), all contained in a single hermetic package. Solid state relays are essentially a hybrid device containing discrete semiconductor components which should be individually tested and controlled.

The problem is presently being compounded by the various types of multichip and high frequency (i.e., R.F. and microwave) modules being introduced. Multichip modules (MCM) are taking advantage of integrated circuit, hybrid and printed wiring board (PWB) technologies, and are being used to fabricate state-of-the-art high performance products.

It is specifically recommended that packaged items be reviewed to uncover potential "hidden hybrids" as shown in Table 7.2-2. Once located, the appropriate component procurement approach (such as MIL-PRF-38534) should be used to ensure reliable and quality products. Incorporation of appropriate evaluation, audit and testing requirements could eliminate costly testing and corrective action procedures at a later date, while ensuring customer satisfaction.

TABLE 7.2-2: HIDDEN HYBRID CHECKLIST

Analyze:
- Fabrication Process - uses hybrid microcircuit assembly techniques
- Technology - contains microcircuits and/or semiconductors
- Packaging - potted/encapsulated modules

Take Action:
- Testing Requirements - per applicable test procedure

7.2.2.4 Device Specifications

Part electrical, mechanical and physical characteristics should be defined in a device specification to be used for design and procurement. Applicable device electrical performance parameters that ensure product performance objectives are met for all operating conditions should be specified, including reliability parameters. This information may be available in vendor catalogs/data sheets. Special care should be taken for electrical parameters that are "guaranteed but not tested," or other special features which should be discussed and agreed to with each vendor. The part specification should be based on several factors, including operating environments, worst case stress levels, and quality requirements.

The Defense Electronic Supply Center (DESC) Standard Microcircuit Drawing format is an example of how to prepare a company specification for microcircuits and other applicable components. This format has been reviewed and coordinated with industry and can be used to develop a specification that provides realistic, clearly stated requirements. Details are provided in MIL-HDBK-780 "Standardized Microcircuit Drawings." Reference [3] “Analog Testing Handbook (ATH)” provides information for the specification for analog and mixed mode (analog/digital) devices.

7.2.2.5 Screening

Screening, or 100% testing similar to that included in MIL-PRF-38534 and -38535, is recommended, pending vendor validation and use experience. Data from appropriate in-process and reliability testing can be used to justify deletion of end-of-line tests. Vendor/customer trust and alliances can result in practical cost-effective testing.

7.2.2.6 Part Obsolescence and Diminishing Manufacturer Sources (DMS)

Obsolescence occurs when parts that are required for system support are no longer manufactured (usually due to insufficient market demand). It is a common occurrence within the DoD for systems to have lifetimes greater than the life cycle of their supporting part technologies. Hence, part obsolescence is typically more of a problem for military systems than for commercial systems. Also, parts qualified for military use have historically represented more mature technologies relative to those used in non-military applications. The potential for diminishing manufacturing sources, causing parts that are not yet obsolete to become unavailable, must also be considered. This unavailability can be the result of the manufacturer experiencing limited orders, downsizing, market instability, or the result of other business decisions to exit the market for a particular technology or device. Regardless of the reason, the part is unavailable, and the effect is essentially the same as if the part had become obsolete.

Part and vendor obsolescence management should be a basic part of a company's operating, design, and manufacturing procedures (i.e., best commercial practices) and be substantially product independent, evolve around needed components, operating environments and package styles. Implementation of an effective PMP requires diligent management in maintaining part availability for system support, including taking the actions necessary to maintain availability of parts that are, or will be, obsolete during the equipment life cycle. Such actions can be grouped into two categories: management and technical.

Management solutions to part availability problems include preventive measures to alleviate the use of potentially obsolete parts, detection of the use of potentially unavailable parts, and identification of the need to procure an adequate quantity of parts to support the equipment throughout its life cycle. Management solutions include the use of a PPL and the lifetime purchase of parts to ensure part availability in the event that they become obsolete. This latter solution carries its own risks and burdens (for example, provisions for storing the parts in a sufficiently benign environment that precludes the occurrence of storage-related failure mechanisms).

Technical solutions include replacement of the unavailable part with an equivalent part, device emulation, and system redesign. If there is a direct replacement available, substitution is usually the easiest and least costly solution. There are several semiconductor information sources that can assist in the identification of equivalent parts. These include the IC Master and Part Master (available from International Handling Services), and Computer Aided Product Selection (CAPS) (available from Cahners Publishing).

Early notification of part/vendor end-of-life status provides time to select an acceptable solution that will minimize the identified problem's impact on manufacturing. External sources such as the Defense Logistics Agency/Defense Supply Center Columbus (DLA/DSCC), Government Industry Data Exchange Program (GIDEP) and vendors, as well as management of the company's internal PPL, can be used to provide early notification. Figure 7.2-2 illustrates a process flow for short and long term solutions that takes place when obsolete part notification is received. The major difference between short and long term solutions is that, in the long term solution, even when a part or vendor exists or another solution is found, the effort does not stop. As mentioned, it is critical that the solution is not just a stop gap and that long term support issues are addressed. Therefore, a trade study using the factors indicated in Figure 7.2-2 is performed to ensure a long term solution is not required in the future. (This concept is further described in reference [4] "767 AWACS Strategies For Reducing DMS Risk").

When a device has been identified as needed but unprocurable, the most practical solution is emulation. Device emulation is a process by which a direct replacement is designed and fabricated for an unavailable part. The design task may include reverse engineering, simulation, or direct design and fabrication (if original schematics and drawings are available). The Defense Logistics Agency (DLA) currently leads such an emulation program, referred to as the Generalized Emulation of Microcircuits (GEM).

System redesign is also a possible technical solution to alleviate the dependence on unavailable parts. Device emulation and system redesign can be very costly solutions to the unavailability problem. Implementation of preventive measures early in the part selection process can provide significant cost savings as the system reaches end-of-life.

The VHSIC Hardware Description Language (VHDL) is a valuable tool that can assist in the emulation or redesign of devices. VHDL is fast becoming the hardware description language of choice because it is an IEEE standard and has technology process and vendor independence, CAD tool support, and top-down design methodology capability. What is required is a VHDL behavioral description of the obsolete device or printed wiring assembly. The next step is to produce a structural VHDL description of the design to be emulated, which can then be processed by logic and layout synthesis tools of choice. This emulated design can then be processed by a compatible wafer foundry processing capability and packaged appropriately for insertion.

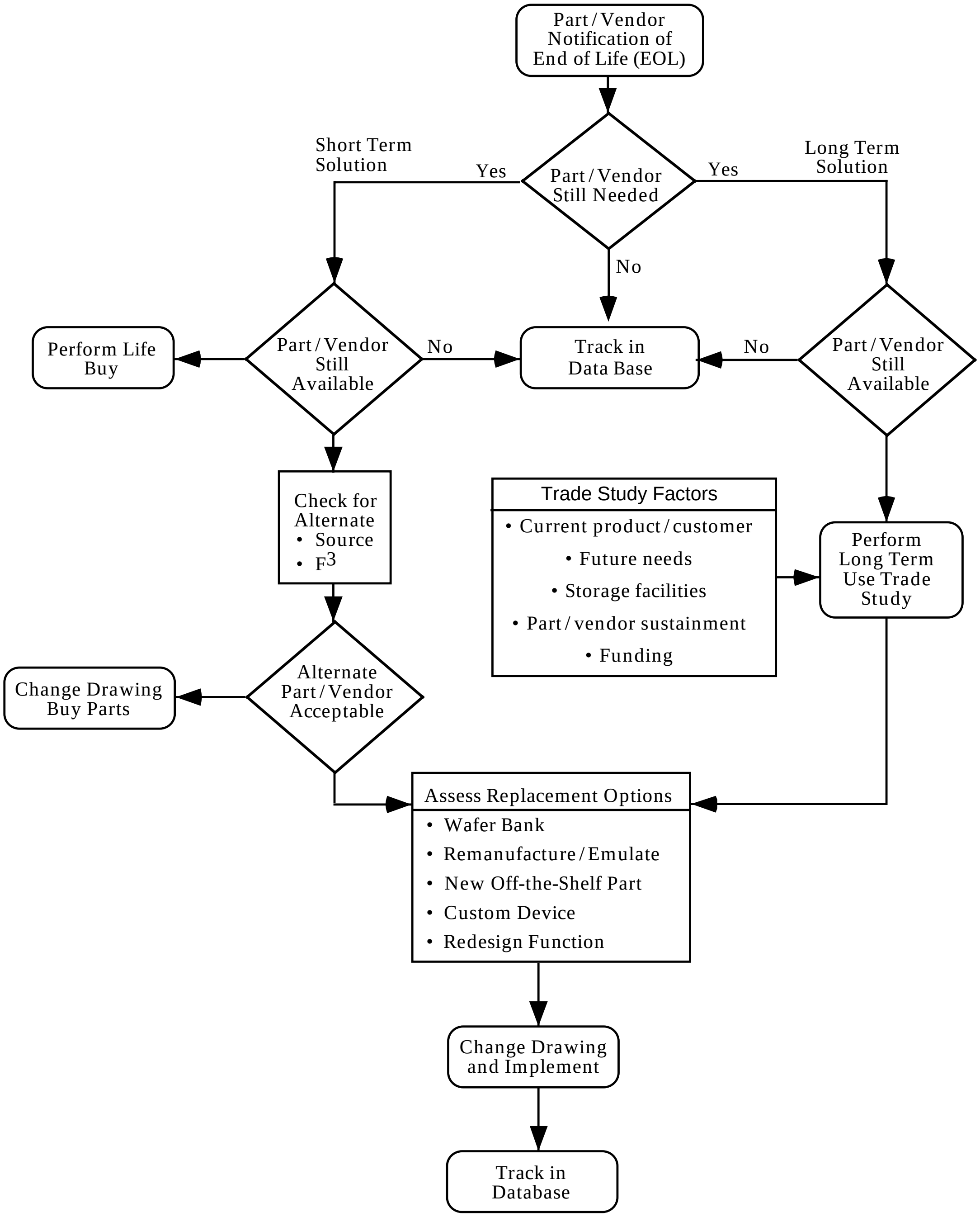


FIGURE 7.2-2: PART OBSOLESCENCE AND DMS PROCESS FLOW

7.2.2.7 Failure Reporting, Analysis, And Corrective Action System (FRACAS)

FRACAS is a management tool established to identify and correct deficiencies in equipment and thus prevent further occurrence of these deficiencies. It is based upon the systematic reporting and analysis of failures during manufacturing, inspection, test and use. The closed-loop feature of FRACAS requires that the information obtained during the failure analysis be disseminated to all of the decision making engineers and managers in the program. See Section 8 for more information on FRACAS.

7.2.3 Design for Reliability

In Section 7.2, the elements of a traditional Parts Management Program were discussed. This section discusses some of the methodologies that can be used to ensure that systems are designed and parts are applied in a robust manner. It presents an overview of the analytical tools that can be used to ensure a robust design and discusses several considerations for ensuring a manufacturable product. Although this material is not part of a traditional parts management program, it is relevant since the manner in which a part is used is as important as ensuring an adequate part is obtained. This observation illustrates the inseparability of part selection and application in the design and manufacture of a reliable system, and illustrates the necessity of using a concurrent engineering approach.

In the course of developing a system or equipment, suppliers must determine the market the product will serve and the need they will fulfill (i.e., environment to be used in, quality/reliability to satisfy customer, guarantee/warranty, and performance when compared to competition and cost). Once this is determined, requirements for part quality levels, design guidelines, temperature range and packaging can be defined. Assembly procedures must be defined to determine the appropriate component packaging (i.e., surface mount, through-hole, etc.). Design guidelines for manufacturing producibility must be available to determine package lead pitch vs. printed wiring board capability, specification of component drift parameters and the many other factors leading to robust design. Once they are determined, the PMP function can attempt to provide the components and materials necessary to satisfy them. The output of this function should be company-specific procedures containing:

(1) Guidelines for choosing component quality levels
(2) Design guidelines
  (a) Performance
  (b) Environmental/temperature
  (c) Assembly procedures
(3) Manufacturing/assembly procedures
(4) Performance/reliability demonstration plan

Correct application of parts means "using the best part for the job in an optimum/cost effective manner." Hence, electrical and electronic parts must be selected with the proper temperature, performance, reliability, testability, and environmental characteristics to operate correctly and reliably when used in a specific application. Parts and materials should be selected based on their ability to meet functional requirements for a given period of time under the expected extremes of operating stresses, including shock/vibration, temperature, temperature cycling, humidity, contamination, mechanical stress, electrical stress, radiation and electromagnetic interference. Factors to be considered in optimum parts application are both numerous and complex, and should address each of the factors included in Table 7.2-3. Many of these part application factors can be specifically addressed by performing a reliability assessment.

Design for reliability is the process of selecting a part or material and applying it in such a manner that results in high reliability under the worst case actual use conditions. Such an effort requires a structured approach during the part selection and design process. This process should include:

(1) Definition of operating environments
(2) Establishment of lifetime requirements
(3) Use of reliability models to estimate lifetime under use conditions
(4) Estimates of reliability during the useful life
(5) Stress derating
(6) Analysis and design modifications to ensure robustness

Several analytical techniques are useful in robust design. These include derating, failure mode and effects analysis (FMEA) (with or without criticality analysis), fault tree analysis (FTA,) and finite element analysis (FEA) (see 7.3, 7.8, 7.9, and 7.14, respectively).

7.2.3.1 Electronic Part Reliability Assessment / Life Analysis

A reliable product requires that the applicable part reliability and life requirements be adequately defined. This effort requires accurate quantification of the environmental and operational stresses that the part will experience during use and an assessment of part reliability and life under these conditions. Typical stress profiles are frequently used, but worst case stress values may often be better suited for this assessment.

TABLE 7.2-3: GENERIC PART APPLICATION FACTORS

Operating Temperature Range - parts should be selected which are rated for the operating temperature range to which they will be subjected.

Electrical Characteristics - parts should be selected to meet EMI, frequency, waveform and signal requirements and maximum applied electrical stresses (singularly and in combination).

Stability - parts should be selected to meet parameter stability requirements based on changes in temperature, humidity, frequency, age, etc.

Tolerances - parts should be selected that will meet tolerance requirements, including tolerance drift, over the intended life.

Reliability - parts should be selected with adequate inherent reliability and properly derated to achieve the required equipment reliability. Dominant failure modes should be understood when a part is used in a specific application.

Manufacturability - parts should be selected that are compatible with assembly manufacturing process conditions.

Life - parts should be selected that have "useful life" characteristics (both operating and storage) equal to or greater than that intended for the life of the equipment in which they are used.

Maintainability - parts should be selected that consider mounting provisions, ease of removal and replacement, and the tools and skill levels required for their removal/ replacement/repair.

Environment - parts should be selected that can operate successfully in the environment in which they will be used (i.e., temperature, humidity, sand and dust, salt atmosphere, vibration, shock, acceleration, altitude, fungus, radiation, contamination, corrosive materials, magnetic fields, etc.).

Cost - parts should be selected which are cost effective, yet meet the required performance, reliability, and environmental constraints, and life cycle requirements.

Availability - parts should be selected which are readily available, from more than one source, to meet fabrication schedules, and to ensure their future availability to support repairs in the event of failure.

Part test data is generally used to assess part reliability under specific operating stresses. Such data can take many different forms. Useful reliability assessment data are often gleaned from the analysis of life tests, performed either by the part manufacturer or by the user of the part. Helpful part reliability and life information may also be found in the literature. In any case, the data used for this assessment must address the specific predominant failure mechanisms applicable to that particular part and the specific materials used in the construction of that part. The use of appropriate data can help in ensuring adequate part life in a specific application, as well as in projecting anticipated part reliability. On the other hand, using inappropriate part life and reliability assessment data can give a false degree of confidence in the life estimate and thus provide a potential for early field failures or poor long term reliability.

Part failure mechanisms can generally be grouped into two categories: common cause and special cause. These two types of mechanisms have very different failure characteristics. This difference must be recognized, and properly addressed, in the data collection, analysis and assessment effort.

Common cause failures are due to inherent failure mechanisms; they have the potential of affecting the entire population of parts. These mechanisms are typically addressed through the design of the part itself and the part's fabrication process controls. These contributions help to ensure that the device is sufficiently robust to operate reliably for a given period of time. For these types of mechanisms, a physics-of-failure based reliability assessment is appropriate, since it is possible to gain a good understanding of the failure mechanisms. Such an assessment requires a fundamental knowledge of the device fabrication process, the appropriate process controls, and applicable materials data.

Special cause failure mechanisms result from defects or from specific events. An example of such a mechanism might be: capacitor failures resulting from a defective dielectric or from electrical overstress. Since special cause failure mechanisms are defect or event related, rather than process related, they tend to occur randomly. For such mechanisms, a purely physics-based assessment may not be appropriate, due to the random nature of failure occurrence. For these failure mechanisms, statistical analysis of the data is usually the more appropriate assessment approach.

Clearly, it is important that a combination of both a physics-based approach and a statistical analysis approach be used in any part reliability and life assessment. Because of the differences in the potential failure mechanisms involved, either approach used alone is unlikely to yield correct conclusions regarding part reliability or life assessment.

7.2.4 Design for Manufacturability

Of equal importance to the selection and application of parts is the manufacturing process to be used for their implementation. The best part used in a good design is useless, unless it can be processed in a reliable and reproducible manner. Manufacturing process evaluation is especially important for new or immature technologies, or for technologies for which the manufacturer has little or no experience. Therefore, the manufacturability of equipment designs should be given equal weight with the part selection and application efforts. Reference [5] “Best Practices - How to Avoid Surprises in the World’s Most Complicated Technical Process” is a good reference source for this task.

Procedures are required today to not only procure acceptable parts and materials, but also to ensure that the process steps from shipping to assembly do not destroy good components. It is not enough to qualify components to a standard qualification procedure, because some current assembly processes impose greater stress than those used in the past. A classic example is surface mount technology, which uses soldering processes (i.e., vapor phase, infrared heating) that provide a very fast temperature transition to 220°C, creating a thermal shock which is greater than that used for component verification testing. This is exemplified by the use of plastic surface mount packages which, in some cases, have resulted in the "popcorn effect." This refers to a phenomena in which moisture is absorbed by the plastic encapsulant material and, upon exposure to the soldering thermal shock, the moisture vaporizes, causing the package to delaminate or crack due to the resulting high internal pressures.

In order to determine if components will perform reliably after exposure to handling and assembly stresses, a preconditioning procedure emulating these processes should be developed and applied. Reference [6] describes a procedure generated to ensure that surface mount components can withstand severe printed wiring board assembly conditions and still meet expected reliability requirements. It can be used as a guide to define each test/procedure/operation/ material that is used in component handling and fabrication/assembly for establishing a process requirements procedure. This procedure should emulate all steps, from receipt of material through manufacturing. Additional or different preconditioning may be necessary for a specific process. After exposure, these devices should be subjected to appropriate testing to determine if performance degradation has occurred. Common tests for a molded plastic package include “85°C/85RH,” Highly Accelerated Stress Testing (HAST), Autoclave, and Dye Penetrant. For a hermetic device, seal testing should be part of the test procedure. Residual Gas Analysis (RGA) is also sometimes performed.

7.2.5 Parts Management Plan Evaluation Criteria

The following paragraphs provide guidelines which customers can use to evaluate a supplier's PMP. This evaluation includes an assessment of the quality improvement program, quality assurance, assembly processes, and design criteria. These guidelines are based on industry-accepted quality standards and are practiced by world-class organizations. These paragraphs are provided to express the level of detail desired, highlight the subjects of interest, and provide concrete guidelines. It is intended that suppliers clearly describe their own processes and explain how these processes develop, maintain and improve the reliability of equipment.

7.2.5.1 Quality Improvement Program

Quality is defined as providing customers with products and services that consistently meet their needs and expectations. But quality goes beyond that of the product to include quality of work, service, information, processes, people, and management. Control of quality in every aspect of the contractor's operation is the basic approach to total quality management.

A quality improvement program should be instituted to apply quantitative methods and human resources to control processes. The objective is to achieve continuous improvement in the selection and application of components, their installation in subassemblies, and in end user satisfaction. Each supplier should document their plan for achieving the goal of continuous improvement in the development, manufacture, reliability, administration, and distribution of products and services that meet the customer's needs and expectations.

7.2.5.2 Quality Assurance

Quality assurance is the corporate effort that is specifically aimed at reducing process variation by improving process controls during product development and manufacture, and by taking measures to prevent recurrence of detected problems. Quality assurance also addresses those techniques that will give the customer confidence that future components and assembly processes will have equivalent or better reliability than current components and assembly processes.

Assurance of component and assembly quality should be established before the part and assembly process is approved for use. Suppliers should have procedures for verifying that selected components will operate reliably in the intended environment for the life of the equipment. Component qualification processes should be documented or referenced in the PMP. Testing should be conducted to verify that the proposed components will satisfactorily operate in the intended environment with acceptable reliability. This verification usually takes the form of a qualification test and screening. However, other methods may be proposed, such as extensive field experience of the proposed parts in similar applications or previous contractor qualification data which is valid for the intended application. Furthermore, evidence of quality assembly processes should be demonstrated and documented to ensure reliability at higher levels of integration. The supplier should ensure that the component quality is maintained during the manufacture of equipment. The component reliability should not be compromised by equipment manufacturing processes such as handling or assembly.

7.2.5.2.1 Part Qualification

Qualification is used to verify that components are able to function for the specified life in the intended environment. The goal of qualification should be to ensure long term mechanical and electrical integrity. Qualification requirements may be satisfied by similarity to existing qualified devices of similar packaging and technology. The process for the disposition of failures during the qualification procedures should be at the discretion of the supplier, but should be identified in the PMP. The following items should be accounted for during component qualification:

(1) Hermetic and hygroscopic nature of unique package types
(2) Operating characteristics over entire temperature range
(3) Packaging capability for handling thermal shock
(4) Internal circuitry and connection resistance to contamination and corrosion (passivation)
(5) Internal connection fatigue life
(6) Levels of inherent contamination in packaging
(7) Solderability of leads

Detailed qualification processes should be documented or referenced in PMP and should address, as a minimum:

(1) Goals/objectives
(2) Procedures
(3) Test reports
(4) Pass/fail criteria
(5) Failure detection and analysis
(6) Requalification criteria
(7) Failure resolution/corrective action procedures

Qualification Testing- Accelerated environmental qualification testing may be proposed for all components if adequate field data does not exist to indicate long term reliability has been achieved. The environmental testing is to verify that reliability performance is within specification. Electrical characteristics for all potential environmental conditions should be verified through qualification testing if it is not already verified by the manufacturer.

Field Data- Component field data can be used in lieu of qualification testing when the data verify an acceptable failure rate. Component failure rates are generated by dividing total accumulated component failures by total accumulated hours. Component failure rates may also be calculated using industry-accepted prediction methodologies, such as those presented previously. Component types used for failure rate calculations should be of similar families, technologies or package types. The sample size should be statistically significant, with adequate field experience per component.

Component Manufacturer Data - Component manufacturer qualification data may also be used in lieu of field data, provided the data are adequately documented, statistically significant, and indicates that the components should function in the environment for the specified time. These data demonstrate that processes are in statistical process control and accelerated component testing data can be correlated to the intended application environment.

Component Reliability Assessment - Suppliers should have a plan for performing component reliability assessment. The formulas, data and assumptions used to generate the reliability assessments should be documented or referenced in the PMP. When required by contract, the supplier should explain to the customer how part reliability will allow the resulting product to meet or exceed the reliability requirements of the respective equipment performance specification.

When components are selected for use in an intended environment, a component quality and reliability assessment is necessary. The assessment technique and source of reliability data should be clearly defined or referenced in the PMP. The following reliability sections address only component reliability, and not assembly, LRU or system reliability assessments.

Reliability Analysis - A preliminary reliability analysis for each component should be performed prior to the preliminary design review and, as a minimum, should consist of a clear example of the content and format of the reliability analyses being proposed. The supplier is encouraged to base component reliability predictions on field data or other acceptable technical evaluations. Further, suppliers are encouraged to modify component reliability assessments based on methods used to improve the quality of components, such as component manufacturing process control, screening, qualification or other provisions. Failure rates based on the supplier's experience and modifications based on quality provisions should be available for customer review when required by contract.

A final reliability analysis for each component should be required at the critical design review. This analysis should be completed as early as possible, so that potential problems with parts selection or system architecture can be uncovered in time for a cost-effective correction.

Reliability Tracking - In order to perform root cause failure analysis and provide a basis for quality improvement, the component reliability and quality assessments should be verified on a continual basis. A verification should be made to show that the measured reliability exceeds the predicted reliability. This may include tracking field reliability measurements and analysis, tracking screen yield, and/or monitoring manufacturing floor rejects. Failure rate assessments should be updated for future reliability predictions, particularly when part reliability is measured to be less than predicted.

When field data are used to perform a part reliability assessment, the following information should be provided:

(1) Component MTBF formulas correlating field performance to MTBF should be available to the customer. Details of operating hours should be included, as well as component part numbers, equipment part numbers, and failure analysis results.

(2) Data submittals, if required, should include a summary of part types and failure mechanisms, and should include or reference the raw data used to arrive at these conclusions.

(3) Component failure rates should be generated by dividing total accumulated component failures by total accumulated hours. Component types used for failure rate calculations should be of similar families, technologies or package types.

(4) Accumulated operating hours and failures should be statistically significant to provide accurate failure rates. Suppliers should establish confidence intervals for the calculated failure rates using statistical techniques similar to the chi-square method.

(5) Continue to track component in-service data on an ongoing basis until equipment production is completed.

Requalification - Requalification of the component should be performed when significant changes (i.e., form, fit or function) are made to the package or internal circuitry. The following are examples of significant changes that would require requalification, but do not constitute a complete list.

(1) Changing the package material or component size.
(2) Changing the component fabrication process.
(3) Changing component materials.
(4) Changing lead finish/material.
(5) Internal circuit redesign.
(6) Changing the assembly plant.
(7) Substantial rejections from the field or infant mortality during testing.

The extent of the requalification should correspond to the changes made to the component. Partial qualification testing should be allowed, provided changed features are tested thoroughly. It is the purchaser's responsibility to establish the means of communication with component manufacturers such that major changes are identified to the purchaser in a timely fashion. The determination to requalify may be difficult if parts are procured through distributors, where design or material changes to the part may be unknown.

7.2.5.2.2 Production Quality Assurance

Components should perform initially with minimal infant mortality and latent defect rates. Verification should be provided that current and future component reliability is not compromised by unpredictable variations in the manufacturing process. Suppliers should strive to continuously detect and eliminate component flaws that result in infant mortality failures, or changes which may unpredictably degrade future lot quality.

Screening - Product assurance can be accomplished by 100% part screening, but alternative processes may be proposed, such as analyzing key process measurements of the component during manufacture or sample screening. The screening procedures, if applicable, can be performed by either the purchaser, the part vendor or a qualified screening house. Periodic screening failure reports should be available to customers.

Reduced Screening - Reduced screening may be considered when screening, factory and in-service rejections are measured and are found to consistently exhibit an acceptable defect density. Available data, including those from the device manufacturer, should be provided to indicate that the current level of screening is not required. Reduced screening may consist of sample screening, or a reduction of electrical testing and/or burn-in. However, to eliminate screening, some kind of quantitative measure of lot quality should be offered to ensure continuing quality. Approval of an alternate assurance method should be based upon scientific techniques and statistical analysis.

Historically, screening data indicates that part quality may change over time. Future part quality can be adequately assessed by measuring past part quality performance. The reduced screen criteria is aimed at measuring the level of part defects over a period of time, and then making a determination as to whether the level of defects is acceptable. The criteria stated in this section represents one possible baseline. Changes to the criteria can occur based upon experience and a partnership with vendors that would allow other innovative approaches to be considered. A generalized process flow appears in Figure 7.2-3.

At the start, parts should be qualified and screened. All failures before, during, and after screening should be recorded. These failures can be used to determine the level of defects in the tested parts.

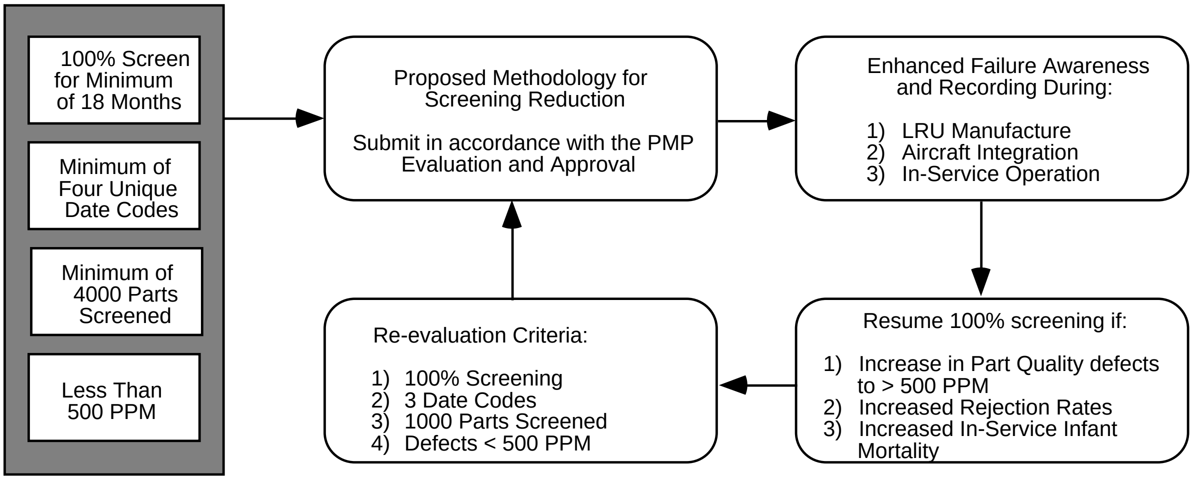


FIGURE 7.2-3: REDUCED SCREEN FLOW

In order to identify whether the part manufacturer consistently produces low defect levels in all lots (including future lots) and maintains configuration control of part specifications, the following data can be collected:

(1) Parts being screened should come from a minimum of four separate lots (date codes).
(2) 100% screening should be performed for at least 18 months.
(3) 4,000 parts, minimum, should have been screened.

The defects measured should be below 500 parts per million (500 PPM = 1 failure in 2,000 parts). If the sample of parts tested has more than 500 PPM, then the reduction or elimination of screening should not be allowed. Failing this criteria indicates a possibility that future parts may also have more flaws than are acceptable.

As part of the original PMP, a failure recording system should be developed and implemented that will record failures during sample screening, equipment/item assembly, environmental stress screening, and field operation. If these measurements indicate a decline in part quality, 100% screening can be reinstituted until 500 PPM quality is re-established.

Screening Documentation - Detailed screening processes should be documented in the supplier's program plan and should address, as a minimum:

(1) Goals and objectives
(2) Test methods
(3) Data collection techniques
(4) Test reports
(5) Pass/fail criteria
(6) Failure detection and analysis
(7) Failure resolution and corrective action procedures

Data Retention - Results of all screening, qualification, tests, inspections, field history and failure analysis should be recorded and maintained on file. It is recognized that some yield, qualification or screening data may be proprietary to the component manufacturer. Suppliers should still collect and retain some evidence of component quality.

7.2.5.3 Assembly Processes

Equipment manufacturing processes contribute to equipment reliability. Thus, when reviewing the process for selecting components, an assessment of the ability to manufacture the assembly using the proposed technology should be accomplished. The overall goal is to ensure that manufacturing processes are mature.

Processes In Manufacturing - A verification should be made that all manufacturing processes involving electronic components are mature. Further, the supplier should implement continuous improvement goals and quality assurance requirements.
This portion of the parts management plan should include a definition of the manufacturing processes used, how the piece parts flow through these processes, and where process controls are used. The use of statistical process control, design of experiments, and other methods of process control should be documented.

Process Maturity - Suppliers should document their ability to use the proposed processes successfully. If the proposed manufacturing techniques have been used on other products, identification of these existing processes, and a simple statement that these processes are in control and capable, should be adequate. If new techniques are being proposed (such as a change to surface mount technology), demonstration of process control and capability should be required. Suppliers should list the activities performed to identify all of the key process parameters, measurement criteria, and manufacturing procedures needed to minimize the learning curve during production. Examples of these activities include:

(1) Development of manufacturing procedures
(2) Personnel training
(3) Identification of process measurements
(4) Development of pass/fail criteria
(5) Design of experiments
(6) Product life testing after assembly

Process Control and Capability - The next item that should be demonstrated is a supplier's ability to keep their processes in statistical control and capable. Guidelines are provided in Reference [7].

Component Packaging - Maintainability of equipment can be enhanced through the use of standard part package types. Therefore, suppliers are encouraged to procure components with standard package outlines. Standard package outlines are contained in Reference [8].

Component Marking - Components which undergo screening should be permanently marked to indicate the individual component has received quality assurance testing. Markings should be visible when components are mounted on the PC board. This helps prevent components without quality assurance from being accidentally installed at the factory or at a remote repair facility.

Components should also be permanently and legibly marked by the manufacturer with the following information, where space allows:

(1) Manufacturers name, trademark or logo
(2) Manufacturers part number
(3) Inspection lot identification or date code
(4) Pin 1 locator or orientation designator

Components without adequate space for marking should have provisions to preclude accidental replacement with a different part. All component marking should be able to withstand normal use in the planned environment. For military products, marking should be able to pass the resistance-to-solvents test of MIL-STD-883, Method 2015.

Component Handling - Component quality assurance measures can be easily compromised by improper handling. Thus, the contractor PMP plan should reflect practical and proven handling procedures to maintain component quality. Considerations may include ESD prevention, lead formation maintenance and automated handling practices for standard and non-standard packages and humidity control (i.e., PEMs).

All components should be shipped in appropriate packing material. The program plan should address component handling issues, such as ESD, installation orientation, mounting techniques (tube, reel, etc.), contamination and humidity.

Procurement and Shipping Procedures - Component quality should not be degraded during handling or screening. Handling or screening provisions placed in effect with third party participants, such as manufacturers, distributors or screening facilities, should be identified or referenced. Suppliers should be encouraged to eliminate unnecessary processing which may degrade component quality.

The component manufacturer or screening house should obtain and keep on file written certification of specified shipments of components. The shipment certificate should include:

(1) Manufacturer name and address
(2) Customer or distributor name and address
(3) Component type
(4) Date code and latest re-inspection date, if applicable
(5) Quantity of components in the shipment
(6) Level of screen and specification reference

To maintain quality as shipped from the part manufacturer, date codes should be no older than 12 months from receiving date of the purchase order by the manufacturer.

The supplier should ensure that the components are the correct type, date code, screening level, and package type prior to stocking. Incoming inspection is one way to ensure that the components are received with the proper information. This procedure should be properly identified in the PMP.

Discrepancy controls for non-conforming materials should be implemented. These controls should include flow charts describing corrective actions, actions taken to prevent recurrence of discrepancies, etc.

Storage Procedures - The PMP should also address relevant storage and stocking procedures. For instance, plastic packages absorb moisture over time, which may cause package cracking during the solder process. Dry storage may be necessary up until the time of soldering. An alternative process would involve a thermal pre-bake to drive out excessive moisture.

References [9] and [10] can be used in determining the sensitivity of particular ICs to moisture-induced package cracking.

Rotation of stock is also an important function of the storage process. The supplier's plan should identify how their process controls stock flow (i.e., First In/First Out, Last In/First Out, etc.).

Modification and Repair of PCBs and Assemblies - Repair and modification techniques for surface mounted components can be complicated, and may require special tooling and processes. Thus, the program plan should identify the governing documents and procedures for the modification and repair of PC boards.

7.2.5.4 Design Criteria

A reliability program should provide Line Replaceable Unit/Line Replaceable Module (LRU/LRM) design guidance and control early in an equipment design program. Misapplication of any part can affect the reliability and performance of that part. Many parts have unique packaging and performance characteristics that should be accounted for in the design of the equipment.

LRU/LRM design guidance should address such issues as thermal stresses, contamination and electrical derating. Appropriate industry standards or proven "in-house" standards should be followed rigorously. The parts management plan should reference these design standards and analytical methods. Design criteria should embody lessons learned by the supplier.

The reliability of components can be greatly improved by using the best equipment design standards and techniques available. Early equipment design analyses not only gives the customer confidence in the product, but gives the supplier time to implement design changes in an orderly and cost-effective manner. This early analysis is an essential practice with the use of new technology parts.

Many of the design requirements are already defined by the equipment specification or other design guidelines. The objective is to have the existing design standards referenced, rather than being detailed in the contractor's PMP.

Electronic Parts Selection ListThe contractor should prepare a parts selection list. The list should be initially submitted at the preliminary design review. This list is considered preliminary and should be updated as the design matures. The parts list should specify whether "preferred" or "non-preferred" parts (definitions follow) are being used.

Preferred Parts Selection - Preferred parts are those parts for which the contractor has demonstrated a successful history of use.

Non-Preferred Parts (NPPs) - Many component manufacturers are now producing high quality new technology components. If the reliability of these new technology parts can be shown to be acceptable in the intended environment, adequate quality assurance provisions exist which will ensure future production, and application of these NPPs will meet or exceed current reliability performance requirements, then these parts can be considered.

Component Descriptions **-** Components can be procured under a variety of product descriptions which include commercial item descriptions (CIDs), program-specific documents, and defense detail specifications (MIL-DTL). The selected component description should provide configuration (and interchangability) control such that the manufacturer, supplier or distributor guarantees the electrical operating characteristics and package specifications.

Components should be tested to supplier requirements under control of the component specification. Lot tolerance percent defective, or other quality and performance guarantees, can be specified in the component description, and should be contracted with the screening house or vendor. The PMP plan should also identify the disposition of failed lots. Tips for selecting and developing product descriptions are presented in reference [11] "Buying Commercial and Nondevelopmental items: A Handbook."

Notification of Change **-** The component should be controlled to the greatest extent possible through a system of change control. Requalification may be necessary based on the significance of the change. Part specifications should be documented and performance to those specifications guaranteed.

The program plan should define the supplier's "notice of change" agreement. The agreement should ensure that the component is under configuration control at all times, and that quality is not compromised by manufacturer process changes. The component description should require the component vendor or distributor to notify the supplier of component process or design changes.

Thermal/Fatigue Analysis and Measurements - The equipment contractor should provide an engineering analysis of the component thermal operating characteristics early in the equipment design process, followed by a thermal test to verify the analysis accuracy. Equipment thermal requirements may necessitate a thermal management program, a thermal analysis, and a cooling evaluation. Thermal cycling fatigue analysis should also be accounted for in the design. This analysis may account for lead compliance during thermal cycling and identify coefficient of expansion mismatches.

7.3 Derating

Derating can be defined as the operation of an item at less severe stresses than those for which it is rated. In practice, derating can be accomplished by either reducing stresses or by increasing the strength of the part. Selecting a part of greater strength is usually the most practical approach.

Derating is effective because the failure rate of most parts tends to decrease as the applied stress levels are decreased below the rated value. The reverse is also true. The failure rate increases when a part is subjected to higher stresses and temperature. The failure rate model of most parts is stress and temperature dependent.

7.3.1 Electronic Part Derating

Achieving high equipment reliability requires that each electronic part be both properly applied and capable of withstanding all of the stresses to which it will be subjected. Thus proper derating of electronic parts is a powerful tool for enhancing equipment reliability.

Electronic part derating is done with reference to the “Absolute Maximum Ratings.” These ratings are defined in the manufacturer's specification or data sheet as those values which: “should not be exceeded under any service or test condition.” There are various "absolute maximum ratings" for each part: voltage, current and power, etc. Each absolute maximum ratings is unique. It is applied individually, not in combination with other absolute maximum rating. Absolute maximum ratings include both operating and storage temperatures, e.g., the maximum junction or hot spot temperature. The “absolute maximum ratings” are typically based upon “DC power conditions measured in free air at 25°C.”

Electronic part reliability is a function of both electrical and thermal stresses. Increased thermal stresses generate higher junction temperatures. The result is increased chemical activity within the part as described by the Arrhenius Reaction Rate Model and thus in an increased failure rate. Electronic part reliability is largely determined by the thermal stress.

The specific parameters to be derated vary with different types of parts as shown in Table 7.3-1. Capacitors are derated by reducing the applied voltage to a stated percentage of the absolute maximum rated. Transistors are derated by reducing applied voltage to avoid voltage breakdown, reducing output current, power dissipation and maximum junction temperature. A sample set of derating values for transistors is shown in Table 7.3-2.

TABLE 7.3-1: PRINCIPLE RELIABILITY DEPENDENT STRESS FACTORS/DERATING FACTORS

COMPONENT FAMILY	TEMPERATURE °C	VOLTAGE	CURRENT	POWER	OTHER
Capacitors	Ambient	Ripple & Transient			
Circuit Breakers	Ambient		Contact		Load type
Connectors	Insert	Dielectric withstanding	Contact		
Crystals				Input	
Diodes	Junction	Reverse & Peak Inverse Voltage	Surge, Forward, Zener	Dissipation	
EMI & RF Filters	Ambient	Maximum Operating	Maximum Operating		
Fuses	Ambient	Maximum Operating	Surge		
Inductive Devices, Transformers	Hotspot	Dielectric withstanding	Maximum Operating		
Microcircuits	Junction	Supply & Input Signal	Output & Load	Dissipation	Frequency Fanout
Relays	Ambient		Contact		Load type Cycle Rate
Resistors	Hotspot	Maximum Operating		Dissipation	
Switches	Ambient		Contact		Load type Cycle Rate
Thermistors	Maximum Operating			Dissipation	
Transistors	Junction	Breakdown, V_{CB}, V_{CE}, V_{BE}	Output	Dissipation	Safe operating area

TABLE 7.3-2: DERATING VALUES FOR TRANSISTORS[1]

PART TYPE	DERATING PARAMETER	DERATING LEVEL I (Space)	II (Airborne)	III (Ground)
TRANSISTORS	On-State Current (I_t - % Rated)	50%	70%	70%
• Thyristors	Off-State Voltage (V_{DM} - % Rated)	70%	70%	70%
(SCR/TRIAC)	Max T (°C)	95°	105°	125°
• Field Effect	Power Dissipation (% Rated)	50%	60%	70%
	Breakdown Voltage (% Rated)	60%	70%	70%
	Max T (°C)	95°	105°	125°
• Bipolar	Power Dissipation (% Rated)	50%	60%	70%
	Breakdown Voltage (% Rated)	60%	70%	70%
	Max T (°C)	95°	105°	125°

It is imperative that derating be cost effective. If derating is excessively conservative (e.g., lower than necessary part stresses are applied) part costs rise severely. At optimum derating, a rapid increase in failure rate is usually noted for a small increase in temperature or stress. However, there is usually a practical minimum derating value. Below this minimum stress level, circuit complexity increases drastically, offsetting any reliability gain achieved by further derating.

Derating helps to compensate for many of the variables inherent in any design. Electronic parts produced on an assembly line are not all identical. There are subtle differences and variations from one part to the next. Proper part derating helps to compensate for part-to-part variations and alleviate their impact upon equipment reliability. Electronic parts with identical part numbers may be purchased from a variety of suppliers. While these items are "electrically interchangeable" there may be significant design, material and manufacturing differences between them. Derating also compensates for these differences. Furthermore, critical part parameters are not necessarily stable over their entire life. Proper derating will help assure proper circuit operation in spite of these part parameter changes.

Data on failure rates vs. stress is available for a number of electronic parts. This data can be used to determine the reliability improvement through derating. The same is not true of mechanical and structural parts, as can be seen in the following subsection.

7.3.2 Derating of Mechanical and Structural Components

For mechanical and structural components, such failure rate versus stress data may be obtainable from the manufacturer or users, but time rate data may not be available. In using a manufacturer's rating and single design stress values, the design engineer must keep in mind that they are really distributions, not single values. Either the worst case "tolerances" for both stress

[1] Rome Laboratory, Part Derating Guide

and strength or a plot of the distributions must be utilized. When there is time dependency for the distributions (e.g., degradation, wear out), the stress and strength distributions must be related in the appropriate manner to the cyclic or time operation in the intended environment.

The classical approach to mechanical and structural design is to give every part enough strength to handle the worst stress it will encounter. Several references, such as MIL-HDBK-5 are available, providing data on the strength of materials. Some of these provide limited data on strength degradation with time, resulting from fatigue. Effective design procedures should provide for evaluating alternative configurations with respect to reliability. Since failure is not always related to time, the designer needs techniques for comparing stress vs. strength, and determining the quantitative reliability measure of the design. The traditional use of safety factors and safety margins is inadequate for providing a reliability measure of design integrity. The concept of stress strength in design recognizes the reality that loads or stresses and strengths of particular items subjected to these stresses cannot be identified as a specific value but have ranges of values with a probability of occurrence associated with each value in the range. The ranges of values (variables) may be described with appropriate statistical distributions for the item. Stress/strength design requires knowledge of these distributions. After the strength and stress distributions are determined, a probabilistic approach can be used to calculate the quantitative reliability measure of the design, including confidence limits.

To illustrate the concept of stress and strength distributions related to reliable design, assume that a large number of tests of the strength of a given manufactured item have been run, with each test being run to failure. A relationship (frequency distribution) between the number failing at any particular value of strength (or band of values) and the value can be determined. Figure 7.3-1(a) shows a generalized frequency distribution of the results. If the exact relationship were known, the probability of a randomly selected specimen failing at a particular value of stress F' could be predicted. It would be that fraction of the population, whose strength was equal to or less than a stress F'. Similarly if a large number of experiments were conducted, and the stress was recorded on each experiment, a relationship between the relative frequency of stresses and the stress can be established. This relationship is shown in Figure 7.3-1(b). If the exact relationship were known, the probability that on any randomly selected trial the stress would exceed a strength S' could be predicted. This would be the fraction of the population (of possible trials) in which the stress exceeded the strength S'. With both of these distributions defined, unreliability is determined as the probability that the stress is greater than the strength. Unreliability can be determined analytically, graphically, by numerical integration or by probabilistic techniques such as "Monte Carlo" provided the form or shape of the two probability distribution functions are known. The curves from Figure 7.3-1(a) and 7.3-1(b) are combined in Figure 7.3-1(c) to illustrate the region of the unreliability given by the shaded area where stress exceeds strength. Figure 7.3-2 illustrates normal (gaussian) stress and strength distributions, where the stress and strength variables are identified as Kips (a thousand pounds).

Looking at Figure 7.3-2, two things may happen with time and repeated stress. The variance of the strength distribution may change; for example the curve may extend from 13 to 23 Kips rather than the original 16 to 20 Kips. This would result in an increased unreliability since the shaded area would now extend from 13 to 20 Kips. This is shown in Figure 7.3-3(a). The other factor that could change with time and stress is that the mean of the strength distribution might be lowered, to say 15 Kips. This, in turn, would result in a decreased reliability as shown by the shaded area of Figure 7.3-3(b).

The purpose of stress strength analysis is to improve the reliability of the design. That is, to find the optimum comparison of stress and strength that will have an acceptable probability of success and compete favorably with other constraints such as weight, cost, and availability of material.

There are four basic procedures the designer may use to increase reliability.

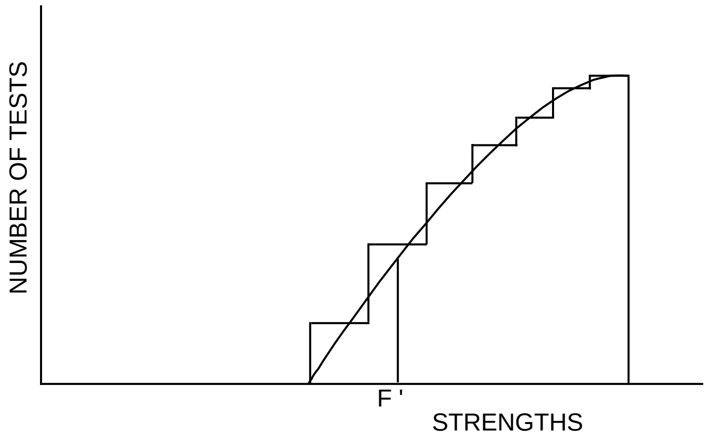
(1) Increase Average Strength: This approach is tolerable if size, weight, and cost increases can be accepted or if a stronger material is available.

(2) Decrease Average Stress: Occasionally the average allowable stress on a component can be reduced without greatly affecting its performance.

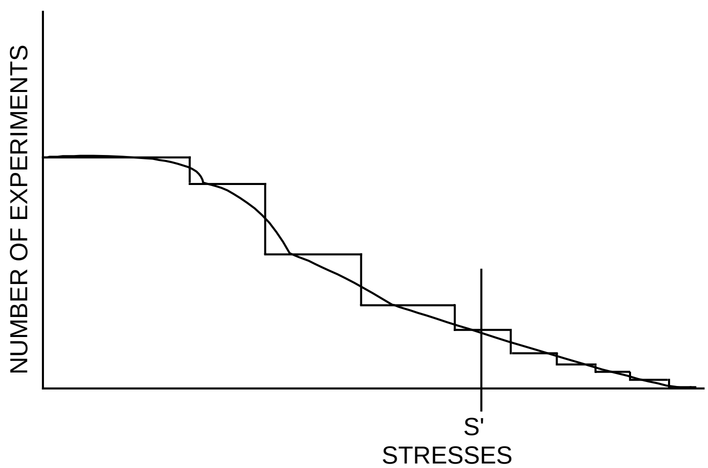
(3) Decrease Stress Variation: The variation in stress is usually hard to control. However, the stress distribution can be effectively truncated by putting limitations on use conditions.

(4) Decrease Strength Variation: The inherent part-to-part variation in strength can be reduced by improving the basic process, holding tighter control over the process, or by utilizing tests to eliminate the less desirable parts.

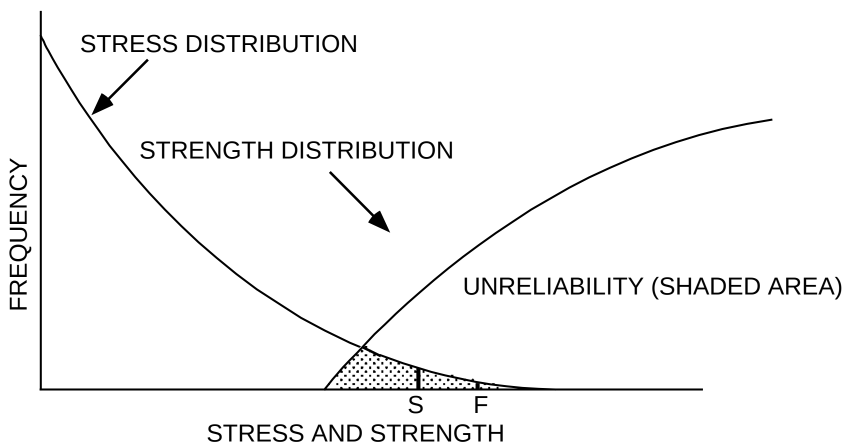
References [12], [13] and [14] provide more details on this procedure and its application to mechanical and structural components.



a) Strength Frequency Distribution



b) Stress Frequency Distribution



c) Probability of Stress Exceeding Strength

FIGURE 7.3-1: STRESS-STRENGTH DISTRIBUTIONS AND UNRELIABILITY IN DESIGN

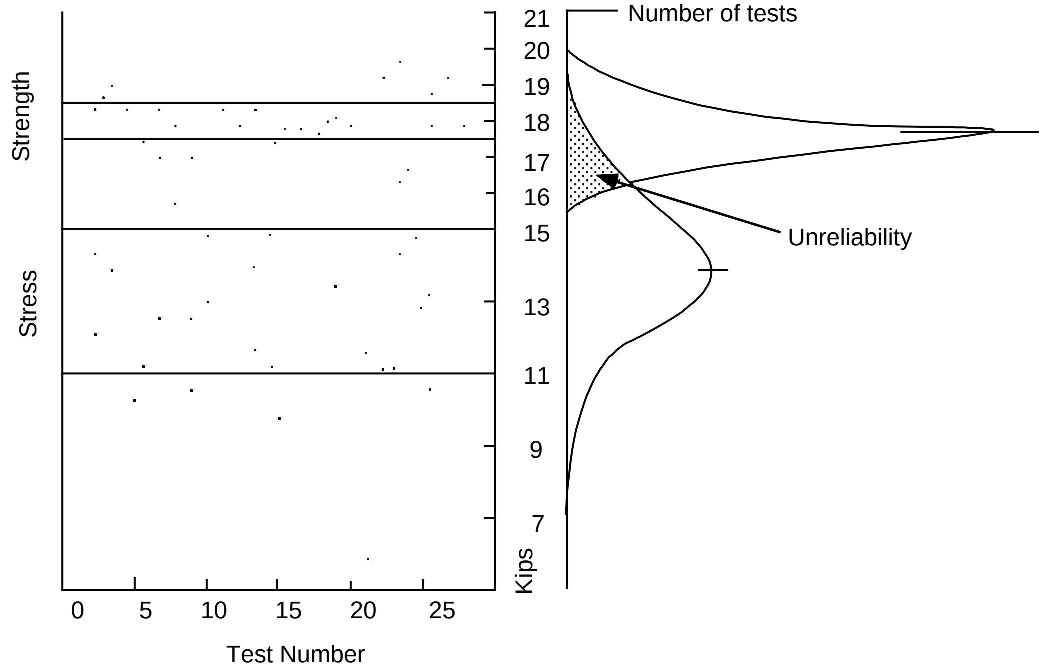
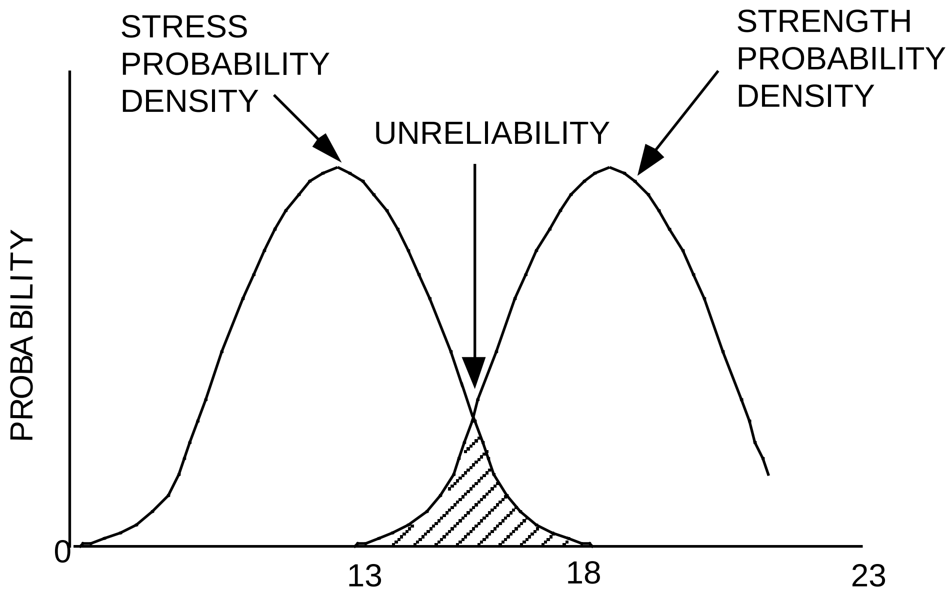
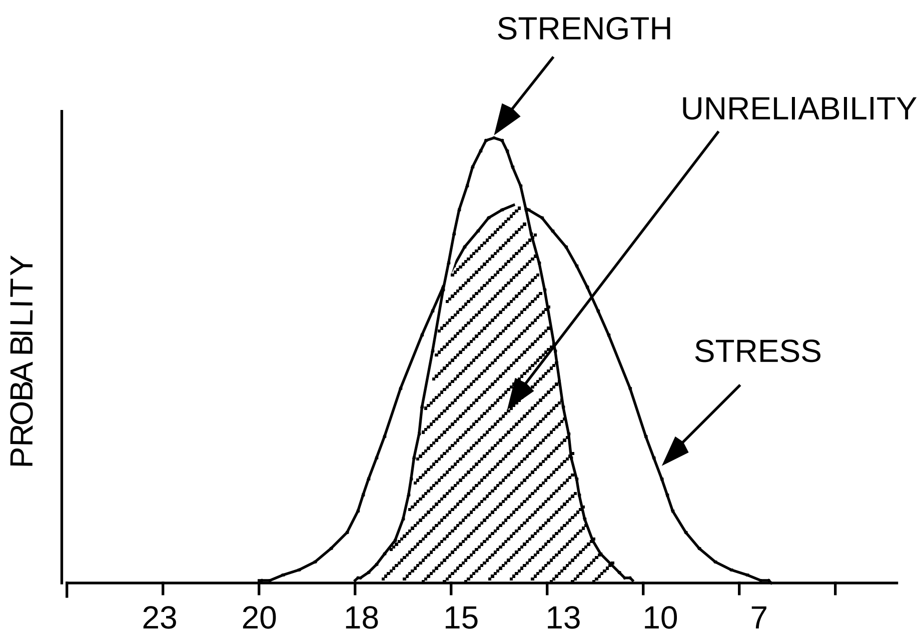


FIGURE 7.3-2: NORMAL (GAUSSIAN) STRESS-STRENGTH DISTRIBUTIONS AND UNRELIABILITY IN DESIGN



(a) Result of Increase of Variance in Strength with Time & Stress



(b) Result in Decrease in Strength with Time & Stress

FIGURE 7.3-3: FACTORS AFFECTING UNRELIABILITY

7.4 Reliable Circuit Design

This section cannot possibly cover all of the aspects of circuit design. In addition to a number of design textbooks, there are handbooks available (e.g., References [15] and [16]) which can be used to solve almost any circuit design problem.

The only thing that this section can accomplish in the limited space available is to outline some of the circuit design methods available to ensure high reliability. They are by no means comprehensive; circuit designers should consult their own organizations' design rules, component application notes, the cited references and other relevant sources. The methods outlined in this section are intended as a guide to the points which reliability engineers and circuit designers need to consider.

In order to produce a reliable circuit design, the designer must consider the following reliability design criteria:

(1) Component derating (discussed in the previous section)

(2) Proper use of parts (discussed in 7.2)

(3) Transient and overstress protection

(4) Parameter degradation and analysis

(5) Fundamental design limitations

Except for component derating, which was discussed in the previous section and parts use, which was discussed in 7.2, the following paragraphs discuss each of the listed criteria.

7.4.1 Transient and Overstress Protection

Electronic components are often prone to damage by short-duration voltage transients, caused by switching of loads, capacitive or inductive effects, static electricity, power supply ripple, testing, etc. Small semiconductor components are particularly vulnerable, owing to the very low thermal inertia of their wire bonds. MOS devices are very vulnerable to static electricity, and require special protection.

The subject of electrostatic discharge (ESD) is treated very thoroughly in other sources, and will only be summarized here. It is becoming an increasingly important and recognizable problem with the trend toward the development of integrated circuits of greater complexity and higher component densities. Some of today's microcircuits can be damaged by ESD voltages as low as 20 volts. The smaller the part, the less power it can dissipate or the lower the breakdown voltage, and the more likely it is to be damaged by an electrostatic discharge (ESD). Certain parts are considered highly susceptible and their chances for damage are great. These include metal oxide semiconductor (MOS) parts with a direct access to the MOS junction, high frequency parts produced by the Schottky barrier process, many bipolar and field-effect microcircuits like RAMs, ROMs, and PROMs utilizing small active area junctions, thin dielectrics, metallization crossovers, and N+ guard ring structures, precision film resistors and similar parts. A detailed list of electrostatic discharge sensitive (ESDS) parts and their voltage sensitivity ranges are provided in MIL-STD-1686 and MIL-HDBK-263. They also describe control programs that can be applied to minimize component failures due to ESD.

In addition to ESD, the designer must cope with the other causes of transient generation described in the first paragraph.

Semiconductor device circuit malfunctions can arise from two general sources: (1) transient circuit disturbances and (2) component burnout. Generally, transient upsets are the controlling factors, because they can occur at much lower energy levels.

Transients in circuits can prove troublesome in many ways. Flip-flop and Schmitt triggers can be inadvertently triggered, counters can change count, memory can be altered due to driving current or direct magnetic field effect, one-shot multivibrators can pulse, the transient can be amplified and interpreted as a control signal, switches can change state, semiconductors can latch-up, requiring reset, etc. The effect can be caused by transients at the input terminals, output terminals, on the supply terminals, or on combinations of these. Transient upset effects can be generally characterized as follows:

(1) Circuit threshold regions for upset are very narrow. That is, there is a very small voltage amplitude difference between signals which have no probability of causing upset and signals which will certainly cause upset.

(2) The dc threshold for response to a very slow input swing is calculable from the basic circuit schematic. This can establish an accurate bound for transients that exceed the dc threshold for times longer than the circuit propagation delay (a manufacturer's specification).

(3) Transient upsets are remarkably independent of the exact waveshape, and depend largely on the peak value of the transient and the time duration over which the transient exceeds the dc threshold. This waveform independence allows relatively easy experimental determination of circuit behavior with simple waveforms (square pulse).

(4) The input leads (or signal reference leads) are generally the ones most susceptible to transient upset.

Logic devices which interface with inductive or capacitive loads, or which "see" test connections, require transient voltage protection. This can be provided by a capacitor between the voltage line to be protected and ground to absorb high frequency transients, by diode protection to prevent voltages from rising beyond a fixed value (clamping), or by series resistances to limit current values.

The transient voltage levels which can cause failure of semiconductor devices are referred to as VZAP. VZAP values depend upon transient duration. Passive devices can also be damaged by transient voltages, but the energy levels required are much higher than for small semiconductor devices. Therefore, passive devices do not normally need individual protection.

7.4.1.1 On-Chip Protection Networks

On-chip protection networks for integrated circuits incorporate many of the principles that apply to equipment protection. It is appropriate therefore to discuss some of these principles before describing discrete devices that are used for protection from transients. The basic approach is to utilize clamps and attenuators that reduce current and voltage transients and protect internal components from excessive thermal dissipation or voltage levels capable of rupturing insulating layers.

A simple yet very effective protection network consists of a diode connected between an input terminal and ground. The diode is designed to clamp the input voltage to the gate to a level below the breakdown voltage for any input less than the design goal. Figure 7.4-1 shows the diagram and illustrates the voltage transfer function between the voltage source V_s and the internal gate being protected.

For negative values of input voltage V_s the voltage surge or ESD transient is conducted to ground with the gate voltage V_G increasing to one forward diode voltage drop. For positive voltages less than the diode breakdown voltage V_{BR} the protection diode is transparent and the gate voltage responds as it would to any signal voltage (note that any noise could be interpreted as a signal). If the transient exceeds V_{BR}, the diode goes into the reverse breakdown region, and further increases in the input voltage are attenuated by the voltage divider consisting of the diode incremental resistance and the source resistance R_S. The object is to prevent V_G from reaching the destructive level BV_{oxide} for any anticipated value of the input voltage V_S.

On-chip protection is the least expensive and best way to improve device and system level hardness and provide maximum customer satisfaction. Nevertheless, sensitive parts do exist, and their use compels the equipment designer to employ effective techniques for best performance and reliability.

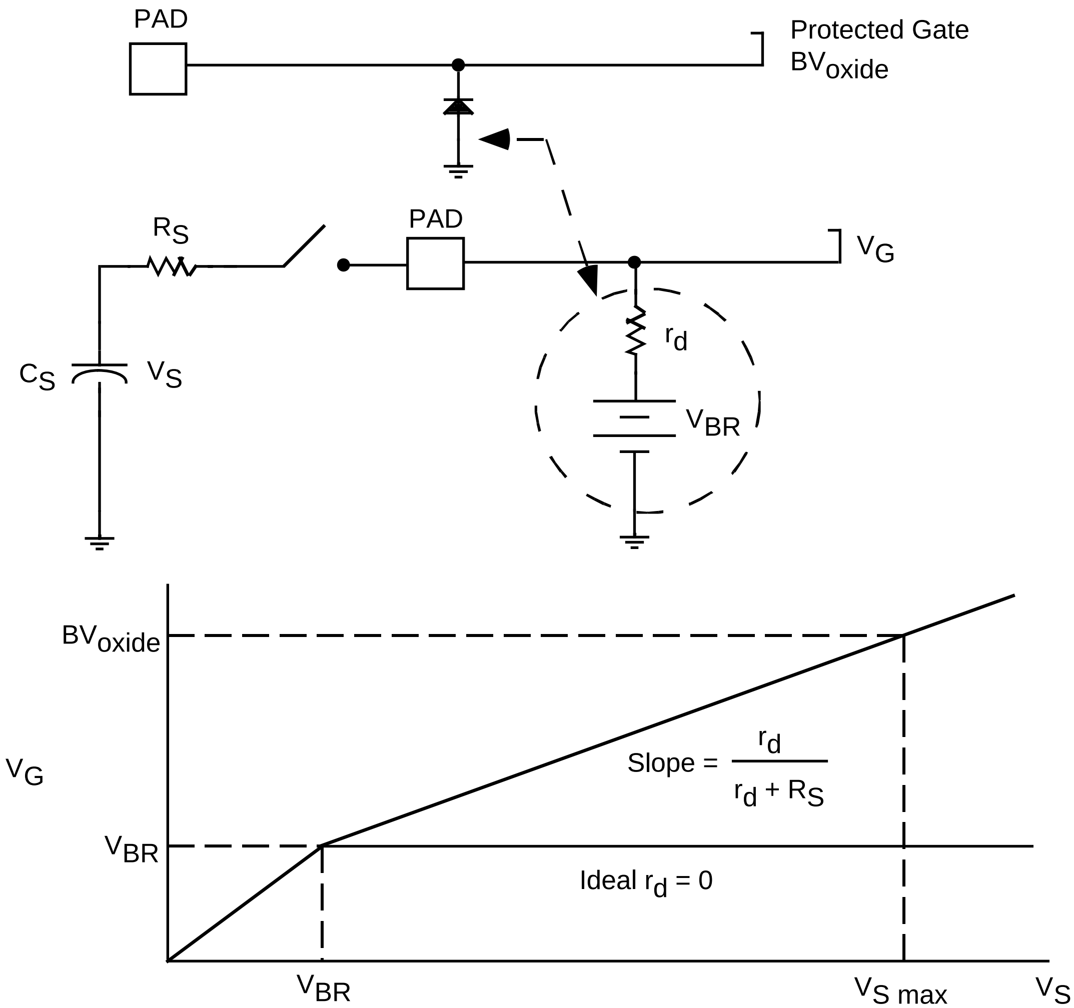


FIGURE 7.4-1: ON-CHIP DIODE PROTECTION CIRCUIT

7.4.1.2 Metal Oxide Varistors (MOVs)

A varistor is a variable resistor whose value depends on the voltage (or current). It is, in other words, a non-linear resistive element that serves the dual function of clamping and energy absorption. Invented in Japan, it is a widely used, effective, and low cost solution to the problem of controlling voltage surges, especially on power lines.

The metal oxide varistor, or MOV, consists mostly of ZnO with a small amount of other oxide additives. The powdered material is pressed and sintered to form various shapes and sizes according to the intended application. ZnO is conductive, and the individual grains are separated by an insulating layer that produces a p-n junction-like characteristic. At about 3 volts across the insulating layer a tunneling mechanism provides electrons for carrying the current. The nominal "clamping" voltage is determined by the number of grain boundaries between the electrodes.

The conduction is highly non-linear and the current and voltage are related by the equation:

$$I = CV^{\alpha}$$

where C depends on the cross section area and thickness and α is a constant between 20 and 50. A plot of this equation on a linear scale resembles the characteristics of a back-to-back diode configuration.

The current rating of an MOV depends on its area, and the energy absorbed depends on the volume. Since energy is absorbed uniformly throughout its volume, the energy rating can be substantial. The speed of these devices is excellent, limited primarily by the lead inductance. There is a small aging effect. MOVs generally have low leakage current, even at elevated temperatures, and good tolerance to radiation environments.

MOVs are available in many sizes, shapes, and ratings, for applications in both ac and dc systems ranging from several volts to several thousand volts. Peak current ratings range from a few tens of amperes to several tens of kiloamperes, with energy capabilities from fractions of a joule to 10 kilojoules. When one compares these energies with the millijoules in an ESD pulse capable of destroying an integrated circuit, it becomes clear that MOVs are not a good choice for protecting the inputs of integrated circuits from ESD. In addition, the capacitance of even the smallest MOVs is fractions of a nanofarad and would degrade the high frequency performance. Nevertheless, protection of power ICs or of circuit boards is a possible application, and MOVs are available as coaxial inserts for connector pins.

The largest application for MOV surge suppressors is protection against power line surges. Many computers, appliances, power cord extensions, etc., are equipped with MOVs for protection from routine power line transients.

7.4.1.3 Protective Diodes

Discrete diodes and diode networks can be used in a multitude of applications for protection in power lines, signal lines, and equipment from transients caused by lightning or lightning induced effects, power line surges, inductive switching, ESD, or EMP. The sophisticated processing technology of the silicon semiconductor industry makes possible such a wide variety of applications.

Turn-on time for silicon pn junctions is extremely short, claimed to be a picosecond or so in theory, with actual response time limited by lead inductance. The series resistance is very small, leading to a minimal increase in clamping voltage with current. Clamping voltage ranges from several volts to several hundred volts. Pulse power rating at 100 ns ranges from a few kilowatts to over 10 megawatts. From 100 ns to 10 ms the power rating follows the classical Wunsch-Bell model, decreasing as $t^{-1/2}$. Power is derated linearly above 25°C to zero at 175°C. Since the series resistance is so small virtually all of the power is dissipated in the depletion layer of the pn junction, which ranges from small fractions of a micron at lower voltage ratings to several tens of microns at 500 volts. Diode capacitance can be as low as several tens of picofarads for high voltage diodes, and an order of magnitude larger at lower voltages. Many different packaging styles are available, including arrays in ceramic dual-in-line packages, hermetically sealed.

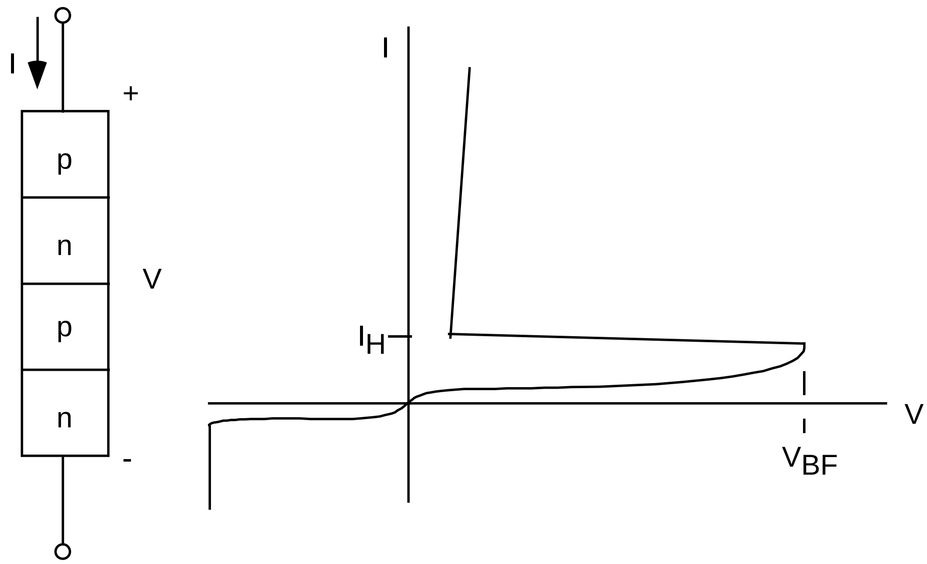
7.4.1.4 Silicon Controlled Rectifier Protection

A silicon controlled rectifier (SCR) is a four-layer silicon device employing regenerative feedback to achieve a snap-back characteristic that offers excellent circuit protection and low power dissipation. High power devices have been in use for many years as a control device for lighting, motors, and voltage control. Their use in protection networks for integrated circuits is more recent.

The four-layer structure of an SCR is shown in Figure 7.4-2 together with its current-voltage characteristic. For positive values of voltage a very small leakage current flows until the forward breakdown voltage V_{BF} is reached, whereupon the device switches into a low-voltage, high-current conduction state. It remains in this state until the transient surge decreases to a low value where the SCR current falls below the holding current I_H. The SCR then reverts to its normal blocking state.

In the blocking state the np junction is reverse biased and sustains the large voltage. Once triggering occurs the junction voltage collapses, and both transistors saturate. This leads to a small voltage across the device in the conducting state. By adjusting the doping levels the breakdown voltage can be varied over a wide range. In integrated circuit form there are large parasitic resistors that complicate the design; nevertheless the basic ideas are the same.

The holding current is a very important parameter in any SCR. The device cannot resume its non-conducting state unless the current falls below I_H. In an ac circuit this is accomplished by the normal reversal of voltage every half cycle. In surge suppression this requires that the transient be reduced to a level where the source resistance or other resistance be sufficient to limit the current to less than I_H when the transient has subsided, otherwise the continued dissipation may destroy the device (as in latch-up).



(a) (b)

FIGURE 7.4-2: (A) FOUR-LAYER STRUCTURE OF AN SCR
(B) CURRENT - VOLTAGE CHARACTERISTIC

The key that makes the SCR such an effective clamping device is its very low voltage in the conducting state, together with its very low incremental resistance. As a result it can conduct large currents with very little power dissipation. In effect, it is a “crowbar” device once triggered.

Besides its use in on-chip protection, discrete devices and arrays are available for a large variety of applications, from protecting integrated circuits and circuit boards to surge protection and control in high-voltage, high-current environments.

7.4.1.5 Passive Component Protection

Discrete components can also be useful in reducing susceptibility to transient electrical overstress. To be effective, they must function in concert with other impedances. For example, a resistor in series with the input impedance to an integrated circuit can form a voltage divider network to attenuate the transient and absorb part of the energy. Similarly, a resistor across the line could act with the source resistance to attenuate a surge. On the other hand, if the source were a true voltage source, then shunt elements would have no effect. Furthermore, since a linear device affects desired signals as well as transients, it may not be feasible to use a purely linear network, especially if the frequency spectrum of the signals overlaps the spectrum of the disturbance. Nevertheless, resistors, capacitors, and inductors can be an inexpensive means to achieve the desired result.

There are several types of resistors, the common ones being carbon composition resistors, film resistors, and wirewound resistors. The conducting element in a carbon composition resistor is made of a silica-loaded resin with dispersed carbon particles that is formed into a slug or pellet. Because the thermal mass of the slug is relatively large, it can absorb a considerable amount of energy. Experimental data show that a 100 Ω, 1/8 W carbon composition resistor can dissipate 1 Mw of power for 1 ms before exhibiting a resistance change greater than 5%. The damage threshold follows a t^{-1} dependence to 100 ms, where the threshold is about 10 watts. The energy absorbed in this range is several orders of magnitude greater than the threshold for integrated circuits.

Note that the power level in the preceding paragraph corresponds to a voltage of 10 kV across the 100 W resistor, far in excess of the rated value of 150 volts. Nevertheless, unless the power dissipation results in catastrophic thermal failure or flashover, the resistor will remain functional and continue to offer protection.

At high frequencies the capacitance and inductance of the resistor must be considered. A typical value for the parasitic capacitance is 1.6 pF. For a low value resistor, less than 100 ohms, the capacitive reactance is negligible below several hundred MHz. For higher values the upper cutoff frequency can be as low as 10 MHz. If the resistor is used in a shunt arrangement the capacitance would aid in transient suppression by blunting very fast wavefronts; on the other hand, in a series arrangement the capacitance would be deleterious, exposing a sensitive integrated circuit to the full leading edge spike.

The parasitic inductance of a carbon composition resistor depends on the size of the conducting slug and the length of the leads. In a practical sense the larger the conductor the lower the inductance, with hollow or square conductors the best shape. A typical measured value of inductance for carbon composition resistors is 20 nH, with leads adding to this by about 20 nH per inch. However, with short leads and except for very low values of resistance the capacitive reactance from the lead terminations and the capacitance between conducting particles dominates the high frequency performance and total impedance decreases at high frequencies faster than with film resistors.

Film resistors consist of evaporated films of thickness from 0.005 to 1 mm, or thicker films up to 100 mm deposited from a resistive ink, or, in the case of carbon film resistors, deposited from the pyrolitic decomposition of an organic gas. Sheet resistances for the different types vary from 10 to 10,000 ohms per square. The films are spiral-cut to trim the resistors to final value. The spiraling increases the total inductance; nevertheless, the high frequency performance is dominated by capacitance.

The ability to absorb energy from a transient pulse depends on the thermal mass of the resistive element and on the maximum temperature that can be tolerated before permanent damage occurs.

Film-type resistors, especially carbon film, have a much higher critical temperature than carbon composition resistors. However their much smaller thermal mass makes them inferior to carbon composition resistors for this application. In fact, ESD pulses are known to cause permanent damage to film resistors. The spiraling causes a non-uniform power dissipation in the film that can lead to thermal damage at the ends of the spiral cut, and high voltages can cause an arc across the spiral cuts. Both types of damage have been observed.

Wirewound resistors are made by winding a resistance wire on a substrate or bobbin. Although the thermal mass is large, and, in fact, pulse-handling capabilities compare to those of carbon composition resistors in some cases, the very large inductive property limits their suitability for fast transient suppression.

There are two main classes of capacitors, electrolytic and electrostatic. Electrolytic capacitors include aluminum and tantalum types, characterized by large capacitance values, up to 1 F, but limited to voltages below 600 volts. They are made of high purity aluminum foil or tantalum, anodized to form the dielectric layer. This layer has unidirectional properties similar to those of a diode. Unless both electrodes are anodized the capacitor is unipolar and the instantaneous voltage must always remain of one polarity. At voltages roughly 50% higher than the rated voltage additional electrode material is anodized and a substantial current can flow.

One of the principal applications of electrolytic capacitors is in power supply filtering, where they also perform the useful function of suppressing surges that are coupled through from power lines. They are also used on printed circuit boards to decouple circuits connected to power busses, or to shield sensitive ICs from noise generated on the board.

Electrolytic capacitors are limited by their poor frequency response. They have a large inductive component that limits the self resonant frequency to 10 kHz or so. For this reason electrolytics are often paralleled with a 0.1 or 0.01 mF electrostatic capacitor that acts as a low impedance to high- frequency signals.

The main electrostatic capacitor types include plastic, ceramic disk, ceramic multilayer chip capacitors, and glass and mica capacitors. Plastic capacitors are made by evaporating a thin layer of aluminum onto a thin plastic film of polyester, polystyrene, polycarbonate, or other plastic. They exhibit the interesting property of self-healing, whereby voltage breakdown at a site is cleared by the evaporation of the aluminum around that site, restoring operation to an equal or higher voltage capability.

Ceramic capacitors make use of the high dielectric constant of ferroelectric materials to achieve large capacitance values in a small package. For disk capacitors the appropriate combination of powders is mixed and pressed into the desired form. These are sintered at high temperature, then electrodes are screened on, and the device encapsulated. By forming a very thick disk, high voltage ratings can be achieved. Multilayer chip capacitors are made from a slurry containing the mix of dielectric powders, and cast onto a stainless steel or plastic belt. After drying, electrodes are screen printed, the layers are stacked, then cut apart into individual units. The capacitors are

7.4.1.7 Protection Design For Parts, Assemblies and Equipment

The best way to build robust equipment is to use parts that are themselves robust. This is the most efficient, cost-effective, and practical approach. It protects integrated circuits during fabrication, testing, shipping, assembly, and equipment operation and repair. It requires no extra components or board space and imposes no degradation in frequency response. Nonetheless, sensitive parts do exist and at times must be used; in any event no matter how well a system performs it can always be improved by good design.

If an item of equipment were battery operated, shielded from electrical and magnetic fields and radiation, and had no input or output ports it might be impervious to anything except perhaps a direct lightning stroke. Realistically, equipment has a direct connection to power lines, signal ports, output ports, switches and other controls, apertures for cooling air, etc. The equipment will encounter transients at the inputs, radiated fields, both direct and indirect ESD discharges, handling by inexperienced personnel and repair persons, etc.

TABLE 7.4-1: COMPARISON OF PROTECTION DEVICES

	ADVANTAGES	DISADVANTAGES	RANGE OF MAXIMUM VALUES*	COMMENTS
MOV	low cost high energy capability low leakage current radiation hard	high capacitance aging effect	E = 0.06J - 10kJ I_{peak} = 25A - 100kA V_{DC} = 3.5V = 6kV C = 11pF - 11nF	especially useful for power line transients, board protection
Diode	low series resistance low capacitance	higher cost	P = 400 - 6500W V = 5 - 270V I_{peak} = 2 - 600A C = 3 - 500pF	fail short or open
SCR	nearly ideal characteristics low leakage current	higher cost limited availability turn-off requirements need to be addressed	V = 30 - 270V I_{peak} = 2 - 600A C = 3 - 90pF	fail short
R, L, C	low cost readily available	linear devices often not transparent to signals	V = 100V - 10kV p = 0.1 - 1mW	often used in conjunction with other transient suppressor

*Values in this column apply to specific transient waveforms or other special conditions. Consult specification sheets for details.

There are many anecdotes about computer malfunctions on cold winter days when the humidity is low and furnishings, and thoroughly train all staff members in proper procedures, the best approach is to make the equipment fool-proof.

To protect against direct-discharge ESD, it is necessary either to insulate the equipment or to provide a safe alternative path for the discharge current. If the discharge is indirect, then the equipment must be properly shielded to prevent magnetic or electrostatic coupling to interior circuitry. Protection must also be provided for maintenance, upgrading, or repair operations. This requires on-board protection and good layout.

There are four categories of techniques for system hardening.

(1) Board layout

(2) Shielding and grounding

(3) Use of transient protective devices

(4) Use of passive components and filters

Because of the wide overlap between these categories it is not possible to treat them as entirely separate areas; however all of them will be covered in what follows.

7.4.1.8 Printed Wiring Board Layout

Arrangement of parts on a printed wiring board should give priority to sensitive ICs. These should be placed near the center of the board where they are less likely to be contacted during handling. To further protect the components a guard ring should be in place around the periphery. The guard ring should be connected to the pull-out lever or board latch and to frame ground, but not directly to circuit ground. To prevent arc-over after the board has been installed the guard ring should be connected to circuit ground through a high resistance to bleed off static charges that might accumulate. To avoid electromagnetic interference (EMI), noisy circuitry should be segregated from sensitive circuits, and analog circuits segregated from digital. Edge-triggered logic should be avoided where possible. In extreme cases Faraday shields might be needed. To avoid coupling between an ESD discharge path (such as the guard ring) and sensitive inputs, parallel traces should be avoided. It is best to remember that any circuit that is a good radiator is also a good receiver for EMI. Whenever a trace becomes a significant fraction of a wavelength it has the potential of becoming a good antenna. Since light travels a foot per ns in free space (slower in materials), fast risetime transients can be efficiently radiated from short lines; therefore leads should be kept short, or shielded where necessary.

Inductive coupling is minimized by small loop areas. For example, power and ground traces should be close together. This can be accomplished by running multiple power and ground lines, on different layers, and transposing them at intervals. It is preferable to run each set of these multiple feeders back to a common point near the connector rather than form one long, continuous run. To maximize the capacitance between power and ground, a ground plane, and ground fill, should be utilized - no area should be left unused. To minimize coupling between signal lines it is preferable to alternate them with ground or power buses, but this may not be feasible from a real estate and frequency performance point of view.

Power supply filtering and decoupling is such an important concern that many boards use several capacitors to reduce noise and glitches on the power bus caused by digital circuits. Clearly even a low magnitude transient is capable of causing upset. Even though the power supply itself has a very low output impedance, cabling within the cabinet and the impedance of interconnects provide opportunities for EMI. One recommendation is that each board use an electrolytic capacitor (> 50μF), a 0.01μF capacitor for high frequency suppression, and a ferrite bead on the power lead, all as close as possible to the connector. Also recommended is a 0.01 μF capacitor for each IC. Such filtering will also help to attenuate surges on the power lines that couple through the power supply. At high frequencies the power supply busses are a significant fraction of a wavelength, and the characteristic impedance of the transmission line formed by the power supply trace and ground can contribute significantly to the noise. One way to lower the characteristic impedance is to use a power supply trace that is separated from the ground plane by a thin insulator. In lieu of this the inductance of the supply trace should be minimized by making the buss as wide as possible, and by providing multiple paths for the supply current.

7.4.1.9 Shielding

Equipment enclosures rely on the reflection of incident electromagnetic waves or their absorption by I^2R losses in the material to prevent the transmission of electromagnetic energy, either into the equipment, or from the equipment. In the first case we wish to protect the interior circuits from radiation caused by indirect ESD, lightning, or EMI. In the second case we wish to prevent emission from the equipment itself in order to avoid adding to the EMI background, and to comply with regulatory requirements.

An electromagnetic wave incident on a conducting surface generates currents in the material that produce electromagnetic fields opposing the incident wave. The stimulated wave is observed as a reflection of the incident wave, and the stimulated current produces losses in the body of the material that represent an attenuation of the wave as it progresses through the conductor. If the illuminated body is a perfect conductor the wave is totally reflected; there is no penetration of the shield, which acts like a Faraday cage. When the conductor is less than ideal only a portion of the wave is reflected, and non-uniform conduction currents flow in a layer near the surface. This is the so-called skin effect. The skin depth is the distance in which the induced currents or fields decrease by a factor of 1/e, to 37% of their original amplitude.

The shielding effectiveness due to absorption depends on the thickness of the material, and the decibel loss in any given material increases with the square root of frequency. (Ref. [101], RADC-TR-78-156, “Electromagnetic Properties and Effects of Advanced Composite Materials: Measurement and Modeling” and Ref. [102], RADC-TR-81-162, “Electromagnetic Shielding Effectiveness for Isotropic and Anisotropic Materials.”)

The other component of shielding refers to the reflection of the electromagnetic wave at front and back surfaces of the shield. A plane wave or an electric field is reflected very effectively from the surface of a good conductor. A magnetic field tends to be reflected from the back surface of a shield as it re-enters the atmosphere. The magnetic field that is reflected is virtually 100% of the incident wave, *but is reflected in phase,* and adds to the incident wave, again producing a standing wave.

Shielding effectiveness must be re-examined when the incident wave is not a plane wave (far-field condition). When the source is close to the shield (near-field region) the wave impedance is not 377 ohms and important differences exist.

The transition between near-field and far-field regions occurs at a distance d = $\lambda/2\pi$. At greater distances the fields are plane waves, the wave impedance is 377 ohms, and both the E field and H field decrease with distance as 1/r.

In the near-field region the reflection conditions are different because the wave impedance is different. For electric field sources the reflection losses are even greater than those with a plane wave. Consequently, shielding is not a problem. Even a thin, evaporated coating on a plastic layer is effective, although making a good electrical contact to ground the shield is problematic. (Grounding of any shield or conducting surface within the enclosure is necessary to prevent secondary arcs within the equipment.) When the source generates a primarily magnetic field, shielding is more of a challenge. Reflection losses are *smaller* than those of a plane wave, and *decrease* as the frequency decreases. Since absorption losses are small at low frequencies, it becomes a challenge to design shielding against low-frequency, near-field, magnetic sources.

The common methods of shielding against low-frequency magnetic fields are the use of ferromagnetic shields, such as “mumetal“ and the use of the “shorted turn”. Some ferromagnetic materials have very high permeabilities below 1 kHz and are particularly effective in confining magnetic fields. The “shorted turn” method uses a closed, conducting loop perpendicular to the magnetic field to generate an opposing magnetic field within its area. It is useful in subduing emissions from motors, transformers, etc.

Indirect ESD sparks and other arcing sources are usually high impedance, high E field sources. Magnetic sources are those that involve large currents as in power lines, ground return wires, conductors carrying a discharge current, transformers, etc.

Once appropriate shielding material has been selected, any apertures must be given proper attention. These are required for input and output signals, power, switches and controls, ventilation, and access. The general rule is that the largest dimension of an aperture must be a small fraction of a wavelength at the highest harmonics present. Some experts recommend $\lambda/10$, others as small as $\lambda/50$. With digital clock frequencies at 100 MHz and harmonics approaching 1 GHz the appropriate size of apertures to limit emissions would be of the order of 3 cm. To shield from ESD arcs of 10 ns duration would likewise require on aperture of 3 cm. With microwave equipment even these dimensions are too large. Note that any apenture with a wire through it acts like a section of coaxial cable with good transmission.

The 3 cm limit refers to the largest dimension of the aperture. If an opening is long and narrow it acts as a slot antenna, receiving or emitting frequencies corresponding to its length. When a larger opening must be used it can be subdivided into smaller openings or covered with a wire mesh. Access panels and doors must have closure seams protected by gaskets or interrupted by screws at least every 3 cm.

All conductive surfaces within the equipment should be grounded. Otherwise a secondary arc between the surface and another part of the cabinet could occur, or, worse yet, to a sensitive part or a trace on the printed wiring board. An arc occurring within the cabinet is itself a source of close range EMI confined to the cabinet enclosure. Direct ESD injection to internal circuitry must be prevented. No parts of the internal circuitry should be accessible to hands or fingers, and any direct discharge must be confined to the cabinet or shielding only. There should be no accessibility through apertures; switches and other controls should have grounded cases or should be insulated and sufficiently separated from circuits to preclude the possibility of arcing.

7.4.1.10 Grounding

Grounding refers literally to the electrical connection between equipment and a conducting rod driven into the earth. Figure 7.4-3 shows how grounding is accomplished at the service entrance to a building.

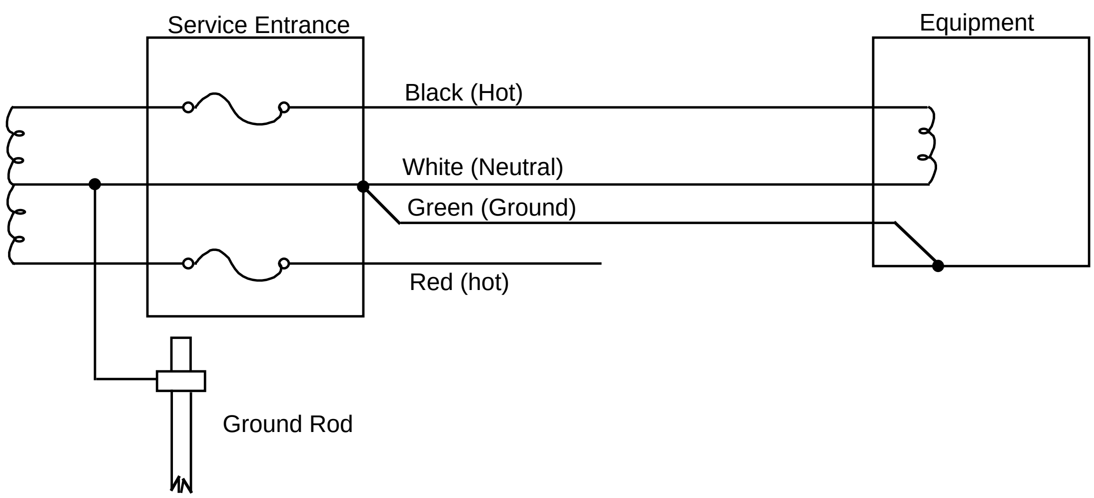


FIGURE 7.4-3: GROUNDING PRACTICE AT A SINGLE PHASE SERVICE ENTRANCE

If the “ground” for every circuit in every piece of equipment were at the same potential as the building earth ground, as intended, the circuit and systems designers’ jobs would be much easier. The reason this is not the case is because all ground conductors have impedance associated with them that renders different parts of the ground system at different potentials whenever ground currents are present.

An extreme case is associated with a nearby lightning stroke. Even if the stroke is conducted harmlessly to "ground", the flow of the extremely large currents through the finite resistance of the earth would cause one building ground to be at a different potential from another. If two items of equipment were located in two different buildings, connected by a shielded cable "grounded" at each end, a large current would flow through the shield and through the equipment.

One can immediately sense that grounding may be as much an art as a science; nevertheless there are important general principles that are effective in minimizing grounding problems. The overriding concern is to prevent large ground currents from flowing through impedances (especially inductive impedances) that raise parts of the system ground to higher voltages, the so-called "ground bounce" problem.

In Figure 7.4-4 several subsystems have their ground returns "daisy-chained". This invites problems. The noisy currents from the block of digital circuits flow through Z_{G2} and Z_{G1} and together with the large and erratic currents from the power electronics block that flow through Z_{G1} raise the potential of the ground bus of the low level and sensitive analog block, raising the noise level in that system. It is far better to keep the grounds from each block separate, and connect them all at a common point. Figure 7.4-5 shows separate grounds returned to a common point.

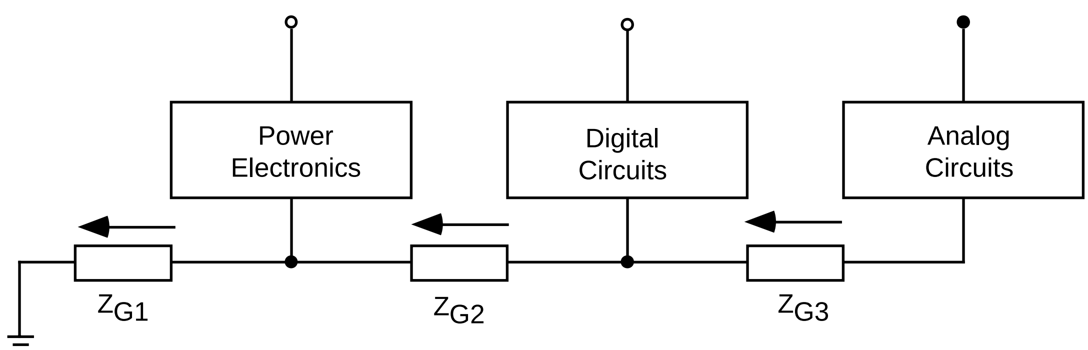


FIGURE 7.4-4: CIRCUIT SUBSYSTEMS WITH GROUND CONNECTIONS "DAISY-CHAINED" INVITES PROBLEMS

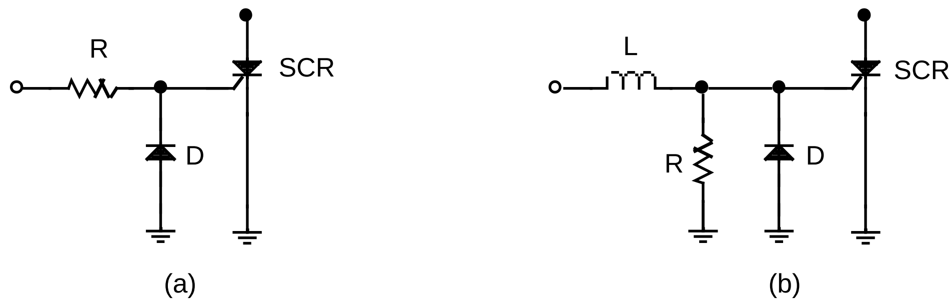


FIGURE 7.4-8: DIODE PROTECTION FOR SILICON CONTROLLED RECTIFIERS

Figure 7.4-9 illustrates diode protection for a TTL circuit. Diode D_1 clamps the negative transient to ground (using the output impedance of the driving circuit), and D_2 prevents the input from going more positive than V_{CC}. D_3 clips any negative surges on the output bus and the capacitor filters the V_{CC} bus.

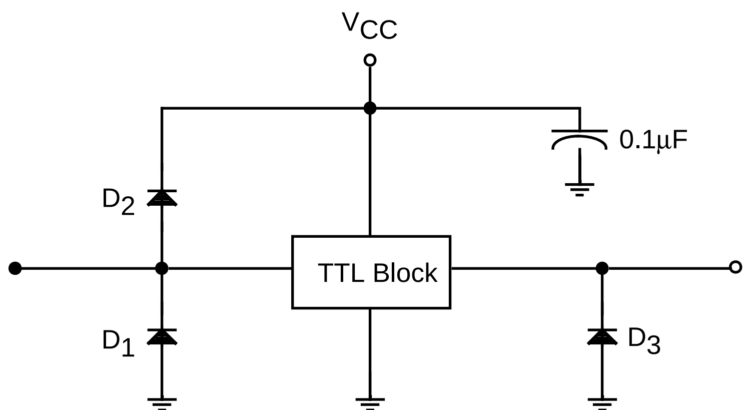


FIGURE 7.4-9: TRANSIENT PROTECTION FOR A TTL CIRCUIT USING DIODES

In Figure 7.4-10 a simple scheme for protecting CMOS circuits is shown. In part (a) the diode clamps positive pulses to V_{DD} and limits negative voltages to V_{DD} minus the diode reverse breakdown voltage. In part (b) the protection circuit is more elaborate, patterned after a commonly used on-chip protection scheme. D_1 is selected to have a larger reverse breakdown voltage than D_2 or D_3. Resistor R_2 in conjunction with the on-chip protection network at the input of the CMOS circuit provides a third stage of protection. Each of the three stages would turn on sequentially as the input transient becomes larger and larger.

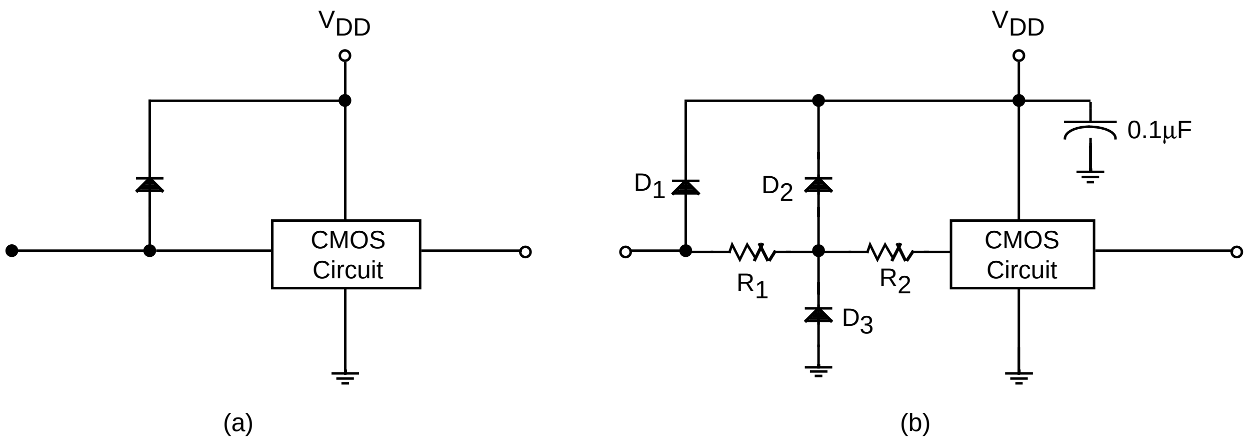


FIGURE 7.4-10: TRANSIENT PROTECTION FOR A CMOS CIRCUIT

Line transients are one of the causes of failures in switching mode power supplies. A major failure mode is shorting of the switching power transistors caused by power line surges. The inrush of current in conjunction with the equivalent series resistance and inductance of the filter capacitors produces an overstress of the power switches that leads to failure. The best remedy is to suppress the transients at the input to the power supply.

Figure 7.4-11 shows an effective method of suppressing line transients that uses a hybrid scheme employing both an MOV and clamping diodes. The MOV is a high energy absorbing device that provides the main protection. The clamping diodes provide a more precise limit to the voltage excursion. Suitable values of L and R are of the order of 100 µH and 1 ohm, respectively.

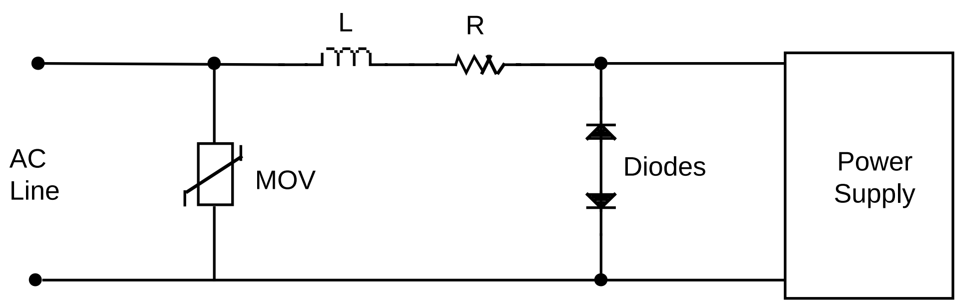


FIGURE 7.4-11: INPUT PROTECTION FOR POWER SUPPLIES

Diodes are compatible with on-board protection of components, and diode arrays can be effectively utilized to provide protection for data lines and power buses. Figure 7.4-12 shows an arrangement where the diode array is located adjacent to the edge connector. Some designers prefer to place the protection devices close to the components being protected; others try to avoid surge currents being propagated around the circuit board where they become sources of rf fields, and high voltage transients can cause arcs to nearby components or traces. For the latter case, the diodes are located at the connector as shown in the figure.

Transient voltage suppressor diodes usually fail short, but may also fail open. If they fail open there is no longer protection of susceptible components against subsequent transients. For the more common case of failing short, a significant surge current can occur unless limited by series impedances. This condition is much more serious for power buses than for data lines. The short-circuit current will eventually cause the protection diode to open-circuit. The relationship between the amplitude (squared) of the short-circuit current and the time to fail open is described fairly well by the classical Wunsch-Bell model; in other words the current is proportional to $t^{-1/4}$.

Figure 7.4-13 shows a fuse in a power bus protected by a diode. The fuse must be carefully selected to conduct the expected transient current, but to open when the diode fails short. These requirements are especially difficult to meet for low voltage diodes.

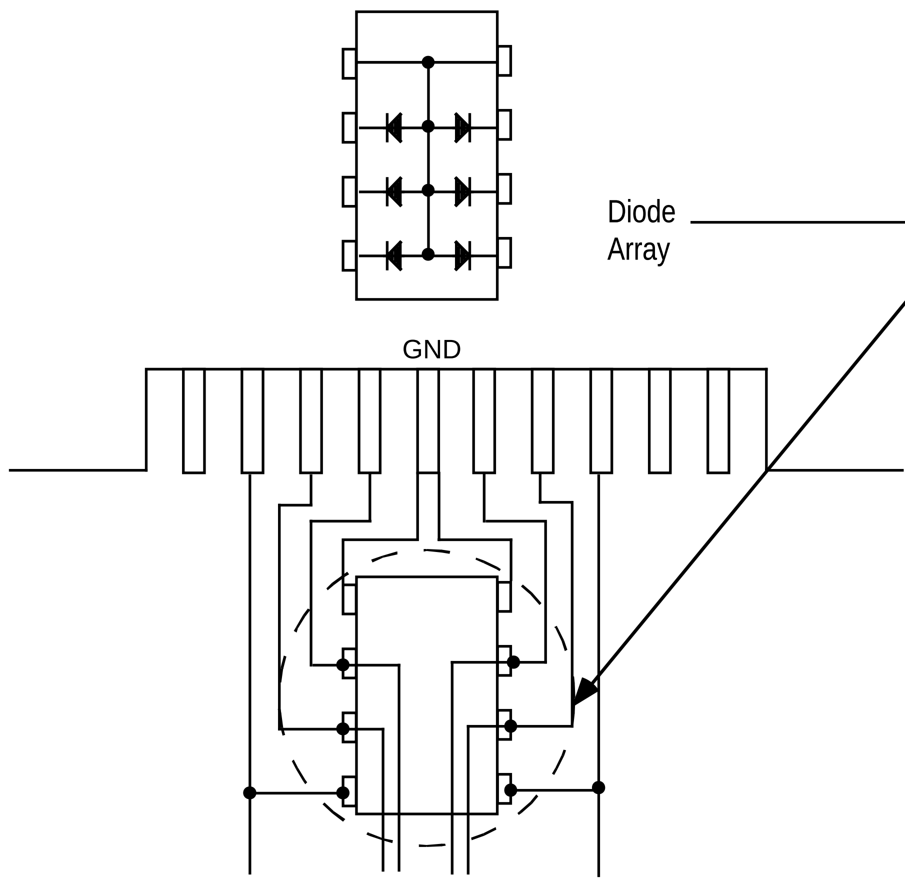


FIGURE 7.4-12: PROTECTION OF DATA LINES OR POWER BUSES
USING A DIODE ARRAY

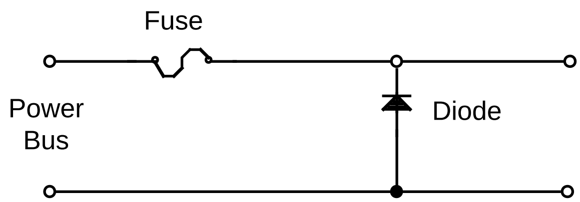


FIGURE 7.4-13: FUSE PROTECTION FOR A TRANSIENT VOLTAGE SUPPRESSOR DIODE

Since the threat from ESD comes mainly during handling of the boards, CRO-BAR® devices that connect all traces together at the edge connector pads would eliminate ESD pulses that are applied at the connector. CRO-BAR® devices automatically disengage when the board is inserted into a slot or when a cable is attached. CRO-BAR® devices on the cable itself also remove any static charges that may accumulate when the cable is left unconnected or unrolled from a spool.

Most of the techniques mentioned are common sense to those with some design experience in EMI, electromagnetic compatibility (EMC), ESD, or transient suppression. A few are still somewhat controversial. The importance of good design becomes increasingly relevant as electronic equipment and appliances proliferate and as devices become smaller and more sensitive. The battle is never finished, and as the electronics industry evolves, design engineers must realize that old remedies may no longer be applicable and new, innovative solutions must be thought of. Once a solution has been found, funding is often discontinued, those who make such decisions not realizing that like the mythical multiheaded dragon, new crises continually arise to frustrate those caught unawares.

7.4.2 Parameter Degradation and Circuit Tolerance Analysis

The values of part parameters, physical and electrical characteristics, are known to vary with time under the effects of aging and stress. Variations in part parameter values, if not considered in the design, can have undesirable effects on circuit performance and are a significant cause of system failure. Even when the variations in the value of a single parameter for a single part have no effect on system performance, the cumulative effect of such changes can degrade system performance to a point where it is no longer acceptable.

In addition to the variations caused by aging and stresses, the values of part parameters vary due to the manufacturing processes used in the manufacture of the parts. These variations can differ by manufacturing lot and can be affected by procedures in which parts are individually selected, using for example, “screens”. Whatever the causes, the variations in a given part parameter can be described by a statistical distribution. The expected value and standard deviation of this distribution represent the nominal or “average” value of the parameter and the variation around this nominal value, respectively. As already indicated, the nature of the distribution is a function of the manufacturing process, aging, and stress. Stress includes elements of the operating environment, such as temperature, vibration, pressure, and so forth. Examples of part parameter variations are shown in Figures 7.4-14, 7.4-15, and 7.4-16. Figure 7.4-14 shows the average variation from the initial value and the standard deviation of the variation plotted over time for the resistance of a particular type of resistor and Figure 7.4-15 shows the same information for the capacitance of a particular type of capacitor. Figure 7.4-16 shows how the variability in the nominal resistance increases under a specified stress and temperature condition for a period of time, for two different levels of power dissipation.

In designing a system, these variations in the values of part parameters must be specifically addressed to ensure that the design is robust. A robust design is one in which substantial variations in the values of part parameters have little or no effect on system performance. In designing for robustness, designers must have a knowledge of the variations expected due to manufacturing, aging, and stress and the expected ranges of those variations. With this knowledge to guide them, designers can work to eliminate or mitigate the effects of variations in parameter values.

Two approaches that can be used to eliminate or mitigate the effects of variations in parameter values are:

(1) Control the device and material parameter variations through process design and control to hold them within specified limits for a specified time under specified conditions. This will be referred to as Parts Control.

(2) Design circuits and systems to be sufficiently tolerant of variations in device and material parameters so that anticipated variations over time and stress do not degrade system performance. This will be referred to as Design Control.

The first approach requires that the parameter value be controlled. Burn-in, preconditioning, and other screening methods, can be used to eliminate or reduce variation in a specific parameter. This screening results in parts having more stable parameters. Controlling the parameters of a part requires detailed testing and control of materials used in the parts. It requires strict control of manufacturing processes, the use of proven designs, and parts testing to collect data on parameter variation over time.

The second approach is to design circuits that are tolerant or insensitive to variations in parts parameters. Three different techniques for designing tolerant circuits are: (1) use feedback to electrically compensate for variations and thereby provide stable performance, (2) ensure that the circuitry provides the minimum required performance even under expected or worst case conditions (i.e., maximum variation), and (3) design the circuit to be insensitive to variations.

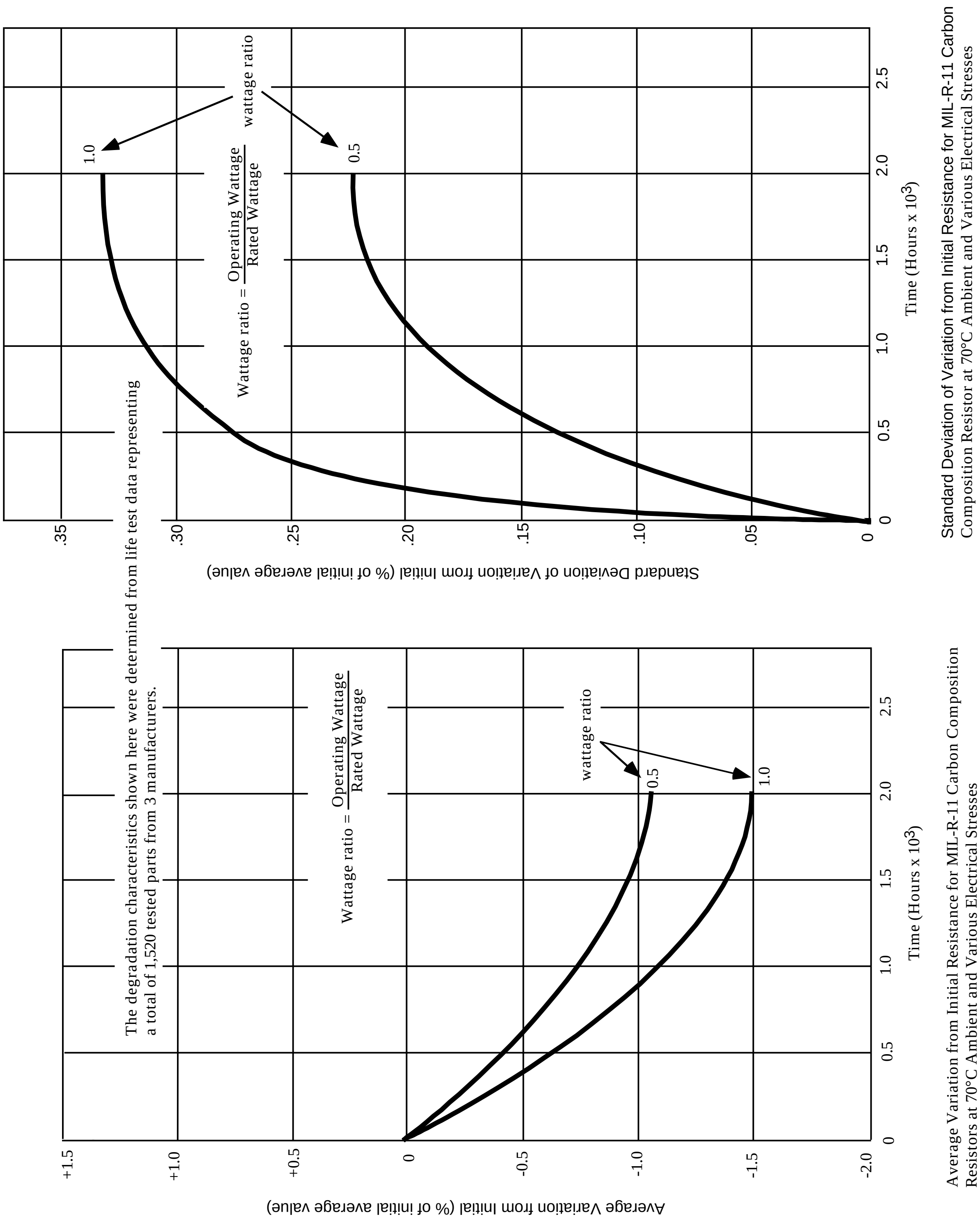


FIGURE 7.4-14: RESISTOR PARAMETER VARIATION WITH TIME (TYPICAL)

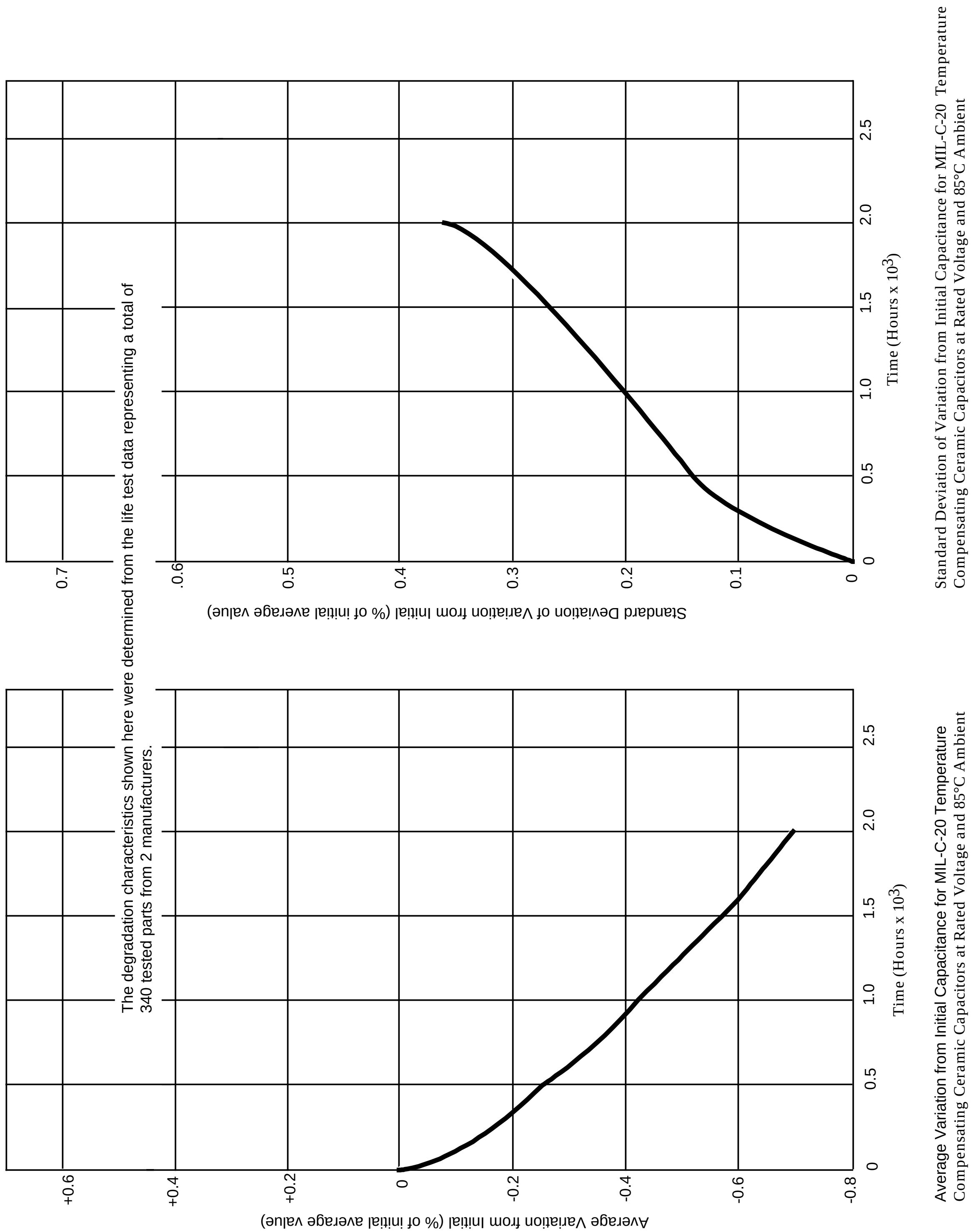
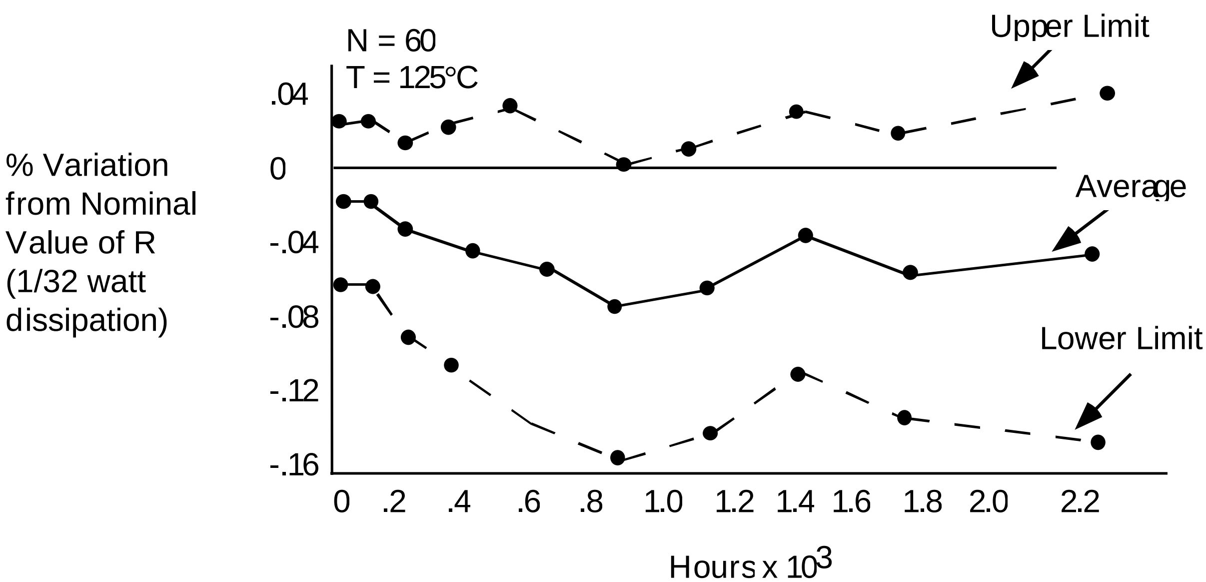
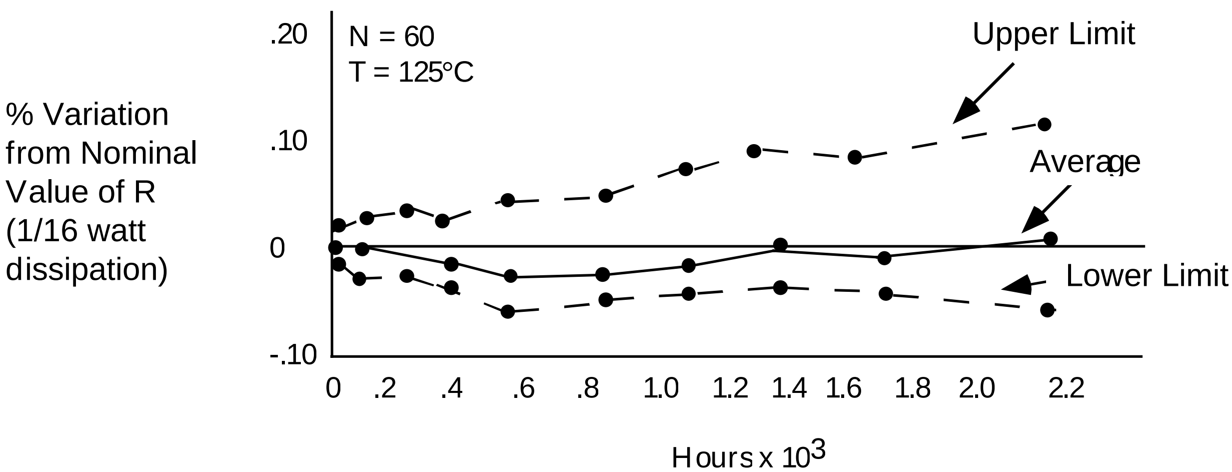


FIGURE 7.4-15: CAPACITOR PARAMETER VARIATION WITH TIME (TYPICAL)





Data shown for 60 resistors of fixed metal film type rated at 1/8 watt during 2000 hours of operation at 125 degrees C and two different levels of power dissipation (stress levels).

FIGURE 7.4-16: RESISTOR PARAMETER CHANGE WITH STRESS AND TIME (TYPICAL)

In designing tolerant or insensitive circuits, the procedures for analyzing variability include:

(1) Worst Case Circuit Analysis (WCCA)

(2) Parameter Variation

(3) Monte Carlo

(4) Design of Experiments (DOE)

(5) Taguchi robust design methodology

These methods are presented in Table 7.4-4 and are described in detail in References [18], [19], [20], and [21]. The ultimate objective of these analytical methods can be one of the following.

(1) To select parts based on a determination of the allowable limits of variation for each part parameter and the anticipated operational environment. (Parts Control.)

(2) To design the circuit to produce the minimum performance under worst case or statistically expected conditions of environment and parameter variation. (Design Control.)

(3) To determine the parameter(s) most critical to proper operation of the part and then to design the circuit in such a way that variations in that parameter(s) do not affect performance. (Design Control.)

The first objective is to match the part to the application. It is seldom possible to find an exact match, so the parts usually have less parameter variability than could be tolerated. The second objective is to design the circuit to operate properly under the worst possible conditions. In so doing, the cost of the design could offset the advantages gained or make the design unaffordable. The last objective is to design a circuit in which the variation in critical part parameters is overcome by the nature of the design. Consider the following example from Reference [21].

TABLE 7.4-4: COMPARISON OF VARIABILITY ANALYSIS METHODS

Analytical Method	Type of Analysis	Statistical?	Output	Objectives
WCCA	Mathematical	No	Worst-case values for inputs with all parameters at cumulative worst-case limits	Determine if failure is possible and, if so, under what conditions
Parameter Variation	Mathematical	No	Range of variability data for Schmoo plots	Establish realistic tolerance limits for parameters
Monte Carlo	Mathematical	Yes	Output histograms	Reliability estimates
DOE	Mathematical	No	Significant (critical) parameters and optimal values	Minimize number of experiments needed to establish relationship between parameters and performance
Robust design	Mathematical	Yes	Component values	Less variability (better quality)

A circuit is required to provide a specified output voltage, V_o, that is primarily determined by the gain of a transistor and the value of a resistor. As shown by the graphs of output voltage versus transistor gain for resistance values R_1 and R_2 in Figure 7.4-17, the transistor is a non-linear device. Assume the prototype circuit achieves the required voltage, indicated by the diamond, with resistance R_1 and transistor gain G_1. The inherent variability in the gain of transistor 1 is depicted by the bell-shaped curve centered on G_1. The amount of variability in R_1 causes a large variation in V_o as shown by the bell-shaped curve marked *a*.

Trying to reduce the variation in V_o by reducing the variability of G_1 may be very difficult or expensive. An alternative to selecting a higher quality (i.e., less variability) transistor is to develop a more robust design. We can do this in the following manner. First, operate the transistor at a higher gain, G_2. Note that the variance of G_2 is now larger as indicated by the bell-shaped curve centered on G_2. However, the non-linear relationship between gain and V_o results in a smaller variation of V_o as indicated by curve *b*. V_o, however, is now too large. By choosing resistance R_2, we reduce the voltage to the proper level, as indicated by curve *c*. V_o is again equal to the target value but with a much smaller variability. Since transistor gain is somewhat affected by ambient temperature and V_o is now less sensitive to gain, an added benefit of this design is that V_o now is less sensitive to ambient temperature.

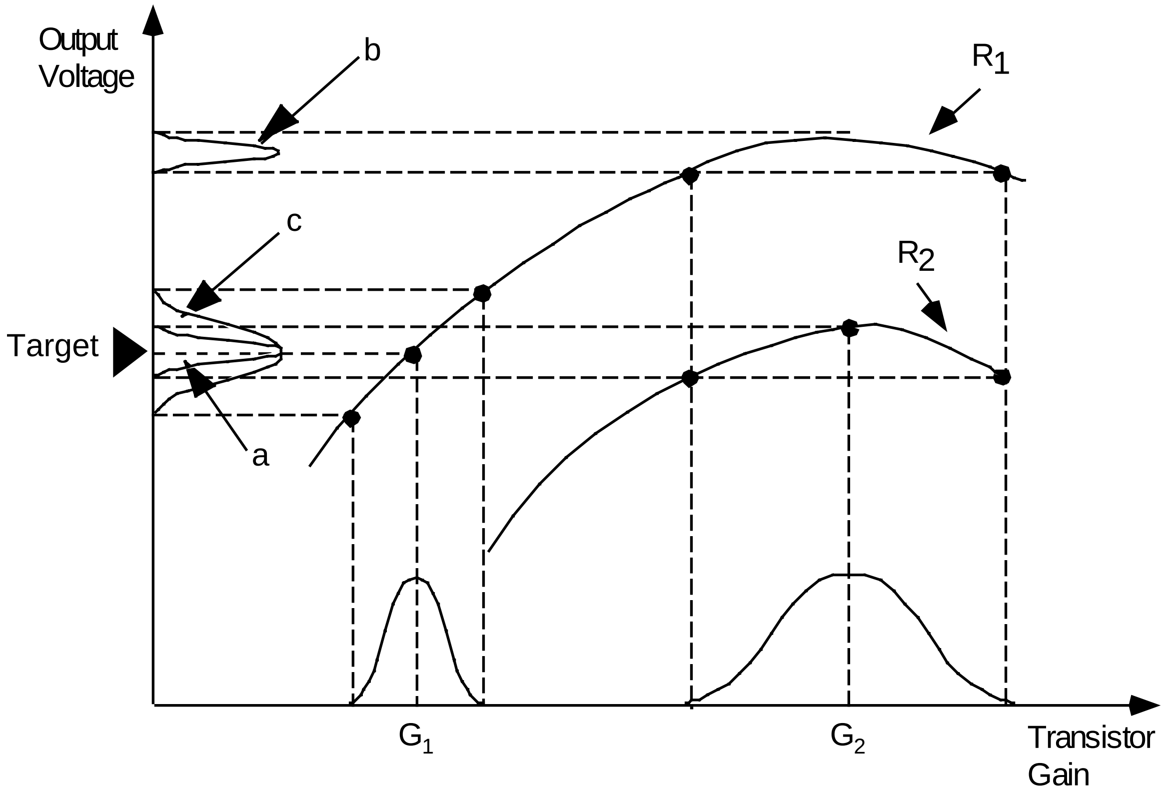


FIGURE 7.4-17: OUTPUT VOLTAGE VERSUS TRANSISTOR GAIN
BASED ON A FIGURE APPEARING IN TAGUCHI TECHNIQUES
FOR QUALITY ENGINEERING (REFERENCE [21])

7.4.3 Computer Aided Circuit Analysis

All of the following material in which computer analysis software packages are identified is presented for information only. The DoD does not endorse their specific use or attest to their quality.

Circuit Simulation: A Brief History

In the early 1960's military requirements led to the development of mathematical simulation of components (capacitors, semiconductors, etc.) to determine their response to pulsed x-ray and gamma radiation. These simulation studies were subsequently extended to small circuits to study their response to the same radiation conditions. This work resulted in the early circuit analysis programs (ECAP, SCEPTRE, CIRCUS, etc.).

Later program capabilities included AC, DC and transient performance simulation - with and without radiation effects. RF and microwave circuit simulation capabilities, sensitivity analysis, Monte Carlo, worst-case analysis and optimization analysis capabilities were also eventually added.

Early simulations were run overnight, in batch mode, on large mainframe computers; it was cumbersome to input the data and the graphics outputs were poor.

These simulation programs quickly migrated to engineering workstations and their capabilities were significantly enhanced by such features as simulation integration and schematic capture. They became more user friendly, included high resolution graphics and gave quick turn-around. These circuit analysis and simulation tools eventually became available for the ubiquitous PCs.

Hardware design and analysis is typically performed on workstations today. At the same time, however, the capabilities of PCs continue to improve. Thus, even the distinctions between PCs and workstations continues to blur with improvements in the state-of-the-art.

The current trends in design and in analysis software are toward portability and standardization, especially the increased use of higher level languages. The trend is to a fully integrated design-analysis environment including:

(1) Schematic Capture
(2) Circuit Simulation
(3) Manufacturing Considerations
(4) Test Vector Generation
(5) Configuration Control

At present, analog circuit analysis and digital circuit analysis usually require different software packages. However, efforts are underway to unify analog and digital simulation software. Several commercial packages are available with mixed analog/ digital simulation capability.

7.4.3.1 Advantages of Computer Aided Circuit Analysis/Simulation

Computer-aided analysis is the standard for large (multi-node), sophisticated circuits, those requiring special analysis techniques (transient, Monte Carlo, etc.) and iterative type analyses (e.g., frequency response). It is very cost effective for programs where the schedule imposes significant time restraints and where an insufficient number of skilled circuit analysts are available for the task. Computer-aided analysis is the only way to handle highly complex circuits accurately.

Computer simulation of circuit performance is a universally accepted technique. It's features are relatively easy to learn, including the ability to adjust (i.e., "tweak") parameter values for re-analysis (temperature changes, BOL (Beginning-of-Life) vs. EOL (End-of-Life) values, etc.). Furthermore, it can provide automatic documentation of the analytical results including: topology listings; calculations of voltage, current, and power; and plots of the variables.

7.4.3.2 Limitations of Computer-Aided Circuit Analysis/Simulation Programs

In general, a single computer program does not provide performance simulation of all classes or types of circuits, i.e., RF and microwave, analog (AC, DC, and transient analysis) and digital (logic and timing analyses). Also, because of the variety of computer platforms, computer programs are typically prepared for use by only one (or a few) computer families.

The accuracy of the circuit simulation results is no better than the accuracy of the model representing the circuit. This, of course, is also true for manual analysis. For each circuit type, special data are required. The accuracy of these data can also affect simulation results. In cases of extreme complexity, even the task of generating the circuit models needed for the computer analysis may be difficult to deal with.

7.4.3.3 The Personal Computer (PC) as a Circuit Analysis Tool

The PC has emerged as a powerful, economical engineering tool with a wealth of application software, particularly for MS-DOS/Windows-based systems. PC software packages typically used in circuit analysis include:

(1) Spreadsheets
(2) Data Base Management Programs
(3) Mathematics packages
(4) Circuit analysis and simulation programs

Spreadsheets

Spreadsheet programs are useful for performing stress analysis and simple calculations. Various different spreadsheet are available. User interfaces for most spreadsheets are similar, thus making it is relatively easy to switch between different spreadsheet packages.

Data Base Management Programs

Data base management programs are very helpful for dealing with large parts data bases. These software packages usually include a built-in command language to facilitate data manipulation and report generation.

Mathematical Software Packages

A variety of general purpose mathematical software packages are currently available. Typical package capabilities include:

(1) Simultaneous equations
(2) Complex variables
(3) Derivatives and integrals
(4) Iterations
(5) Trigonometric and exponential functions
(6) Equations entered in their normal form can generate output plots
(7) Statistical functions
(8) Cubic spline curve fitting
(9) Fast Fourier transforms and inverse vectors and matrices
(10) User-definable functions,
(11) 3-dimensional plotting.

Additional features might include:

(1) A scientific word processor, complete with mathematical function symbols.

(2) Ability to solve both linear/non-linear simultaneous equations with constraints and conditions.

(3) A "root" function which can solve for the zeros of linear and non-linear functions and combinations thereof.

(4) Support for complex arithmetic and matrix data.

(5) Ability to read and write scalar and matrix/vector data to standard ASCII files; or formatted files for matrix or vector data.

(6) Matrix operators: addition, subtraction, multiplication, inversion and powers, (where negative powers of the matrix are powers of the inverse), determinant, transpose, and complex conjugate along with numerous matrix functions.

(7) Vector operators including: scalar multiplication, dot product, scalar division, addition, subtraction, scalar subtraction, magnitude, and complex conjugate along with numerous other vector functions.

(8) Support for various mathematical functions including; trigonometric, hyperbolic, log, exponential, Bessel, complex variables, interpolation, statistical, linear regression and Fourier transform.

Circuit Analysis and Simulation Programs

Analog Circuit Simulation

The different analog simulation programs available typically perform similar circuit analysis functions and program enhancements are implemented regularly. Typical features include :

(1) DC (bias point) and AC (frequency response) steady state analysis
(2) AC and DC Transient (time response) analysis
(3) Noise Analysis
(4) AC and Transient analyses at fixed temperatures
(5) FOURIER Analysis
(6) Worst-Case Analysis
(7) MONTE CARLO Analysis
(8) Component sweeps
(9) Initial condition documentation
(10) MACRO Models
(11) Continuous and piece-wise nonlinearities
(12) Graphic plot or tabular output

A MONTE CARLO analysis option allows multiple repetitive runs to be performed with a random selection of part values (within the tolerance band) for each run. This option can usually be applied to all types of circuit analyses. Data reduction capability then allows deviations from the nominal to be determined and tabulated to determine the probability of proper circuit performance.

Schematic capture interfaces may be included or they are available for most popular packages. Using one of these packages then allows you to go directly from the schematic to the circuit simulation.

Device Models and Parts Data

Device model libraries are usually supplied with the simulation program. Typically they might include: Diodes and Rectifiers, Bipolar Transistors, Power MOSFETs (with Enhanced Curtice and Raytheon models), GaAs MESFETS, Operational Amplifiers, Transformers and Power Inductors (with Jiles-Atherton nonlinear ferromagnetic equations), Voltage Comparators, Switches and miscellaneous parts such as; Voltage-controlled - capacitance, - inductance, - resistance, - conductance. More extensive integrated circuit libraries are also frequently available when needed from the part manufacturers themselves.

Semi-automated processes for creating model libraries may also be included. The parameters are typically estimated from the manufacturers' data sheet parameters. The process is interactive, prompts are provided for the input data, device curves can be presented for verification and the results can be saved in a library file.

Once they are developed, the circuit model libraries provide a repeatable, accurate, basis for analysis. Upon completion of the analysis, the results can easily be integrated into a final report.

Digital Circuit Simulation

Typical program features include:

(1) Schematic capture interface
(2) Extensive model libraries for specific ICs
(3) Logic simulation
(4) Multi-state simulator
(5) Timing analysis
(6) Nominal and worst-case timing
(7) Race, spike, hazard and pulse width analyses
(8) Fault simulation
(9) Grading of test vectors
(10) ATE tester interfaces are available
(11) Ethernet link to workstations and/or mainframes

Digital circuit simulators are very convenient for performing critical timing analyses. Graphic display of the circuit nodes simplifies the analysis of timing relationships. Advanced program features, such as load-dependent delays improve the accuracy of the analysis and eliminate overly conservative delay estimates.

7.4.4 Fundamental Design Limitations

Probably the first and prime step in the establishment of reliability criteria is the establishment of the boundaries which represent the limitations on the controlled characteristics for the component or device in question. Some of the limitations are commonly known: breakdown voltage, power dissipation limitation, current density limitations, and similar factors. Many, however, are either poorly known, or possibly not known at all. Often it is these factor which cause difficulties in circuits.

If one examines the behavior of components in systems, one finds that there normally is a region of operation in which failures are rare or unlikely, but when operating conditions reach a possibly undefinable level, the probability of failure rises substantially. Conversely, with any given configuration, improvements in reliability as a result of redesign may be easy to obtain to a certain level of improvement, and then become progressively more difficult to obtain.

Improvement of reliability in terms of these criteria generally makes more sense than either attempting to attain an excessively high value for all components or being satisfied with an excessively small value based on the poor reliability of the few components. Limiting the collector supply voltage to the minimum provides a very economical way of improving the reliability of a given circuit.

The optimization of the reliability of a system on a circuit-by-circuit basis might appear to be an excessively time consuming and difficult problem. Actually, however, such need not be the case, since it is entirely practical to test at the design state (on paper) the effects of voltage reduction on circuit performance. Since it is necessary to limit voltage gain for reasons of circuit stability, proceeding in this manner might lead to an occasional additional amplifier circuit but it should at the same time lead to substantially reduced power consumption and substantially reduced cooling problems. Both of these are important criteria for reliability.

The following paragraphs discuss some fundamental design limitations which are important to designers of military electronic equipment.

7.4.4.1 The Voltage Gain Limitation

The development of radar brought with it the need to be able to amplify very weak signals in the presence of strong ones, and for the first time made the question of stability and freedom from ringing a prime consideration in tuned amplifiers. These tuned amplifiers frequently were required to have voltage amplifications as great as a million overall, with no change in operating frequency permitted.

The basic criterion which must be satisfied, both for each individual amplifier stage and for the amplifier as a whole, is that the loop amplification of individual elements as well as of the assembled groups of elements must be rigidly limited to assure that stability will not be impaired. This stability problem is essentially a phase-sum problem. If an input voltage is applied to the amplifier or stage in question, then the voltage returned through feedback to be summed into the input voltage is the product of this voltage by the amplification "around the loop" from input back to input

$$K_L = K_\upsilon \cdot K_f \qquad (7.3)$$

where K_υ is the forward voltage amplification to the output, and K_f is the feedback "amplification" from the output back to the input on an open-loop basis.

The modified forward amplification, K'_υ then takes the form:

$$K'_\upsilon = K_\upsilon / (1 - K_\upsilon K_f) \tag{7.4}$$

and the phasor term $(1 - K_\upsilon K_f)$ determines both the variation of the signal amplitude and the signal phase.

Clearly, one of the requirements of any amplifier to which Eq. (7.3) applies is that $|K_\upsilon K_F|$ must be small compared to unity, or a potentially unstable situation can develop. In addition, significant phase shift in the output circuit compared to the input can occur even with relatively small values of $|K_\upsilon K_F|$ values as small as 0.1 or 0.2, for example. In such a situation, as much as 5 to 10 degree phase discrepancy per stage can be encountered.

Where phase stability is of prime importance, it is evident that values of $|K_\upsilon K_F|$ should be less than 0.01 if at all possible, as then there is reasonable chance that the cumulative phase angle discrepancy in a system may be limited to a fraction of a radian. The design of an amplifier meeting this limitation can be both difficult and painstaking, and the mechanical realization of the calculated design can be even more difficult. The design techniques described in Reference [35] offer possibly one of the best ways of achieving the required results.

Early radar experience quickly showed that the limit on per stage gain for K_υ achieving amplitude and phase stability with minimum to modest ringing proved to be approximately 10. (It is possible to get device gains of 100 with common grid or common base circuits, but the required impedance transformation required to match the input circuit for the succeeding amplifier typically reduces the overall stage gain back to approximately 10.) This means that the maximum permitted value for K_f is approximately 0.01 to 0.02, for a power isolation possibly as much as 40 dB. Where phase stability is of primary importance, the maximum permitted value for K_f is nearer 0.001 than 0.01.

It is very important to control and restrain the circulation of carrier frequency currents throughout any multistage amplifier, since if five stages overall are involved, the isolation from output back to input must be about 0.01^5 or 10^{-10}. This is the reason that radar IF amplifiers were designed to receive power in the vicinity of the middle stage, and R-C decoupling was used in both directions for supply voltages, and L-C decoupling for heater currents. All voltage feed points were in addition individually bypassed, and grounds grouped within the channel in such a way as to prevent circulation of carrier frequency currents in the channel.

Clearly, there is really nothing magic about the value of K_υ of 10. The magic number, if one exists, is in fact the "invariant" $K_\upsilon \bullet K_f$ whose value must be sufficiently small to limit the phase and amplitude excursions in the signal. This is the basic stability criterion. But there definitely is an upper limit on the value of K_υ, at least in a practical way, since there is a lower practical limit on how small K_f can be made successfully in production type equipment. The internal stage voltage gain from input to output on control separation amplifiers can be significantly higher, since the input admittances for these devices are sufficiently high that the return feedback gain is severely reduced.

This limitation on voltage gain has very interesting consequences, particularly in design for reliable operation. The voltage gain of a bipolar transistor is given by Eq. (7.5).

$$K_\upsilon = \kappa \Lambda I_C Z_L \tag{7.5}$$

where:

K_υ	=	forward voltage amplification
I_C	=	collector current
Z_L	=	load impedance
κ	=	efficiency factor $\cong$ 1
Λ	=	$q/kT = 40V^{-1}$ at 25°C
q	=	electron charge
k	=	Boltzmann's constant
T	=	absolute temperature

In this equation, it is evident that $I_C Z_L$ is the maximum signal voltage for Class A operation.

It is possible to relate the voltage $I_C Z_L$ to the minimum possible supply voltage V_{CC}, which can be used with the ideal device in question to produce the required operating characteristics. The minimum supply voltage may then be defined in terms of the equation

$$I_C Z_L = \kappa_\eta (V_{CC} - V_{SAT}) \tag{7.6}$$

where κ_η is a parameter which relates the output load voltage to the supply voltage and V_{SAT} is the maximum saturation voltage. κ_η usually has a value between 0.2 and 1.0. Substituting Eq. (7.6) in Eq. (7.5) gives the result:

$$K_\upsilon = \kappa \kappa_\eta \Lambda (V_{CC} - V_{SAT}) \tag{7.7}$$

This equation may be solved for the minimum supply voltage V_{CC} for a device in a circuit to give

$$|V_{CC}| = |K_{\upsilon}| (\kappa \kappa_{\eta} \Lambda)^{-1} + V_{SAT} \qquad (7.8)$$

In Eq. (7.8), the value of κ_{υ} is about 10, typical values of κ_{η} are less than unity, and V_{sat} is a few tenths of a volt. As a result, with $\kappa_{\eta} = .5$, for example, the minimum value of supply voltage required for a circuit can be expected to be roughly a twentieth of the voltage gain. This means that the range of required supply voltage is between 0.5 and 10V, the lower voltage limit applying to the common emitter configuration, and the higher to the common base configuration.

The significance of this relation cannot be overemphasized. The properties of the device and its associated circuitry are controlled largely by the current level selected for operation, and ***there is little point to selecting a supply voltage for the output circuit which is more than marginally greater than calculated by Eq. (7.8).*** Selection of a higher voltage leads either to excessive power dissipation, excessive gain with its inherent instability, or combinations of these conditions. In short, the selected supply voltage should be as small as possible consistent with the demands on the circuits.

This discussion should not be construed to mean that the base supply voltage provided for base bias current and voltage necessarily can be as small as that for the collector. Since crude stabilization of circuits is frequently obtained by controlling the base current in a transistor, the supply voltage provided for this function must be sufficiently large to assure that an adequate constancy of current level can be achieved. This and this alone is the justification for use of a large voltage, yet the current requirement for these circuits is sufficiently small that a substantial decrease in power dissipation and a substantial improvement in reliability could be achieved through the use of separate power sources for these two functions. In comparison, then, one source of high current and low voltage is required, and one of higher voltage but substantially smaller current also is required. Using a common source for both clearly leads to the worst failures of each! Also, use of two power sources allows a better matching of total power and current to the demand resulting in a smaller, lighter, and less expensive solution than with a single power supply.

7.4.4.2 Current Gain Limitation Considerations

The voltage gain limitation is electrostatic, or charge control, in nature. It is particularly important with transadmittance[2] devices, which tend to have a relatively high input impedance and tend to become regenerative by passing through a zero admittance (infinite impedance) condition.

---

[2] Transadmittance for a bipolar transistor is $y_f = \Lambda I_C$

The network dual of the voltage gain limitation is the current gain limitation. It is technically possible for this also to be critical, but at present its consequences are much less severe than its dual. Probably the principal reason for this is the rapidity of decay of magnetic fields associated with currents due to mutual cancellation of opposing components. Additional reasons are the dependence on rate-of-change of current (since only changing fields create voltage and currents), and the nonexistence of true transimpedance devices.

The control of magnetic fields proves to be one of control of fluctuating currents. The more that can be done to keep current fluctuations isolated and out of wires and shielding structures, the more freedom there is from coupling currents and fields. Size of loops carrying fluctuating currents should be kept to an absolute minimum unless the inductive properties of the loop are essential to the operation at hand. Even then the loop or coil should be so designed and so installed that it generates its field efficiently, so that an adequate quality factor, or Q, is obtained, and so that coupled fields and circulating currents induced and generated by the field are limited to regions where they are required and otherwise kept to a practical minimum.

7.4.4.3 Thermal Factors

One of the major problems in the use of transistor circuits is the stabilization of operating conditions so that the circuit can give the required performance over an adequate range of environmental conditions.

There are two principal thermal factors that affect the stability of transistor circuits. The first factor is the reverse leakage current of the collector base junction, the so-called I, and the second factor is the variation of V with temperature. The leakage current increases rapidly as the temperature of the transistor is increased. This effect limits the conditions under which the transistor can provide effective operation (Figure 7.4-18). This current, in conjunction with the current gain of the transistor, limits the minimum usable current through the common emitter amplifier, thereby restricting the available range of operation.

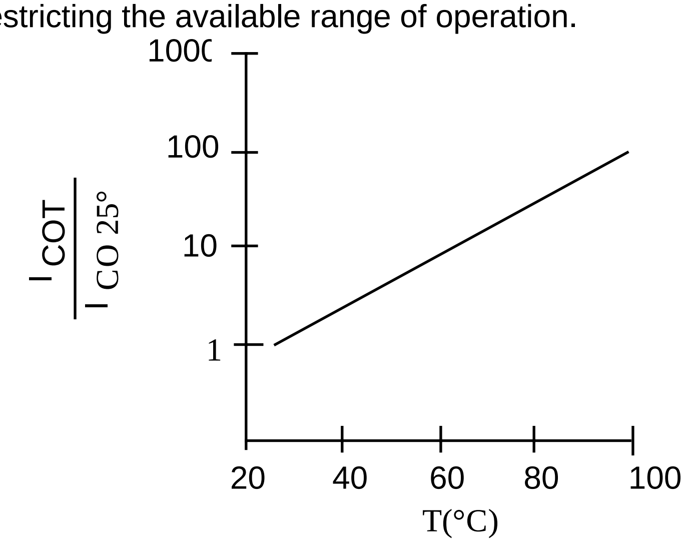


FIGURE 7.4-18: RATIO OF I_{CO} OVER TEMPERATURE T TO I_{CO} AT T = 25°C

Even though it is possible to use the transistor in the common emitter circuit with very small values of currents, the nonlinearity of the device when the base current has a reverse polarity is so pronounced that it is not practical to attempt to do so.

The variation of the base-to-emitter voltage with temperature for fixed values of base and emitter current is the second important thermal property of a transistor requiring compensation. The voltage between base and emitter affects the static operation of the transistor, and it also affects the small signal operation. Because the static, or Q-point for the transistor varies rapidly with temperature if the base voltage is fixed, it is necessary to fix the Q-point in a way to assure that a full range of operating conditions is available over the required range of operating temperature. The static stability must be determined in terms of the practical circuit in use, and the circuit must be designed to provide the required stability.

Reference [6] provides detailed design procedures for thermal stabilization of circuits, as well as design procedures to prevent thermal runaway.

7.5 Fault Tolerant Design

Simply stated, fault tolerant design means providing a system with the ability to operate, perhaps at a degraded but acceptable level, in the presence of faults. The goal of fault tolerance is to intercept the propagation of faults so that failure does not occur, usually by substituting redundant functions affected by a particular fault. Depending on the system operational requirements, and the fault tolerant design techniques being implemented, a fault tolerant system may have to detect, diagnose, confine, mask, compensate and recover from faults. Systems are still being built today where real time reconfiguration, in the presence of a fault, is not required.

These system still need to have the capability to detect and isolate a failure, but may only require manual intervention to reconfigure the system to compensate. Other systems, such as those designed for space applications, may require built-in capabilities to detect, isolate, confine, mask, compensate and recover from faults in the system.

Each of the preceding concepts (i.e., fault detection, isolation, confinement, etc.) are typically related to what is known as redundancy management. Redundancy is typically necessary to achieve fault tolerance, but is not in itself sufficient for fault tolerance. For example, a system may contain redundant elements performing the same function such that in the presence of a fault, at least one of the outputs or results is correct. However, if the user must determine which result is correct, then only the user is performing the fault tolerant function. Only when the system is designed to determine which redundant result or output is correct for the user can we say that the system is fault tolerant. Using this example, redundancy management controls the non-faulty resources to provide the correct result. In this context then, each of the referenced concepts above can now be defined.

(1) Fault Detection: The process of determining that a fault has occurred.

(2) Fault Isolation: The process of determining what caused the fault, or exactly which subsystem or component is faulty.

(3) Fault Containment: The process that prevents the propagation of faults from their origin at one point in a system to a point where it can have an effect on the service to the user.

(4) Fault Masking: The process of insuring that only correct values get passed to the system boundary in spite of a failed component.

(5) Fault Compensation: If a fault occurs and is confined to a subsystem, it may be necessary for the system to provide a response to compensate for output of the faulty subsystem.

7.5.1 Redundancy Techniques

There are essentially two kinds of redundancy techniques employed in fault tolerant designs, space redundancy and time redundancy. Space redundancy provides separate physical copies of a resource, function, or data item. Time redundancy, used primarily in digital systems, involves the process of storing information to handle transients, or encoding information that is shifted in time to check for unwanted changes. Space, or hardware redundancy is the approach most commonly associated with fault tolerant design. Figure 7.5-1 provides a simplified tree-structure of hardware redundancy techniques that have been used or considered in the past.

A detailed discussion of each of the techniques can be found in Section 7.5.3 through 7.5.5.

7.5.1.1 Impact on Testability

As discussed previously, many of today's more sophisticated systems require the ability to not only detect faults, but to diagnose or isolate faults, and to reconfigure the system to avoid system failure. Automated fault detection and isolation has therefore become an essential means of obtaining highly fault tolerant systems. Because of this, the design of the diagnostic system, including any built-in-test (BIT) features and the overall testability of the design are important tradeoffs that need to be made as part of the fault tolerant design process. Table 7.5-1 presents a sample list of hardware fault tolerant design approaches and their impact on diagnostic approaches and BIT.

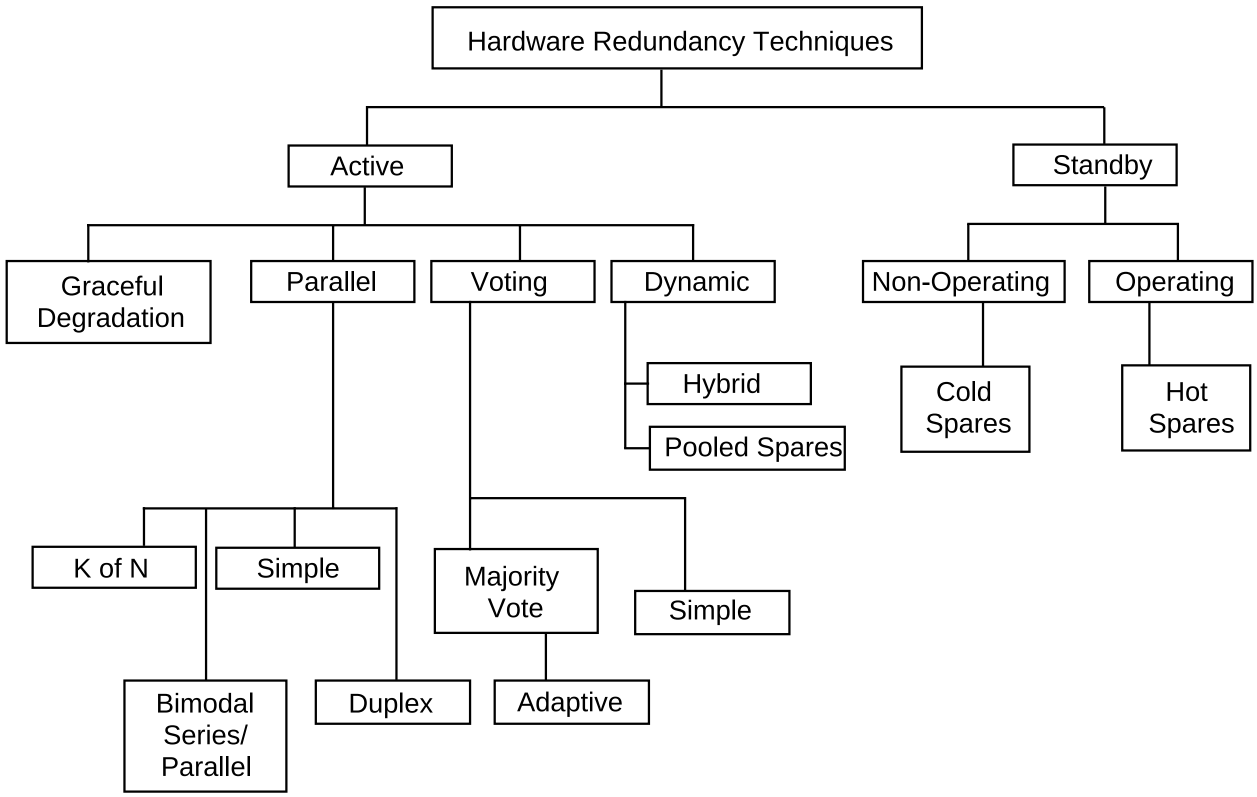


FIGURE 7.5-1: HARDWARE REDUNDANCY TECHNIQUES

TABLE 7.5-1: DIAGNOSTIC IMPLICATIONS OF FAULT TOLERANT DESIGN APPROACHES

FAULT TOLERANT DESIGN TECHNIQUE	DIAGNOSTIC DESIGN IMPLICATIONS	BIT IMPLICATIONS
Active Redundancy	Hardware/Software is more readily available to perform multiple functions.	N/A
Active Redundancy with voting logic	Performance/status-monitoring function assures the operator that the equipment is working properly; failure is more easily isolated to the locked-out branch by the voting logic.	N/A
Stand-by Redundancy	Test capability and diagnostic functions must be designed into each redundant or substitute functional path (on-line AND off-line) to determine their status.	Passive, periodic, or initiated BIT.
Active Redundancy	N/A	Limited to passive BIT (i.e., continuous monitoring) supplemented with periodic BIT.

No matter what technique is chosen to implement fault tolerance in a design, the ability to achieve fault tolerance is increasingly dependent on the ability to detect, isolate, and repair malfunctions as they occur, or are anticipated to occur. This mandates that alternate maintainability diagnostic concepts be carefully reviewed for effectiveness before committing to a final design approach. In particular, BIT design has become very important to achieving a fault tolerant system. When using BIT in fault tolerant system design, the BIT system must:

(1) Maintain a real-time status of the system's assets (both on-line and off-line equipment)

(2) Provide the operator with the status of available system assets

(3) Maintain a record of hardware faults and reconfiguration events required for system recovery during the mission for post-mission evaluation and corrective maintenance.

For fault tolerant systems, it is important that the design's inherent testability provisions include the ability to detect, identify, recover, and if necessary reconfigure, and report equipment malfunctions to operational personnel. Fault tolerant systems often are characterized by complex, non-serial reliability block diagrams, a multitude of backups with non-zero switch-over time, and imperfect fault detection, isolation, and recovery. Therefore it is imperative that effective testability provisions be incorporated in the system design concept. If not, the design, when fielded, will exhibit long troubleshooting times, high false alarm rates, and low levels of system readiness.

7.5.2 Reliability Role in the Fault Tolerant Design Process

The role of the reliability engineer in regards to fault tolerant design requirements is to assure that system reliability requirements are achievable for each of the fault tolerant design approaches being considered. Further, to properly design a fault tolerant system, including a diagnostic scheme, the designer needs to understand how the system can fail, and the effects of those faults. This requires that a failure mode and effects analysis (FMEA) be performed, as a minimum. The FMEA will identify which faults can lead to system failure and therefore must be detected, isolated and removed to maintain system integrity. In general, the reliability design manager must ask the following questions:

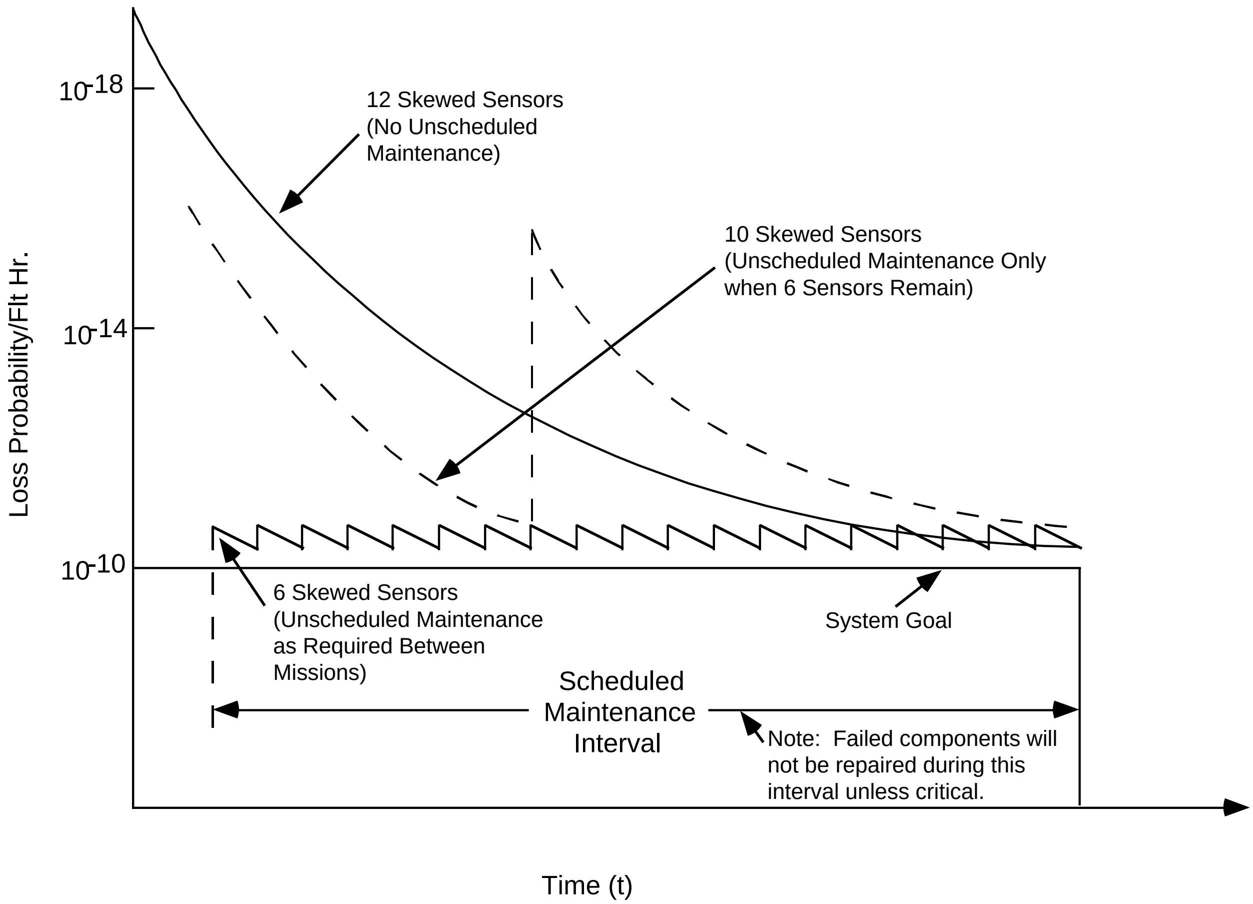
(1) How do the system fault tolerance requirements impact the overall R/M/A requirements?

(2) Where should fault tolerant design methods be applied?
  (a) Which functions involve the most risk to mission success?
  (b) What is the effect of the operating environment
  (c) What maintenance strategy/policy needs to be considered?

(3) What is the effect on Maintainability and Testability?

(4) What are the constraints that affect fault tolerance ?
  (a) cost
  (b) size & weight
  (c) power
  (d) interface complexity
  (e) diagnostic uncertainties

Each of the above questions, and others need to be considered as part of the overall fault tolerant design process. Other reliability tradeoffs to be considered involve analysis of the redundancy approaches being considered for the fault tolerant design. Section 7.5.3 - 7.5.6 provide details on methods of redundancy analysis. In addition to reliability concerns, fault tolerance also requires analysis of the impacts on maintainability and testability. As an example, consider Figure 7.5-2. This figure illustrates a design vs. corrective maintenance tradeoff analysis performed early in the product development phase. In particular, the figure shows the tradeoff of restoration frequency versus the number of sensors being used to meet requirements. This program requires a time period for allocating a scheduled maintenance activity and a probability of less than one in 10 billion per flight hour that a total loss of the skewed sensor function would occur. The tradeoff is made between the number of sensors and the cost of unscheduled maintenance activity associated with each approach. Other tradeoffs, such as cost, power, weight, etc. are also necessary. In general, as in any design analysis support function, the impacts on reliability, maintainability (including testability) and availability of a chosen fault tolerant design approach needs to be performed by the R/M/A professional.



Note: More Frequent Restoration of Redundancy Lowers Fault Tolerance Requirements, But Results in Higher Maintenance Manhours

FIGURE 7.5-2: EFFECT OF MAINTENANCE CONCEPT ON LEVEL OF FAULT TOLERANCE

As a final note on the role of the reliability engineer, consider the following reliability inputs to the development of the specification for a fault tolerant design

(1) Critical mission definition

(2) Quantitative mission reliability

(3) Quantitative maintenance frequency reliability

(4) Description of the storage, transportation, operation, and maintenance environments

(5) Time measure or mission profile

(6) Definition of satisfactory and acceptable degraded system performance

(7) Tolerable failure policy (fail-safe, fail-operational, etc.)

(8) Failure independence

These and other inputs are necessary to ensure that the system specifications are well defined and that they support the ability to clearly define the best approach that also meets R/M/A requirements.

7.5.2.1 Fault Tolerant Design Analysis

The FMEA is a primary reliability analysis, critical to the fault tolerant design process. The reliability engineer will also utilize additional techniques for analyzing the fault tolerant design to verify that it meets reliability requirements. However, many of the evaluation tools used in the past are no longer adequate to deal with more sophisticated fault tolerant designs that include more complex fault handling capabilities. Because fault handling methods include the use of fault detection and fault recovery approaches, any evaluation tool must include the ability to properly account for the effects of imperfect fault coverage (or fault detection) and fault recovery.

Monte Carlo simulation and Markov analysis techniques continue to be used as the primary means of analyzing highly sophisticated fault tolerant designs. These approaches have been modified to incorporate situations where the sequence of failure is important, where the failure is transient or intermittent, or where the response to failure (i.e., detection, isolation, recovery, reconfiguration) is imperfect. In these situations, Markov methods continue to lead the way in evaluation methods. Markov analysis is described in more detail in Section 7.5.6 and will not be discussed in detail here. In general, the Markov approach, which is used to define the specific states that a system can occupy, has been used to incorporate fault handling and recovery. A major limitation to the Markov approach is that the number of system states that must be defined to comprehensively describe a large system and model the behavior of complex fault management schemes can become very large (approaching 10^5 for highly complex systems). A common solution to this problem is to partition the system into smaller systems, evaluate each partition separately, and then combine the results at the system level. However, such an approach is only exact when each partitioned subsystem's fault tolerant behavior is mutually independent of each other. If subsystem dependencies do exist, then an assumption of independence will result in only an approximate solution.

Other approaches that are now becoming more common involve decomposing the system into separate fault-occurrence and fault handling submodels. However, the inputs for this type of approach require knowledge of the distribution and parameter values of: detection, isolation, recovery, rates, etc. The following is a list of assumptions, limitations and sources of error found in existing reliability models:

(1) Solving a fault-handling model in isolation and then reflecting its results in an aggregate model is, itself, an approximation technique. The assumptions necessary to determine a solution typically result in a lower bound (conservative) approximation of the system reliability.

(2) Separate fault-handling models have been assumed to be independent of system state. This requires that the same fault-handling model and choice of parameters be used irrespective of the system's level of degradation. This ignores the fact that for many systems the recovery process is faster if the number of active units is smaller or that the recovery process may be different, depending on the sequence of events in different subsystems.

(3) The common technique of partitioning the system into independent functional subgroups for computational ease is a potential source of error. The magnitude and direction of the error is a function of how truly independent/dependent the subgroups are of each other. If subgroups are assumed independent when in fact they are not, the effect is an overstatement of system reliability/availability. If subgroups are assumed completely dependent when some degree of independence exists, the effect is an understatement of the system's reliability/availability.

(4) Some models assume a constant instantaneous fault-protection coverage factor in lieu of a separate fault handling model. These fail to recognize that during time spent in the intermediate fault-handling states to detect, isolate, and recover/reconfigure, a second item failure could result in system failure. Further, as with fault handling models, these times are generally not constant, but depend on the current state of the system.

(5) Most models require the assumption that the system is perfect at the mission start. Therefore, they cannot evaluate the effects of latent defects (e.g., handling, manufacturing, transportation, prior mission), nor assist in determining the testability payoff or requirements for detection and removing them before the start of the mission.

Models with this limitation cannot be used to evaluate alternate maintenance concepts that include degradation between missions as an acceptable strategy.

(6) Some models require that spares be treated exactly like active units, irrespective of their actual utilization in the system mechanization. This requires that spares are assumed to be "hot" and have the same failure rates and failure modes as the active units. This assumption will cause the model to understate the system reliability in those situations where spares are "cold" or in "stand-by" and/or where their failure rates may be less that those of the active units.

(7) As indicated previously, some models require the assumption that item failure rates are constant throughout time. This will result in an overstatement of system reliability if the items have failure rates that increase with mission time. Some models remove this restriction and permit time-varying failure rates. However, the solution the algorithms employ requires the use of global time (as opposed to local time of entry into a state), thus precluding the use of the model for repairable systems and availability analysis.

7.5.3 Redundancy as a Design Technique

In reliability engineering, redundancy can be defined as the existence of more than one means for accomplishing a given task. In general, all means must fail before there is a system failure.

Thus, if we have a simple system consisting of two parallel elements as shown in Figure 7.5-3 with A_1 having a probability of failure q_1 and A_2 having a probability of failure q_2, the probability of total system failure is

$$Q = q_1 q_2$$

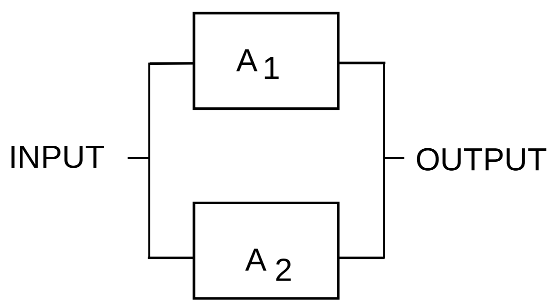


FIGURE 7.5-3: PARALLEL NETWORK

Hence the reliability or probability of no failure is

$$R = 1 - Q = 1 - q_1 q_2$$

For example, assume that A_1 has a reliability r_1 of 0.9 and A_2 a reliability r_2 of 0.8. Then their unreliabilities q_1 and q_2 would be

$$q_1 = 1 - r_1 = 0.1$$

$$q_2 = 1 - r_2 = 0.2$$

and the probability of system failure would be

$$Q = (0.1)(0.2) = 0.02$$

Hence the system reliability would be

$$R = 1 - Q = 0.98$$

which is a higher reliability than either of the elements acting singly. Parallel redundancy is therefore a design tool for increasing system reliability when other approaches have failed. It should be pointed out that while redundancy reduces mission failures, it increases logistics failures.

In general, with n elements in parallel, the overall probability of failure at time t is

$$Q(t) = q_1(t) \bullet q_2(t) \bullet \ldots \bullet q_n(t) \tag{7.9}$$

and the probability of operating without failure is

$$R(t) = 1 - Q(t) = 1 - q_1(t)\, q_2(t) \ldots q_m(t) \tag{7.10}$$

which, because $q_i(t) = 1 - r_i(t)$ for each component, can also be given as

$$R(t) = 1 - [\, 1 - r_1(t)\,]\, [\, 1 - r_2(t)\,] \ldots [\, 1 - r_m(t)\,] \tag{7.11}$$

When each of the component reliabilities is equal, the above equations reduce to

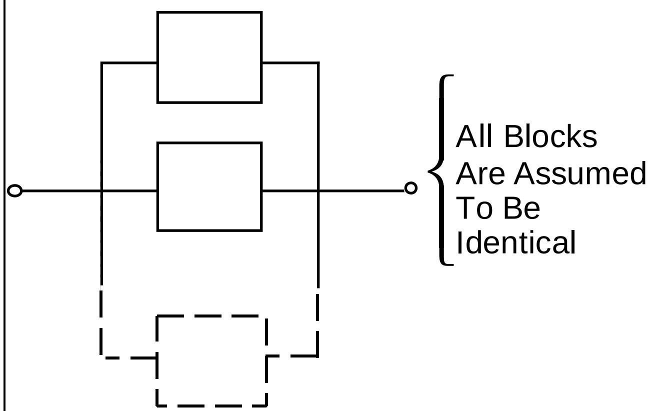
$$Q(t) = [\, q(t)\,]^m \tag{7.12}$$

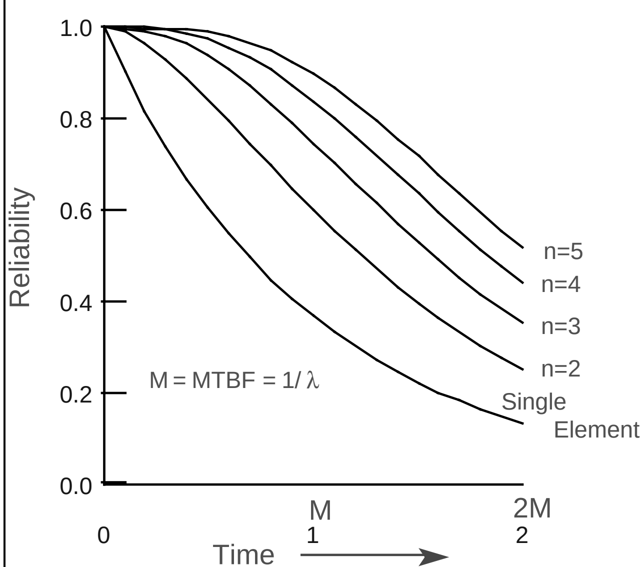
$$R(t) = 1 - [q(t)]^m = 1 - [\, 1 - r(t)\,]^m \tag{7.13}$$

Figure 7.5-4 summarizes the characteristics of simple parallel active redundancy.

So far it has been assumed that parallel components do not interact and that they may be activated when required by ideal failure sensing and switching devices. Needless to say, the latter assumption, in particular, is difficult to meet in practice. Therefore, the potential benefits of redundancy cannot be realized fully. The reader is referred to the cited references, e.g., References [22] and [23], for detailed treatment of redundancy with sensing and switching devices which are most ideal.

Most cases of redundancy encountered will consist of various groupings of series and parallel elements. Figure 7.5-5 typifies such a network. The basic formulas previously given can be used to solve the overall network reliability R_{AC}.





RELIABILITY FUNCTION FOR SIMPLE PARALLEL RELIABILITY

APPLICATION

Provides protection against irreversible hardware failures for continuously operating equipments.

MATHEMATICAL MODEL

$$R = 1 - (1 - e^{-\lambda t})^n$$

APPROXIMATE MODEL

$$R = 1 - (\lambda t)^n$$

for small $\lambda t \leq 0.1$

where:

n	=	number of parallel elements
λ	=	failure rate
R	=	reliability

ADVANTAGES OF PARALLEL REDUNDANCY

- Simplicity
- Significant gain in Reliability from nonredundant element
- Applicable to both analog and digital circuitry

DISADVANTAGES OF PARALLEL REDUNDANCY

- Load sharing must be considered
- Sensitive to voltage division across the elements
- Difficult to prevent failure propagation
- May present circuit design problems

FIGURE 7.5-4: SIMPLE PARALLEL REDUNDANCY: SUMMARY

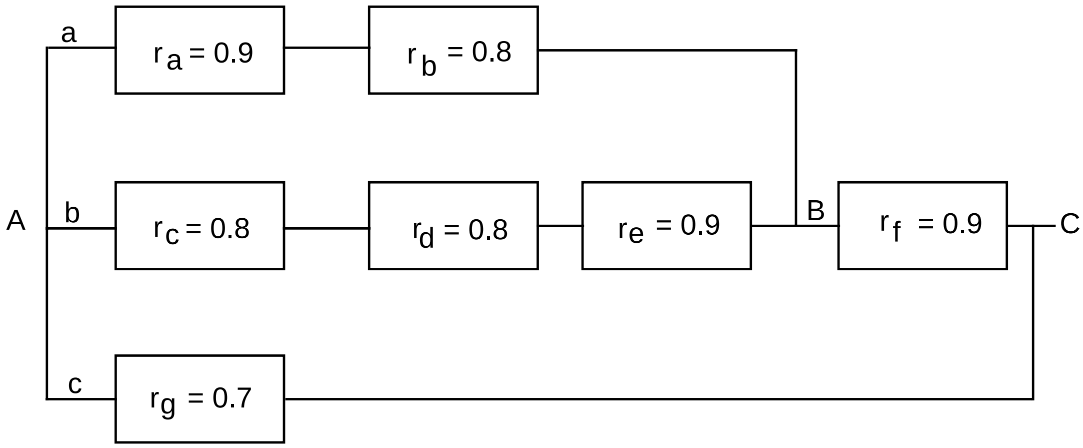


FIGURE 7.5-5: SERIES-PARALLEL REDUNDANCY NETWORK

7.5.3.1 Levels of Redundancy

Redundancy may be applied at the system level (essentially two systems in parallel) or at the subsystem, component, or part level within a system. Figure 7.5-6 is a simplified reliability block diagram drawn to illustrate the several levels at which redundancy can be applied. System D is shown with its redundant alternative D', at the system level. D' is in turn built up of redundant subsystems or components (C_1 and C_2) and redundant parts within components (b_1 and b_2 within Component B). From the reliability block diagram and a definition of block or system success, the paths which result in successful system operation can be determined. For example, the possible paths from Input to Output are:

(1) A, a, b_1, C_1
(2) A, a, b_1, C_2
(3) A, a, b_2, C_1
(4) A, a, b_2, C_2
(5) D

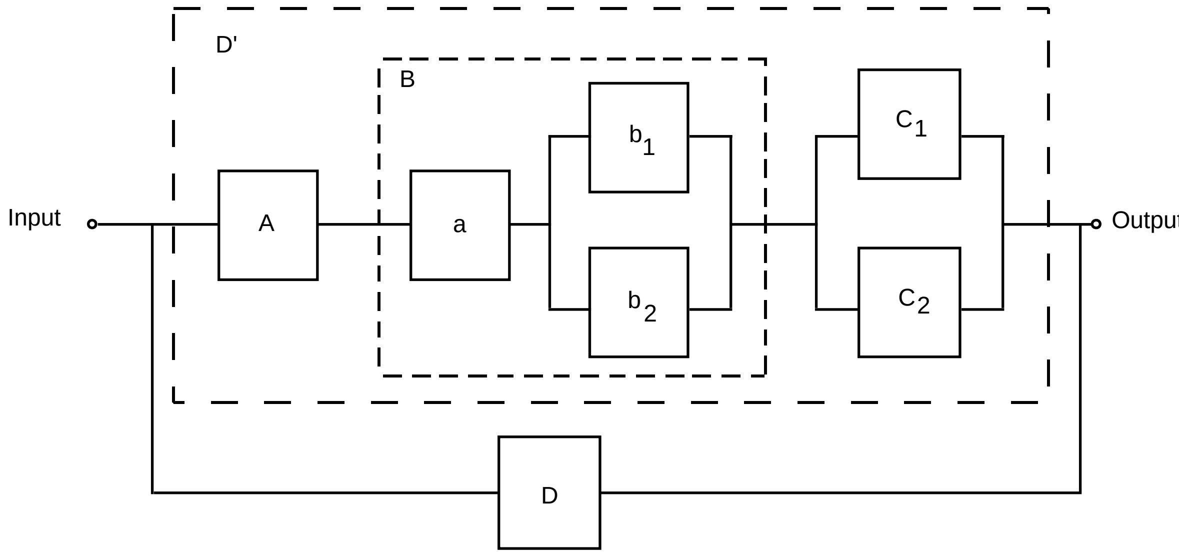


FIGURE 7.5-6: RELIABILITY BLOCK DIAGRAM DEPICTING REDUNDANCY AT THE SYSTEM, SUBSYSTEM, AND COMPONENT LEVELS

The success of each path may be computed by determining an assignable reliability value for each term and applying the multiplicative theorem. The computation of system success (all paths combined) requires a knowledge of the type of redundancy to be used in each case and an estimate of individual element reliability (or unreliability).

7.5.3.2 Probability Notation for Redundancy Computations

Reliability of redundancy combinations is expressed in probabilistic terms of success or failure -- for a given mission period, a given number of operating cycles, or a given number of time independent "events," as appropriate. The "MTBF" measure of reliability is not readily usable because of the nonexponentiality of the reliability function produced by redundancy. Reliability of redundancy combinations which are "time dependent" is therefore computed at a discrete point in time, as a probability of success for this discrete time period. The following notation is applicable to all cases and is used throughout this section:

R = probability of success or reliability of a unit or block

Q = $\overline{R}$ = probability of failure or unreliability of a unit or block

p = probability of success or reliability of an element

q = probability of failure or unreliability of an element

For probability statements concerning an event:

$P(A)$ = probability that A occurs

$P(\bar{A})$ = probability that A does not occur

For the above probabilities:

$$R + Q = 1$$

$$p + q = 1$$

$$P(A) + P(\bar{A}) = 1$$

7.5.3.3 Redundancy Combinations

The method of handling redundancy combinations can be generalized as follows:

(1) If the elements are in parallel and the units in series (Figure 7.5-7), first evaluate the redundant elements to get the unit reliability. Then find the product of all unit reliabilities to obtain the block reliability.

(2) If the elements are in series and the units or paths are in parallel (Figure 7.5-8), first obtain the path reliability by calculating the product of the reliabilities of all elements in each path. Then consider each path as a redundant unit to obtain the block reliability.

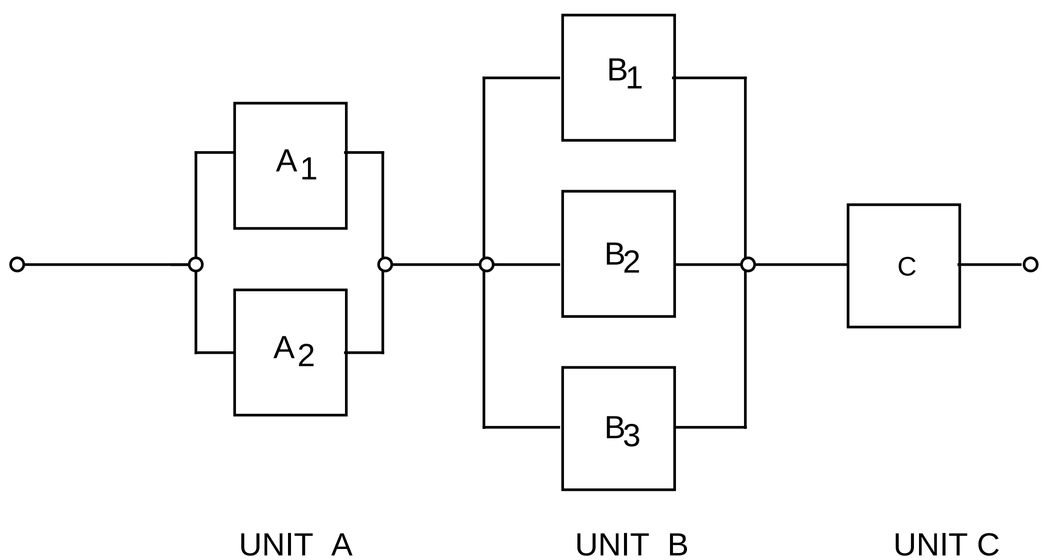


FIGURE 7.5-7: SERIES-PARALLEL CONFIGURATION

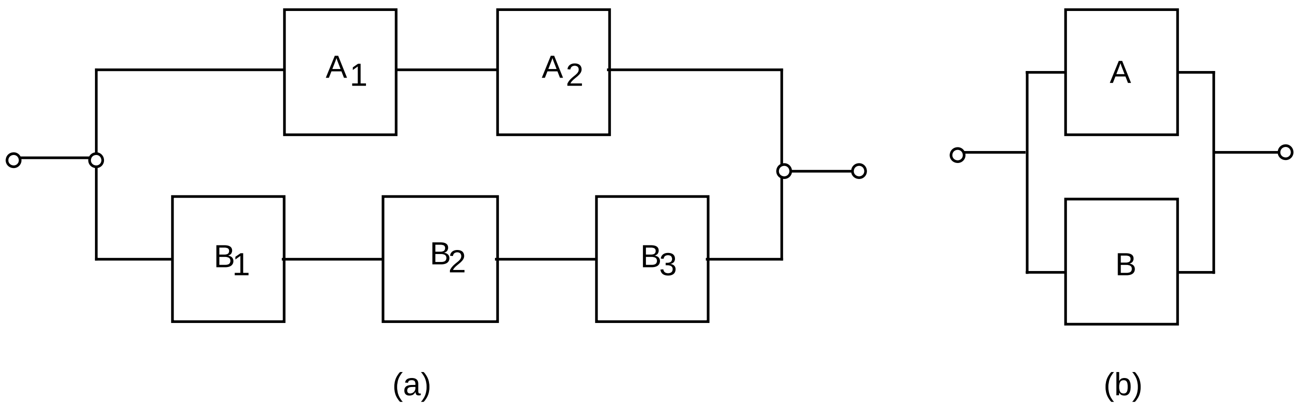


FIGURE 7.5-8: PARALLEL-SERIES CONFIGURATION

In the redundancy combination shown in Figure 7.5-7, Unit A has two parallel redundant elements, Unit B has three parallel redundant elements, and Unit C has only one element. Assume that all elements are independent. For Unit A to be successful, A_1 or A_2 must operate; for Unit B success, B_1, B_2 or B_3 must operate; and C must always be operating for block success. Translated into probability terms, the reliability of Figure 7.5-7 becomes:

$$R = \left[1 - P(\overline{A}_1) \cdot P(\overline{A}_2)\right] \cdot [1 - P(\overline{B}_1) \cdot P(\overline{B}_2) \cdot P(\overline{B}_3)] \cdot P(C)$$

If the probability of success, p, is the same for each element in a unit,

$$R = \left[1 - (1 - p_A)^2] \cdot [1 - (1 - p_B)^3\right] \cdot p_c$$

$$= (1 - q_A^{\,2}) \cdot (1 - q_B^{\,3}) \cdot p_c$$

where:

$$q_i = 1 - p_i$$

Often there is a combination of series and parallel redundancy in a block as shown in Figure 7.5-8. This arrangement can be converted into the simple parallel form shown in Figure 7.5-8 by first evaluating the series reliability of each path:

$$p_A = p_{a_1} p_{a_2}$$

$$p_B = p_{b_1} p_{b_2} p_{b_3}$$

where the terms on the right hand side represent element reliability. Then block reliability can be found from:

$$R = 1 - (1 - p_A) \bullet (1 - p_B)$$

$$= 1 - q_A q_B$$

7.5.4 Redundancy in Time Dependent Situations

The reliability of elements used in redundant configurations is usually time dependent. If the relation between element reliability and time is known, inclusion of the time factor does not change the basic notation and approach to redundancy computation outlined above. As an example, assume two active independent elements in parallel. System reliability is given by:

$$R = p_a + p_b - p_a p_b$$

This equation is applicable for one time interval. To express reliability over a segment of time, the reliability of each element must be expressed as a function of time.

Hence,

$$R(t) = p_a(t) + p_b(t) - p_a(t) p_b(t)$$

where:

$R(t)$ = system reliability for time t, $t > 0$

and

$p_a(t)$, $p_b(t)$ = element reliabilities for time t

The failure pattern of most components is described by the exponential distribution, i.e.:

$$R(t) = e^{-\lambda t} = e^{-t/\theta}$$

where λ is the constant failure rate; t is the time interval over which reliability, R, is measured; and θ is the mean-time-between-failure.

For two elements in series with constant failure rates λ_a and λ_b, using the product rule of reliability gives:

$$R(t) = p_a(t)\ p_b(t)$$

$$= e^{-\lambda_a t} e^{-\lambda_b t} = e^{-(\lambda_a + \lambda_b)t}$$

The system reliability, R(t), function is also exponential. With redundant elements present in the system, however, the system reliability function is not itself exponential. This is illustrated by two operative parallel elements whose failure rates are constant. From:

$$R(t) = p_a + p_b - p_a p_b$$

$$R(t) = e^{-(\lambda_a)t} + e^{-(\lambda_b)t} - e^{-(\lambda_a + \lambda_b)t}$$

which is not of the simple exponential form $e^{-\lambda t}$. Element failure rates cannot, therefore, be combined in the usual manner to obtain the system failure rate if considerable redundancy is inherent in the design.

Although a single failure rate cannot be used for redundant systems, the mean-time-to-failure of such systems can be evaluated. The mean life of a redundant "pair" whose failure rates are λ_a and λ_b, respectively, can be determined from:

$$MTBF = \int_{o}^{\infty} R(t)dt = \frac{1}{\lambda_a} + \frac{1}{\lambda_b} - \frac{1}{\lambda_a + \lambda_b}$$

If the failure rates of both elements are equal, then,

$$R(t) = 2e^{-\lambda t} - e^{-2\lambda t}$$

and

$$MTBF = \frac{3}{2\lambda} = \frac{3}{2}\ \theta$$

For three independent elements in parallel, the reliability function is:

$$R(t) = 1 - \left[(1 - e^{-\lambda_a t})(1 - e^{-\lambda_b t})(1 - e^{-\lambda_c t}) \right]$$

and

$$MTBF = \frac{1}{\lambda_a} + \frac{1}{\lambda_b} + \frac{1}{\lambda_c} - \frac{1}{\lambda_a + \lambda_b} - \frac{1}{\lambda_a + \lambda_c} - \frac{1}{\lambda_b + \lambda_c} + \frac{1}{\lambda_a + \lambda_b + \lambda_c}$$

If

$$\lambda_a = \lambda_b = \lambda_c = \lambda$$

then

$$R(t) = 3e^{-\lambda t} - 3e^{-2\lambda t} + e^{-3\lambda t}$$

and

$$MTBF = \frac{3}{\lambda} - \frac{3}{2\lambda} + \frac{1}{3\lambda} = \frac{1}{\lambda} + \frac{1}{2\lambda} + \frac{1}{3\lambda} = \frac{11}{6\lambda} = \frac{11}{6}\,\theta$$

In general, for n active parallel elements, each element having the same constant failure rate, λ,

$$R(t) = 1 - \left(1 - e^{-\lambda t}\right)^n$$

and

$$MTBF = \sum_{i=1}^{n} \frac{1}{i\lambda} = \sum_{i=1}^{n} \frac{\theta}{i}$$

7.5.5 Redundancy Considerations in Design

The two basic types of redundancy are:

(1) Active Redundancy: External components are not required to perform the function of detection, decision and switching when an element or path in the structure fails. The redundant units are always operating and automatically pick up the load for a failed unit. An example is a multi-engined aircraft. The aircraft can continue to fly with one or more engines out of operation.

(2) Standby Redundancy: External elements are required to detect, make a decision and switch to another element or path as a replacement for a failed element or path. Standby units can be operating (e.g., a redundant radar transmitter feeding a dummy load is switched into the antenna when the main transmitter fails) or inactive (e.g., a spare radio is turned on when the primary radio fails).

Table 7.5-2 summarizes a variety of redundancy techniques. The most important of these are discussed further later in this section.

The application of redundancy is not without penalties. It will increase weight, space requirements, complexity, cost, and time to design. The increase in complexity results in an increase in unscheduled maintenance. Thus, safety and mission reliability is gained at the expense of adding an item(s) in the unscheduled maintenance chain. The increase in unscheduled maintenance may be counteracted by reliability improvement techniques such as design simplification, derating, and the use of more reliable components, as discussed elsewhere in this Handbook.

The decision to use redundant design techniques must be based on analysis of the tradeoffs involved. Redundancy may prove to be the only available method, when other techniques of improving reliability, e.g., derating, simplification, better components, have been exhausted, or when methods of item improvement are shown to be more costly than duplications. When preventive maintenance is planned, the use of redundant equipment can allow for repair with no system downtime. Occasionally, situations exist in which equipments cannot be maintained, e.g., satellites; then redundant elements may be the best way to significantly prolong operating time.

TABLE 7.5-2: REDUNDANCY TECHNIQUES

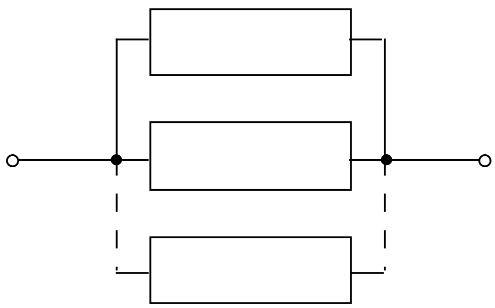
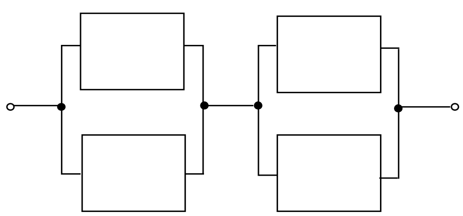
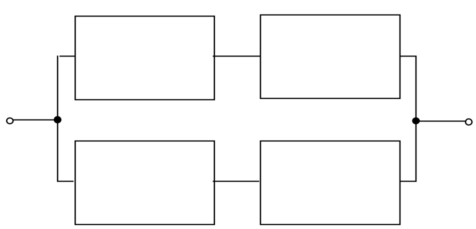
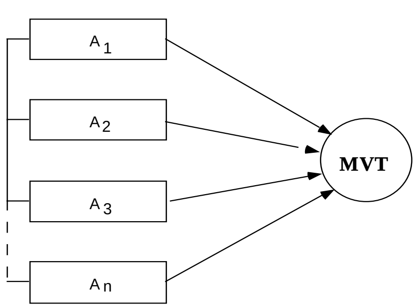
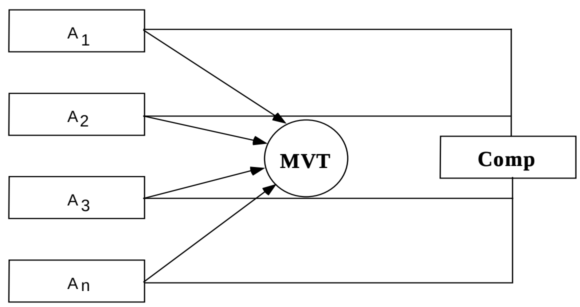
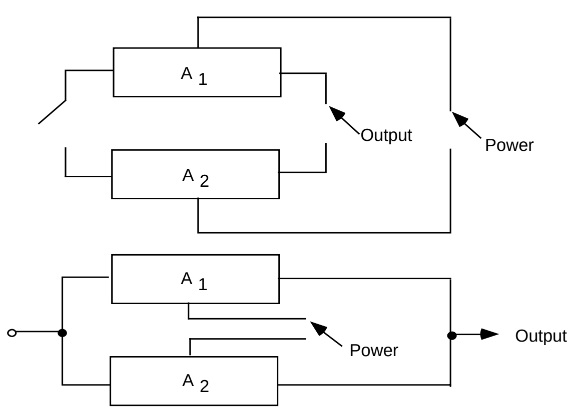
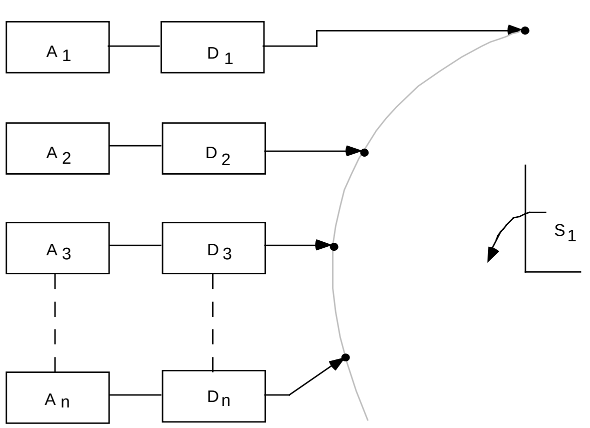
<u>Simple Parallel Redundancy (Active Redundancy)</u> 	In its simplest form, redundancy consists of a simple parallel combination of elements. If any element fails open, identical paths exist through parallel redundant elements.
(a) Bimodal Parallel/Series Redundancy  (b) Bimodal Series/Parallel Redundancy 	A series connection of parallel redundant elements provides protection against shorts and opens. Direct short across the network due to a single element shorting is prevented by a redundant element in series. An open across the network is prevented by the parallel element. Network (a) is useful when the primary element failure mode is open. Network (b) is useful when the primary element failure mode is short.
<u>Majority Voting Redundancy</u> 	Decision can be built into the basic parallel redundant model by inputting signals from parallel elements into a voter to compare each element's signal with the signals of the other elements. Valid decisions are made only if the number of useful elements exceeds the failed elements.

TABLE 7.5-2: REDUNDANCY TECHNIQUES (CONT'D)

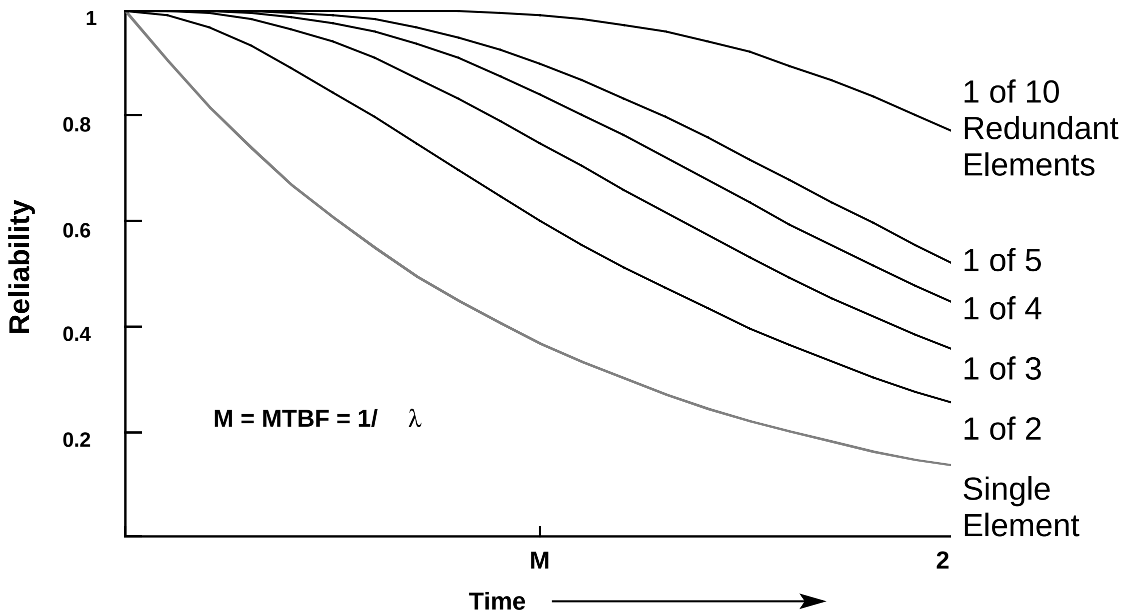
Adaptive Majority Logic 	This technique exemplifies the majority logic configuration discussed previously with a comparator and switching network to switch out or inhibit failed redundant elements.
Standby Redundancy 	A particular redundant element of a parallel configuration can be switched into an active circuit by connecting outputs of each element to switch poles. Two switching configurations are possible. 1) The elements may be isolated by the switch until switching is completed and power applied to the element in the switching operation. 2) All redundant elements are continuously connected to the circuit and a single redundant element activated by switching power to it.
Operating Standby Redundancy 	In this application, all redundant units operate simultaneously. A sensor on each unit detects failures. When a unit fails, a switch at the output transfers to the next unit and remains there until failure.

In general, the reliability gain for additional redundant elements decreases rapidly for additions beyond a few parallel elements. As illustrated by Figure 7.5-9 for simple parallel redundancy, there is a diminishing gain in reliability and MTBF as the number of redundant elements is increased. As shown for the simple parallel case, the greatest gain achieved through addition of the first redundant element is equivalent to a 50% increase in the system MTBF. Optimization of the number of parallel elements is discussed in Section 7.5.5.5.

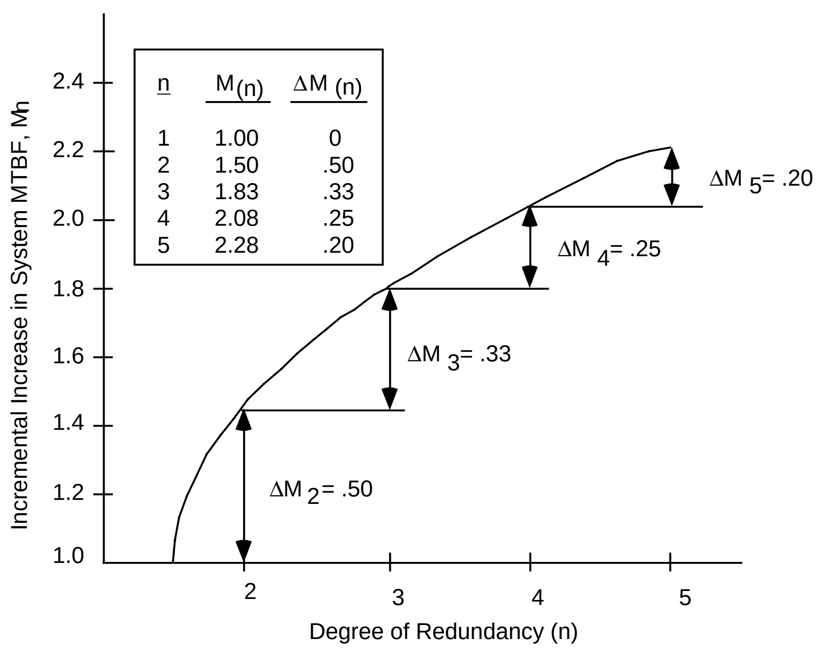
In addition to maintenance cost increases due to repair of the additional elements, reliability of certain redundant configurations may actually be less than that of a single element. This is due to the serial reliability of switching or other peripheral devices needed to implement the particular redundancy configuration. Care must be exercised to insure that reliability gains are not offset by increased failure rates due to switching devices, error detectors and other peripheral devices needed to implement the redundancy configurations. One case where the reliability of switching devices must be considered is that of switching redundancy. This occurs when redundant elements are energized but do not become part of the circuit until switched in after the primary element fails. See Section 7.5.5.2.6 for further discussion.

The effectiveness of certain redundancy techniques (especially standby) can be enhanced by repair. Standby redundancy allows repair of the failed unit (while operation of the good unit continues uninterrupted) by virtue of the switching function built into the standby redundant configuration. Through continuous or interval monitoring, the switchover function can provide an indication that failure has occurred and operation is continuing on the alternate channel. With a positive failure indication, delays in repair are minimized. A further advantage of switching is related to built-in test (BIT) objectives. Built-in test can be readily incorporated into a sensing and switchover network for ease of maintenance purposes.

An illustration of the enhancement of redundancy with repair is shown in Figure 7.5-10. The increased reliability brought about by incorporation of redundancy is dependent on effective isolation of redundant elements. Isolation is necessary to prevent failure effects from adversely affecting other parts of the redundant network. In some cases, fuses or circuit breakers, overload relays, etc., may be used to protect the redundant configuration. These items protect a configuration from secondary effects of an item's failure so that system operation continues after the element failure. The susceptibility of a particular redundant design to failure propagation may be assessed by application of failure mode and effects analysis as discussed in Section 7.8. The particular techniques addressed there offer an effective method of identifying likely fault propagation paths.



(a) Simple Active Redundancy for One of n Element Required



(b) Incremental Increase in System MTBF for n Active Elements

FIGURE 7.5-9: DECREASING GAIN IN RELIABILITY AS NUMBER OF ACTIVE ELEMENTS INCREASES

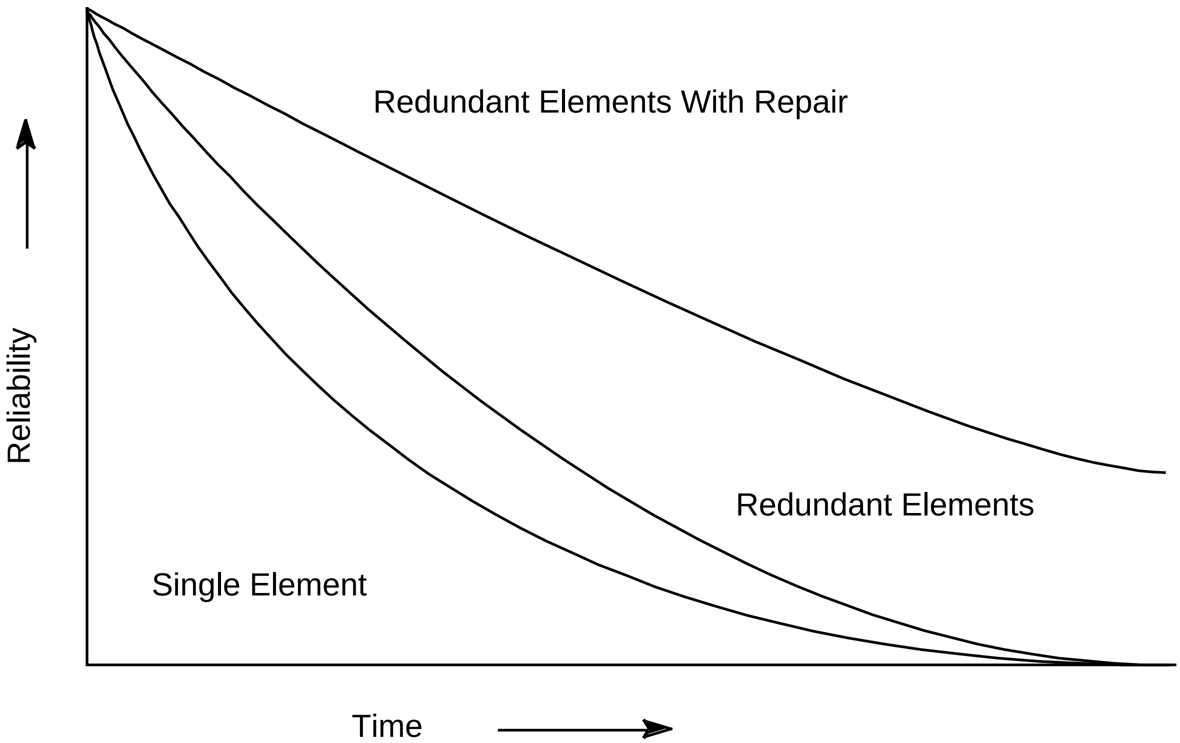


FIGURE 7.5-10: RELIABILITY GAIN FOR REPAIR OF
SIMPLE PARALLEL ELEMENT AT FAILURE

Redundancy may be incorporated into protective circuits[3] as well as the functional circuit which it protects. Operative redundancy configurations of protection devices (e.g., fuse, circuit breaker) can be used to reduce the possibility that the "protected" circuit is not completely disabled should the protective circuit device open prematurely or fail to open due to overcurrent.

The incorporation of redundancy into a design must take into account "checkability." Some items may not be checkable prior to mission start. Such items must then be assumed to be functional at the beginning of the mission. In reality, pre-mission failures of redundant items could be masked. If it is not known that redundant elements are operational prior to mission start, then the purpose of redundancy can be defeated because the possibility exists of starting a mission without the designed redundancy (a reliability loss). The designer must take this into account for built-in test planning, inclusion of test points, packaging, etc., when redundancy is used in system design.

---

[3] It should be noted that the need for or usefulness of modeling reliability at the circuit level is not universally accepted. In particular, many engineers question the value of such modeling for modern technologies. Discussion of circuit-level modeling is included here since it may be of value in some instances.

7.5.5.1 Partial Redundancy

Instances in which the system is successful if at least one of n parallel paths is successful has been discussed. In other instances, at least k out of n elements must be successful. In such cases, the reliability of the redundant group (each with the same Probability of Success, p) is given by a series of additive binomial terms in the form of

$$P(k, n \mid p) = \binom{n}{k} p^{k} (1 - p)^{n-k}$$

Two examples of partial redundancy follow.

Example 1:

A receiver has three channels. The receiver will operate if at least two channels are successful, that is, if k = 2 or k = 3. The probability of each channel being successful is equal to p; then

$$R = P(2, 3 \mid p) + P(3, 3 \mid p)$$

$$R = \binom{3}{2} p^{2} (1 - p) + \binom{3}{3} p^{3} (1 - p)^{0}$$

$$R = 3p^{2} (1 - p) + p^{3}$$

$$R = 3p^{2} - 2p^{3}$$

Use of the binomial formula becomes impractical for hand calculation in multi-element partial redundant configurations when the values of n and k become large[4]. In these cases, the normal approximation to the binomial may be used. The approach can be best illustrated by an example.

Example 2:

A new transmitting array is to be designed using 1000 RF elements to achieve design goal performance for power output and beam width. A design margin has been provided, however, to permit a 10% loss of RF elements before system performance becomes degraded below the acceptable minimum level. Each element is known to have a failure rate of 1000×10^{-6} failures per hour. The proposed design is illustrated in Figure 7.5-11, where the total number of elements is n = 1000; the number of elements required for system success is k = 900; and, the number of element failures permitted is r = 100. It is desired to compute and plot the reliability function for the array.

[4] See any good textbook on probability and statistics.

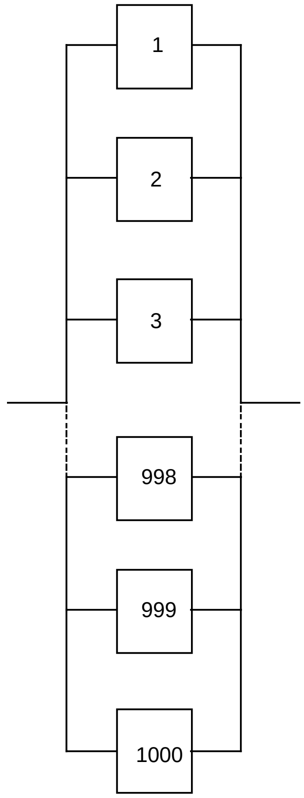


FIGURE 7.5-11: PARTIAL REDUNDANT ARRAY

For each discrete point of time, t, the system reliability function, $R_s(t)$, is given by the binomial summation as:

$$R_s(t) = \sum_{x=0}^{r} \binom{n}{x} p^{n-x} q^{x}$$

$$= \sum_{x=0}^{100} \binom{1000}{x} \left(e^{-\lambda t}\right)^{n-x} (1 - e^{-\lambda t})^{x}$$

where:

$q = 1 - e^{-\lambda t}$

$p = e^{-\lambda t}$

x = number of failures

λ = element failure rate

This binomial summation can be approximated by the standard normal distribution function using Table 7.5-3 to compute reliability for the normalized statistic z.

TABLE 7.5-3: RELIABILITY CALCULATIONS FOR EXAMPLE 2

Time, t	z	$F(z) = R_s(t)$
90	1.57	.942
95	.989	.8389
105	0.0	.500
110	-.42	.337
120	-1.30	.097
130	-2.03	.021

Note that $R_s(t) = F(z)$

where:

$$z = \frac{x - \mu}{\sigma} = \frac{x - nq}{\sqrt{npq}} = \frac{x - n(1 - e^{-\lambda t})}{\sqrt{n(1 - e^{-\lambda t})e^{-\lambda t}}}$$

By observation, it can be reasoned that system MTBF will be approximately 100 hours, since 100 element failures are permitted and one element fails each hour of system operation. A preliminary selection of discrete points at which to compute reliability might then fall in the 80- to 100-hour bracket.

At 80 hours:

$$q = 1 - e^{-\lambda t} = 1 - e^{-(1000 \bullet 10^{-6} \bullet 80)} = .077$$

$$p = e^{-1000} \bullet 10^{-6} \bullet 80 = .923$$

$$\mu = nq = 1000\ (1 - e^{-1000} \bullet 10^{-6} \bullet 80) = 77$$

$$\sigma = \sqrt{npq} = \sqrt{1000\ (.077)\ (.923)} = \sqrt{71.07} = 8.4$$

$$x = 100$$

$$z_{80} = \frac{100 - 77}{8.4} = 2.74$$

$$R_s(80) \quad = \quad F(z_{80}) = F(+2.74) = .997, \text{ from standard normal tables}$$

At 100 hours:

$$\mu = nq \quad = \quad 1000\left(1 - e^{-1000 \bullet 10^{-6} \bullet 100}\right) = 95$$

$$p \quad = \quad e^{-1000 \bullet 10^{-6} \bullet 100} = .905$$

$$\sigma \quad = \quad \sqrt{npq} = \sqrt{86} = 9.3$$

$$x \quad = \quad 100$$

$$z_{100} \quad = \quad \frac{100 - 95}{9.3} = 0.54$$

$$R_s(100) \quad = \quad F(z_{100}) = F(+.54) = .705$$

These points are then used to plot the reliability function for the array, shown in Figure 7.5-12. Also shown in the figure are curves for r=0, 50, and 150.

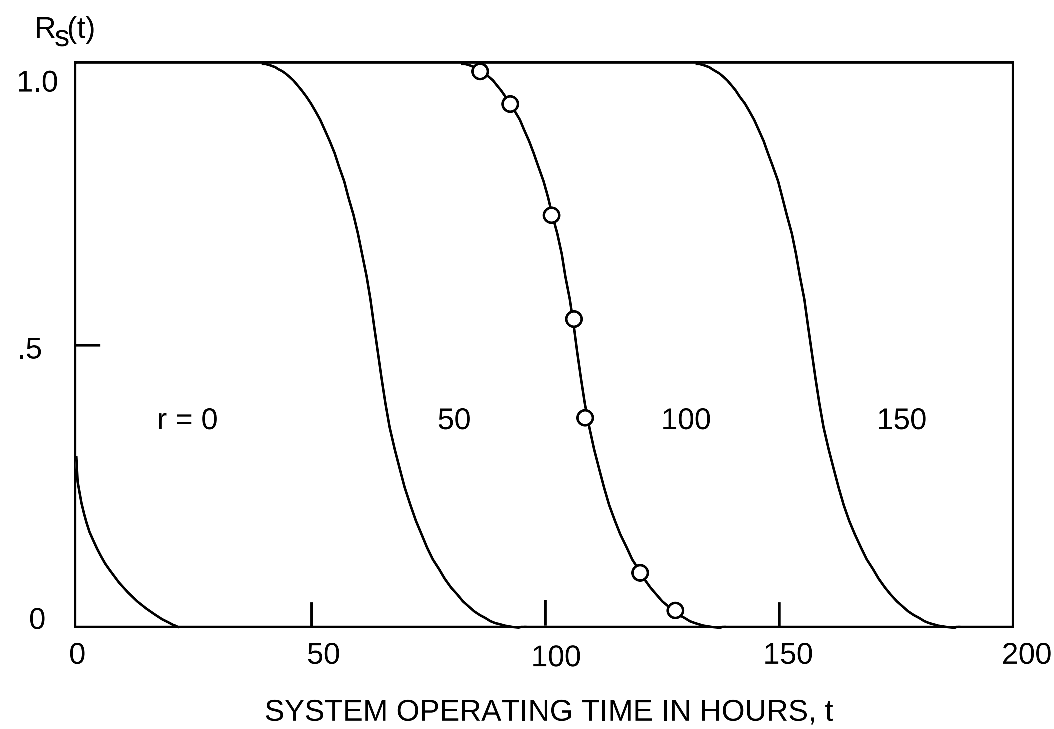


FIGURE 7.5-12: RELIABILITY FUNCTIONS FOR PARTIAL REDUNDANT ARRAY OF FIGURE 7.5-11

7.5.5.2 Operating Standby Redundancy

Until now we have dealt with circuits where it was assumed that switching devices were either absent or failure free. We now deal with circuits whose redundant elements are continuously energized but do not become part of the circuit until switched in after a primary element fails. We will consider two modes of failure that can be associated with the switching mechanism:

a. Type (1). The switch may fail to operate when it is supposed to.

b. Type (2). The switch may operate without command (prematurely).

In the following discussion

q_S = probability of a Type (1) failure

q'_S = probability of a Type (2) failure

7.5.5.2.1 Two Parallel Elements

Consider the system in Figure 7.5-13. There are three possible states that could lead to system failure:

a. A succeeds, B fails, switch fails (Type 2).

b. A fails, B succeeds, switch fails (Type 1).

c. A fails, B fails.

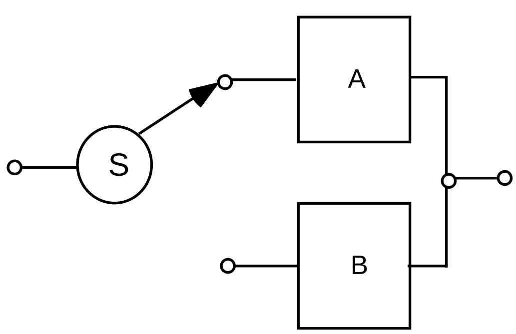


FIGURE 7.5-13: REDUNDANCY WITH SWITCHING

The unreliability of the system, Q, is found from

$$Q = p_a q_b q'_s + q_a p_b q_s + q_a q_b$$

As an example, assume

$$q_a = q_b = 0.2$$

and

$$q_s = q'_s = 0.1$$

Then

$$Q = p_a q_b q'_s + q_a p_b q_s + q_a q_b$$

$$= (0.8)(0.2)(0.1) + (0.2)(0.8)(0.1) + (0.2)(0.2)$$

$$= 0.072$$

$$R = 1 - Q$$

$$= 1 - 0.072$$

$$= 0.928$$

If we are not concerned with Type (2) failures,

$$q'_s = 0$$

and the unreliability is

$$Q = q_a p_b q_s + q_a q_b$$

$$= (0.2)(0.8)(0.1) + (0.2)(0.2)$$

$$= 0.056$$

$$R = 1 - 0.056 = 0.944$$

7.5.5.2.2 Three Parallel Elements

Figure 7.5-14 illustrates this type circuit. It operates as follows: If A fails, S switches to B. If B then fails, S switches to C. Enumerating all possible switching failures shows two kinds of Type (1) failure and four kinds of Type (2) failure:

a. Type (1) Switching Failures:

1. q_{s1} - A fails, S does not switch to B.

2. q_{s2} - A fails, S switches to B, B fails, S fails to switch to C.

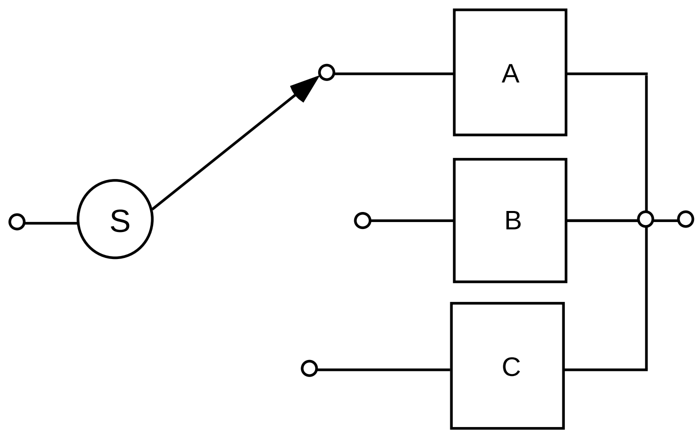


FIGURE 7.5-14: THREE-ELEMENT REDUNDANT CONFIGURATIONS WITH SWITCHING

b. Type (2) Switching Failures:

1. q'_{s3} - A succeeds, but S switches to B.
2. q'_{s4} - A succeeds, S switches to B, B fails, S does not switch to C.
3. q'_{s5} - A succeeds, S switches to B, B succeeds, S switches to C.
4. q'_{s6} - A fails, S switches to B, B succeeds, S switches to C.

The probability of switching failures must be considered in modeling redundancy with switching. The consideration of such failures can be complex. If the switching reliability is high in comparison with element reliability (i.e., switch failure rate is one-tenth that of the element failure rate), it is often possible to simplify the model with an acceptable loss of accuracy by ignoring switch failures. For more detailed information, the reader is referred to textbooks on the subject and Refs. [22] and [24].

7.5.5.2.3 Voting Redundancy

Figure 7.5-15 shows three elements, A, B, and C, and the associated switching and comparator circuit which make up a voting redundant system. The circuit function will always be performed by an element whose output agrees with the output of at least one of the other elements. At least two good elements are required for successful operation of the circuit. Two switches are provided so that a comparison of any two outputs of the three elements can be made. A comparator circuit is required that will operate the two switches so that a position is located where the outputs again agree after one element fails.

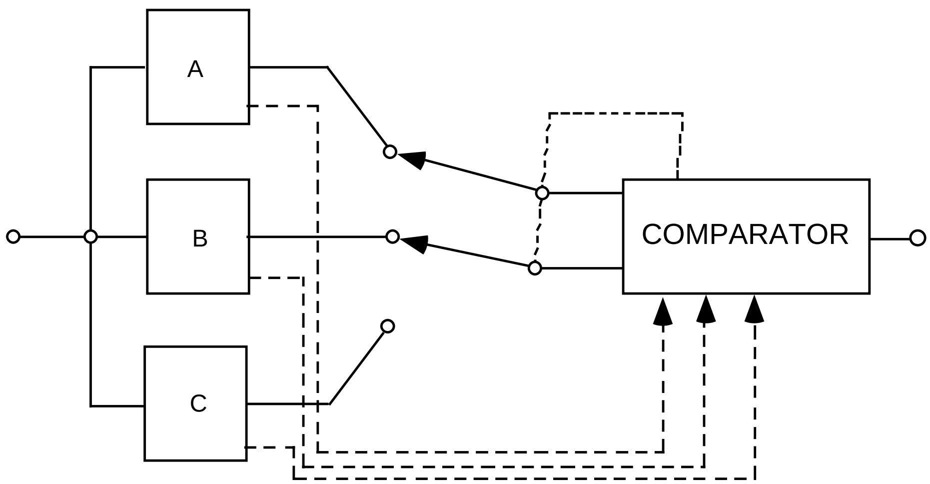


FIGURE 7.5-15: THREE-ELEMENT VOTING REDUNDANCY

If comparison and switching are failure free, the system will be successful as long as two or three elements are successful. In this case,

$$R = p_a p_b + p_a p_c + p_b p_c - 2p_a p_b p_c$$

If failure free switching cannot be assumed, conditional probabilities of switching operation have to be considered. To simplify the discussion, consider the probability of the comparator and switches failing in such a manner that the switches remain in their original positions. If this probability is q_s, then

$$R = p_a p_b + (p_a p_c + p_b p_c - 2p_a p_b p_c)(1 - q_s)$$

Example of a Voting Redundant System

Let all three elements have the same probability of success, 0.9, i.e., $p_a = p_b = p_c = 0.9$. Assume that the comparator switch has a probability of failing (q_s) of 0.01:

$$R = .9^3 + \left[.9^2 + .9^2 - 2(.9)^3 \right] [1 - .01]$$

$$R = .970$$

Information and expressions for the general majority voting case are given in Figure 7.5-16.

7.5.5.3 Inactive Standby Redundancy

In a system with redundant elements on an inactive standby basis (not energized), no time is accumulated on a secondary element until a primary element fails. For a two-element system (see Figure 7.5-13) the reliability function can be found directly as follows. The system will be successful at time t if either of the following two conditions hold (let A be the primary element):

a. A is successful up to time t.

b. A fails at time $t_1 < t$, and B operates from t_1 to t.

For the exponential case where the element failure rates are λ_a and λ_b, reliability of the standby pair is given by

$$R(t) = \frac{\lambda_b}{\lambda_b - \lambda_a} e^{-\lambda_a t} - \frac{\lambda_a}{\lambda_b - \lambda_a} e^{-\lambda_b t}$$

This is a form of the mixed exponential and it does not matter whether the more reliable element is used as the primary or as the standby element.

The mean-time-to-failure of the system is

$$\text{MTBF} = \frac{\lambda_a + \lambda_b}{\lambda_a \lambda_b}$$

$$= \theta_a + \theta_b$$

$$= 2\theta \text{ when } \theta_a = \theta_b = \theta$$

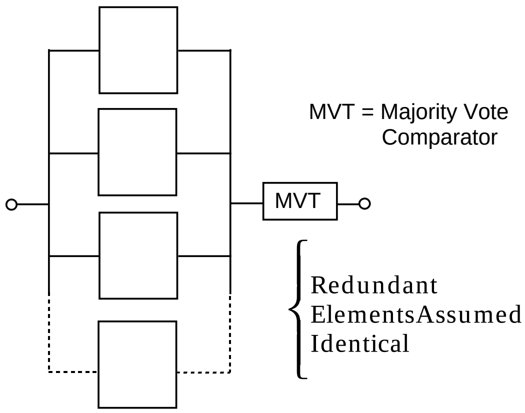
For n elements of equal reliability, it can be shown that,

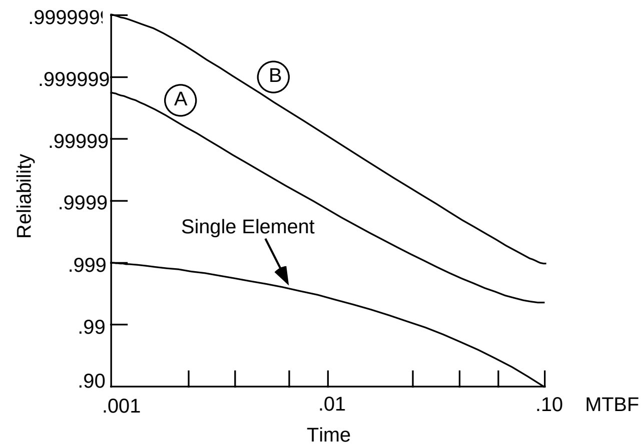
$$R(t) = e^{-\lambda t} \sum_{r=0}^{n-1} \frac{(\lambda t)^r}{r!}$$

where:

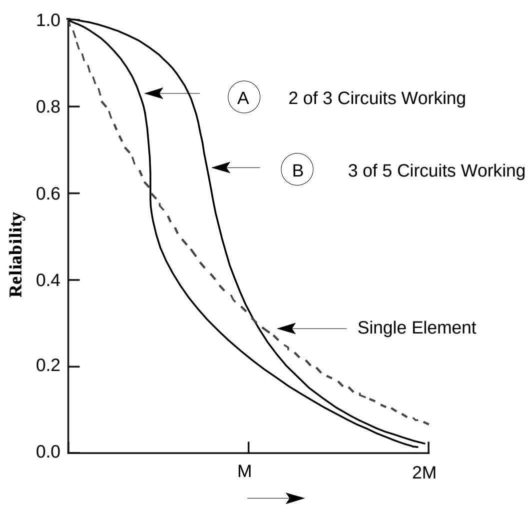
r is the number of failures

$$MTBF = \frac{n}{\lambda} = n\theta$$





Reliability For Mission Time Less Than One MTBF



Reliability Function for Majority Voting Redundancy

APPLICATION

Generally used with logic circuitry for either continuous or intermittent operation. This technique with minor modification, may be put in the form of adaptive majority logic or gate connector redundancy

MATHEMATICAL MODEL

$$R = \left[\sum_{i=0}^{n} \binom{2n+1}{i} \left(1 - e^{-\lambda t}\right) e^{-\lambda t(2n+1-i)}\right] e^{-\lambda_m t}$$

ADVANTAGES

- Can be implemented to provide indication of defective elements
- Can provide a significant gain in reliability for short mission times (less than one MTBF)

DISADVANTAGES

- Requires voter reliability be significantly better than that of element
- Lower reliability for long mission time (greater than one MTBF)

FIGURE 7.5-16: MAJORITY VOTING REDUNDANCY

Figure 7.5-17 is a chart relating system reliability to the reliability of individual operating standby redundant parallel elements as a function of mission time, t/θ. By entering the chart at the time period of interest and proceeding vertically to the allocated reliability requirement, the required number of standby elements can be determined.

Example of Inactive Standby Redundancy

A critical element within a system has a demonstrated MTBF, θ = 100 hours. A design requirement has been allocated to the function performed by this element of R_S = .98 at 100 hours. This corresponds to a 30-to-1 reduction in unreliability compared with that which can be achieved by a single element. In this case, n = 4 will satisfy the design requirement at t/θ = 1. In other words, a four-element standby redundant configuration would satisfy the requirement. Failure rates of switching devices must next be taken into account.

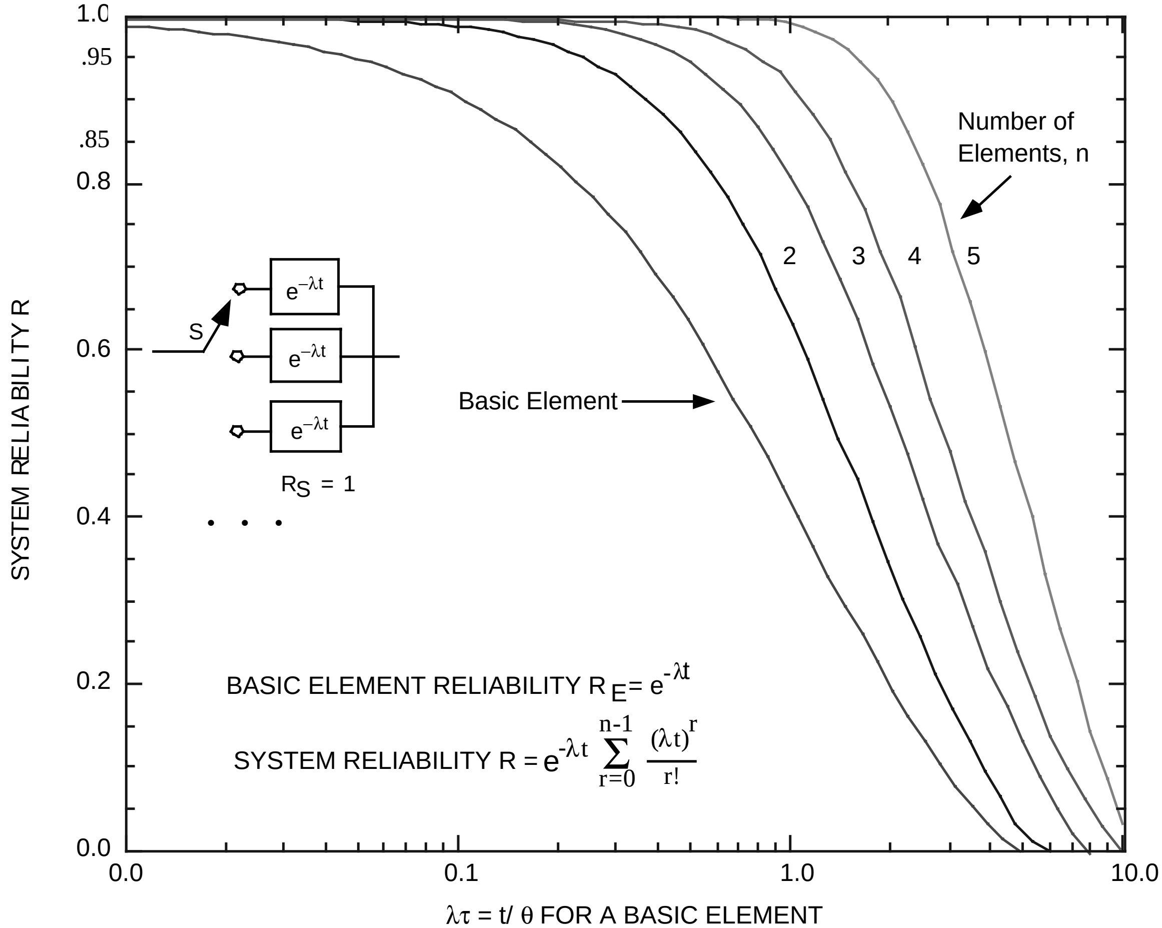


FIGURE 7.5-17: SYSTEM RELIABILITY FOR n STANDBY REDUNDANT ELEMENTS

7.5.5.4 Dependent Failure Probabilities

Up to this point, it has been assumed that the failure of an operative redundant element has no effect on the failure rates of the remaining elements. Dependent failures might occur, for example, with a system having two elements in parallel where both elements share the full load.

An example of conditional or dependent events is illustrated by Figure 7.5-18. Assume elements A and B are both fully energized, and normally share or carry half the load, L/2. If either A or B fails, the survivor must carry the full load, L. Hence, the probability that one fails is dependent on the state of the other, if failure probability is related to load or stress. The system is operating satisfactorily at time t if either A or B or both are operating successfully.

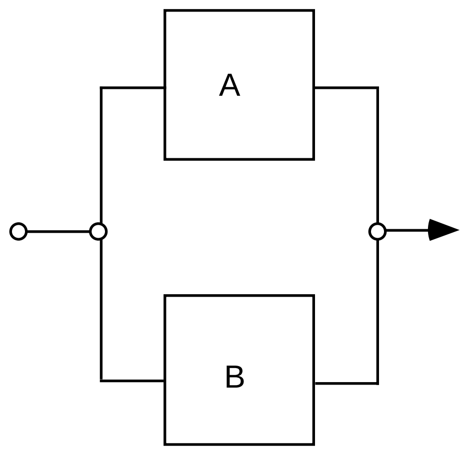


FIGURE 7.5-18: LOAD SHARING REDUNDANT CONFIGURATION

Figure 7.5-19 illustrates the three possible ways the system can be successful. The bar above a letter represents a failure of that element. A primed letter represents operation of that element under full load; absence of a prime represents operation under half load. If the elements' failure times are exponentially distributed and each has a mean life of θ under load L/2 and $\theta' = \theta/k$ under load L where $k \geq 0$, block reliability is given below without derivation:

$$R(t) = \frac{2\theta'}{2\theta' - \theta} e^{-t/\theta'} - \frac{\theta}{2\theta' - \theta} e^{-2t/\theta}$$

System mean life is equal to

$$\theta_s = \theta/k + \theta/2$$

When k = 1, the system is one in which load sharing is not present or an increased load does not affect the element failure probability. Thus, for this case, θ_s is equal to $3\theta/2$.

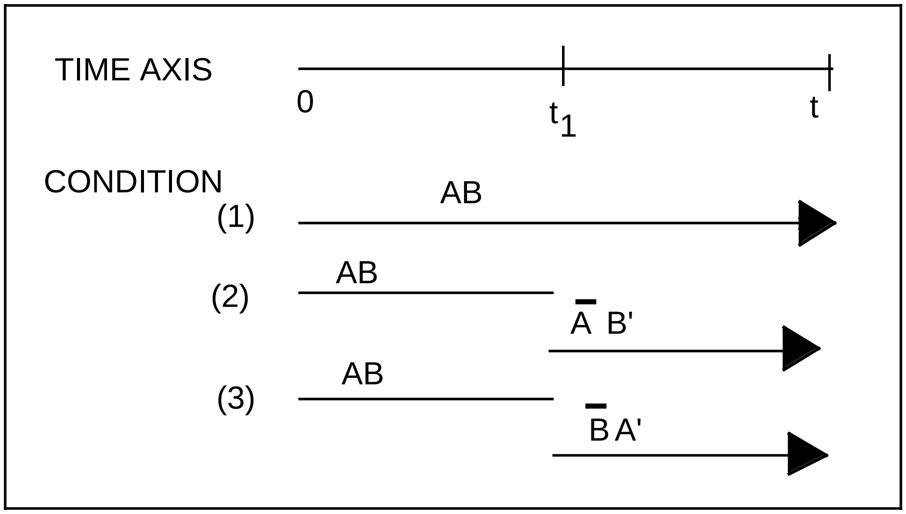


FIGURE 7.5-19: SUCCESS COMBINATIONS IN TWO-ELEMENT LOAD-SHARING CASE

7.5.5.5 Optimum Allocation of Redundancy

Decision and switching devices may fail to switch when required or may operate inadvertently. However, these devices are usually necessary for redundancy, and increasing the number of redundant elements increases the number of switching devices. If such devices are completely reliable, redundancy is most effective at lower system levels. If switching devices are not failure free, the problem of increasing system reliability through redundancy becomes one of choosing an optimum level at which to replicate elements.

Since cost, weight, and complexity factors are always involved, the minimum amount of redundancy that will produce the desired reliability should be used. Thus efforts should be concentrated on those parts of the system which are the major causes of system unreliability.

As an example, assume that we have two elements, A and B, with reliabilities over a certain time period of 0.95 and 0.50, respectively. If A and B are joined to form a series nonredundant circuit, its reliability is

$$R = (0.95)(0.50) = 0.475$$

If we duplicate each element, as in Figure 7.5-20a,

$$R_1 = [1 - (0.50)^2][1 - (0.05)^2] = 0.748$$

Duplicating Element B only, as in Figure 7.5-20b,

$$R_2 = 0.95\ [1 - (0.50)^2] = 0.712$$

Obviously, duplicating Element A contributes little to increasing reliability.

Triplication of B gives the configuration shown in Figure 7.5-20c and

$$R_3 = 0.95\ [1 - (0.5)^3] = 0.831$$

R_3 gives a 75% increase in original circuit reliability as compared to the 58% increase of R_1

If complexity is the limiting factor, duplicating systems is generally preferred to duplicating elements, especially if switching devices are necessary. If another series path is added in parallel, we have the configuration in Figure 7.5-20d, and

$$R_4 = 1 - (1 - .475)^2 = 0.724$$

R_4 is only slightly less than R_1. If switches are necessary for each redundant element, R_4 may be the best configuration. A careful analysis of the effect of each element and switch on system reliability is a necessary prerequisite for proper redundancy application.

7.5.6 Reliability Analysis Using Markov Modeling

7.5.6.1 Introduction

Markov Modeling is a reliability analysis tool which in the past few years has become the most prominent method of computing the reliability (or unreliability) of fault tolerant systems. It is an extremely flexible tool which can be used to predict the reliability of in-flight critical digital electronic systems. It has been used on a number of digital electronic engine controls to compute the probability of events such as aircraft loss due to control system failures, mission aborts, in-flight shut-down of engines, overspeeding of engines and inadvertent thrust reversal. Markov modeling offers many advantages over other reliability modeling techniques, some of which are:

(1) Simplistic modeling approach: The models are simple to generate although they require a more complicated mathematical approach. This is not a problem, however, because the mathematics are well suited for the digital computer.

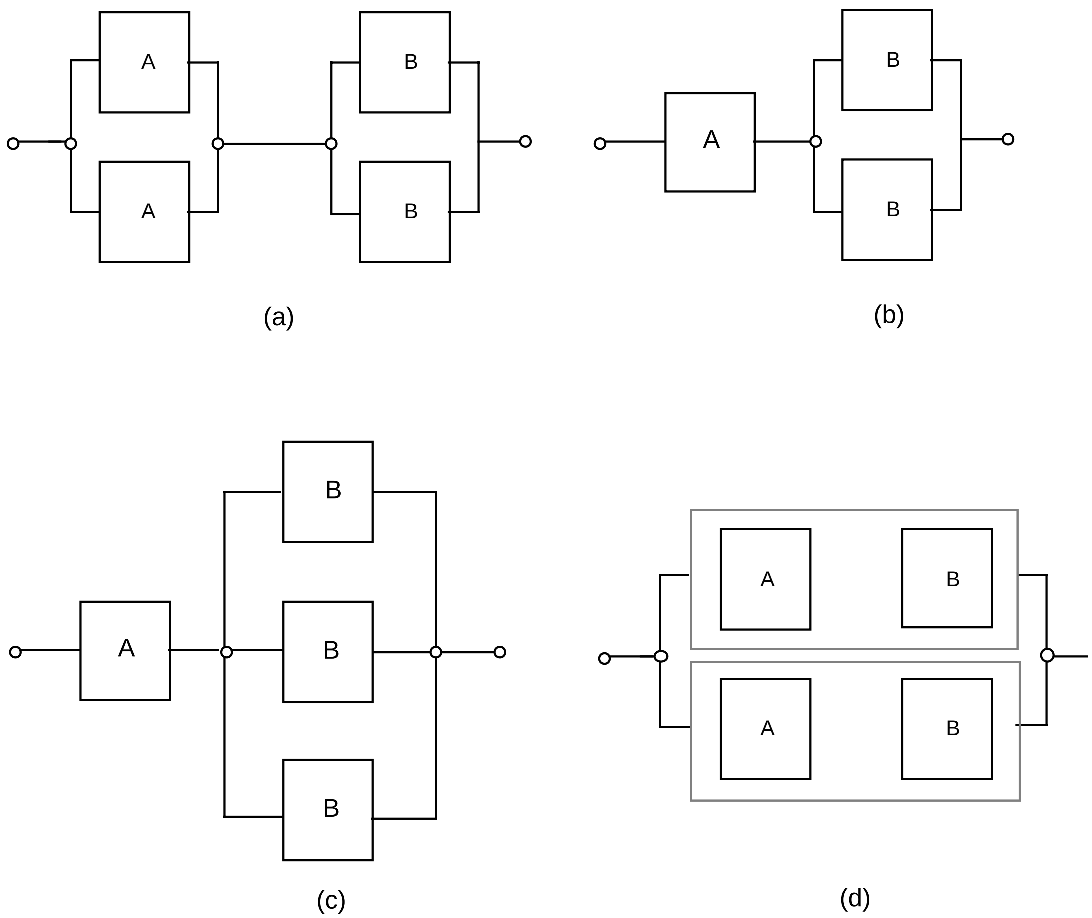


FIGURE 7.5-20: POSSIBLE REDUNDANT CONFIGURATIONS RESULTING FROM ALLOCATION STUDY

(2) Redundancy management techniques: System reconfiguration required by failures is easily incorporated in the model.

(3) Coverage: Covered and uncovered failures of components are mutually exclusive event. These are not easily modeled using classical techniques, but are readily handled by the Markov mathematics.

(4) Complex systems: Many simplifying techniques exist which allow the modeling of complex systems.

(5) Sequenced events: Many times the analyst is interested in computing the probability of an event which is the result of a certain sequence of sub-events. As an example, the probability of an engine overspeed might be desired. This is usually the result of two events, these being loss of overspeed protection and an uncommanded high fuel flow. These must necessarily occur in that order. For if the uncommanded high fuel flow

precedes the overspeed protection failure, an engine shutdown occurs rather than an overspeed. While these types of problems do not lend themselves well to classical techniques, they are easily handled using Markov modeling.

7.5.6.2 Markov Theory

Markov modeling can be applied to systems which vary discretely or continuously with respect to time and space. In reliability we are generally concerned with continuous time, discrete state models. These systems are characterized by randomly varying stochastic processes. Stochastic processes must have two important properties in order to model them with the Markov approach[5].

These are:

(1) The process must be memoryless

(2) The process must be stationary

A memoryless system is characterized by the fact that the future state of the system depends only on its present state. A stationary system is one in which the probabilities which govern the transitions from state to state remain constant with time. In other words, the probability of transitioning from some state i to another state j is the same regardless of the point in time the transition occurs. The states of the model are defined by system element failures. The transitional probabilities between states are a function of the failure rates of the various system elements. A set of first-order differential equations are developed by describing the probability of being in each state in terms of the transitional probabilities from and to each state. The number of first-order differential equations will equal the number of states of the model. The mathematical problem becomes one of solving the following equation:

$$\dot{\underline{P}} = [A]\underline{P}$$

where $\dot{\underline{P}}$ and $\underline{P}$ are n x 1 column vectors, [A] is an n x n matrix and n is the number of states in the system. The solution of this equation is:

$$\underline{P} = \exp[A]t \bullet \underline{P(0)}$$

where exp[A]t is an n x n matrix and $\underline{P}(0)$ is the initial probability vector describing the initial state of the system. Two methods which are particularly well suited for the digital computer for computing the matrix exp[A]t are the infinite series method and the eigenvalue/eigenvector method. Figure 7.5-21 presents a flow chart which illustrates the procedure used to develop a Markov model.

---

5 Extensions of the theory to other processes exist but are beyond the scope of this handbook.

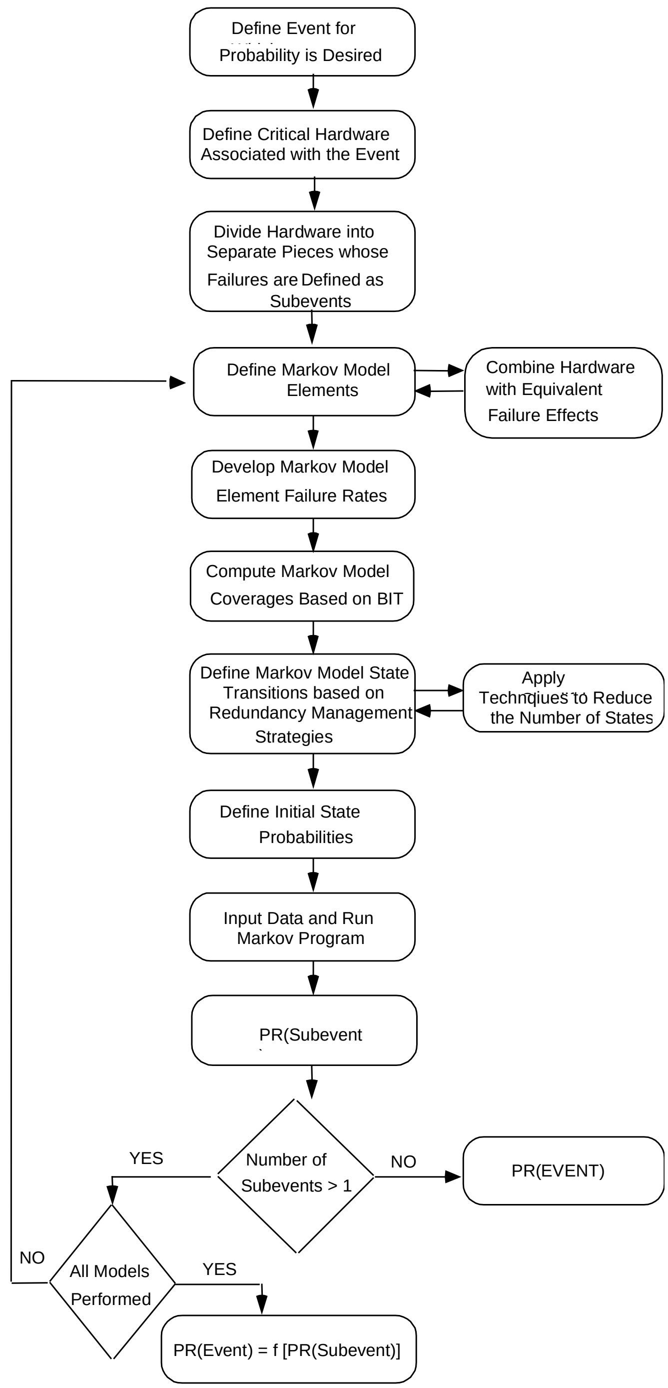


FIGURE 7.5-21: MARKOV MODELING PROCESS

7.5.6.3 Development of the Markov Model Equation

In order to illustrate how the Markov model equations are developed, assume we have a system which is made up of two elements. Each element has two mutually exclusive states - a good and failed. The states of the model are generated based on the elements being in one of these two states. The probabilities that cause transition from state to state are a function of the element failure rates. An element with constant failure rate (1) has a transitional probability which is approximated by $\lambda \cdot \Delta t$. The probability of more than one element failure in Δt is considered negligible. A flow diagram of the two element problem mentioned above is presented in Figure 7.5-22.

We develop the Markov differential equation by describing the probability of being in each of the states at time $t + \Delta t$ as a function of the state of the system at time t. The probability of being in state one at some time $t + \Delta t$ is equal to the probability of being in state on at time t and not transitioning out during Δt. This can be written as:

$$P1(t + \Delta t) = P1(t) \cdot [1 - (\lambda_1 + \lambda_2) \cdot \Delta t]$$

The probability of being in state two at time $t + \Delta t$ is equal to the probability of being in state one at time t and transitioning to state two in Δt plus the probability of being in state two at time t and **not** transitioning out during Δt. This can be written as:

$$P2(t + \Delta t) = P1(t) \cdot \lambda_1 \cdot \Delta t + P2(t)(1 - \lambda_2 \cdot \Delta t)$$

The other state probabilities are generated in the same manner resulting in the following equations:

$$P1(t + \Delta t) = P1(t) \cdot [1 - (\lambda_1 + \lambda_2) \cdot \Delta t]$$
$$P2(t + \Delta t) = P1(t) \cdot \lambda_1 \cdot \Delta t + P2(t)(1 - \lambda_2 \cdot \Delta t)$$
$$P3(t + \Delta t) = P1(t) \cdot \lambda_2 \cdot \Delta t + P3(t)(1 - \lambda_1 \cdot \Delta t)$$
$$P4(t + \Delta t) = P2(t) \cdot \lambda_2 \cdot \Delta t + P3(t) \cdot \lambda_1 \cdot \Delta t + P4(t)$$

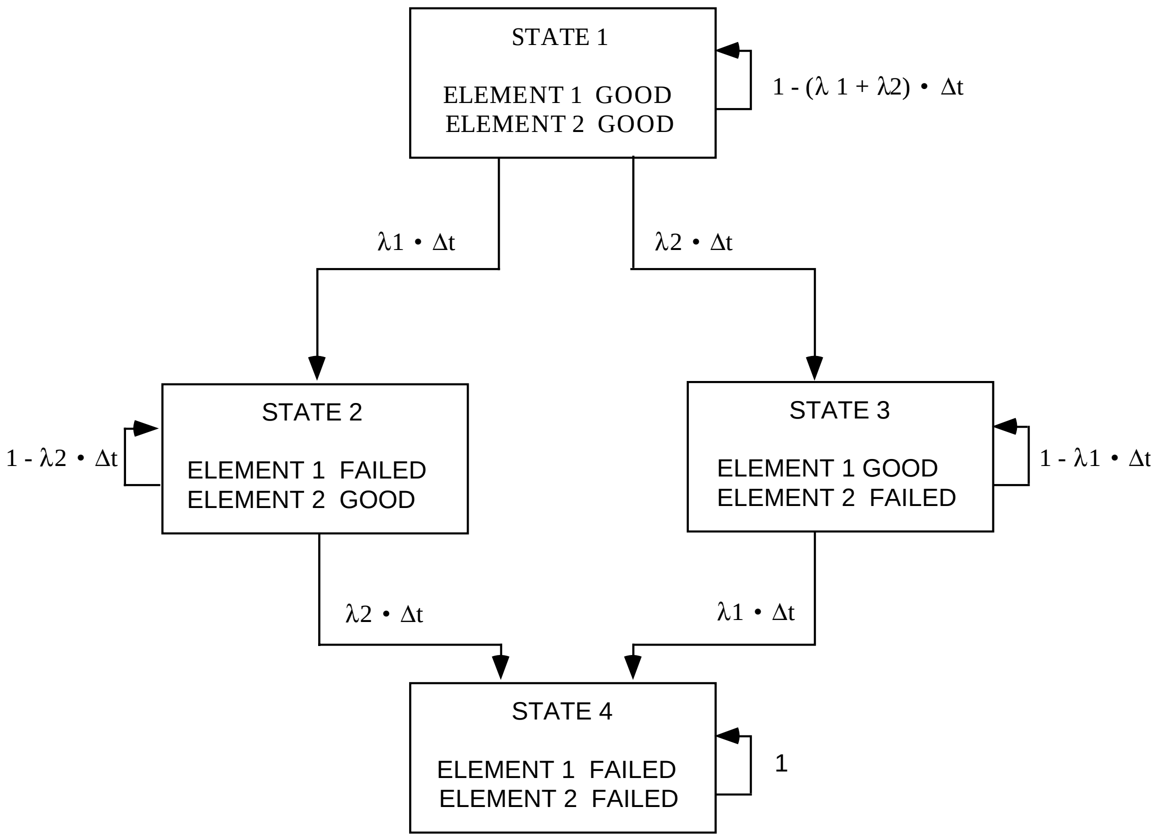


FIGURE 7.5-22: MARKOV FLOW DIAGRAM

Rearranging:

$$[P1(t + \Delta t) - P1(t)]/\Delta t = -(\lambda_1 + \lambda_2) \bullet P1(t)$$
$$[P2(t + \Delta t) - P2(t)]/\Delta t = \lambda_1 \bullet P1(t) - \lambda_2 \bullet P2(t)$$
$$[P3(t + \Delta t) - P3(t)]/\Delta t = \lambda_2 \bullet P1(t) - \lambda_1 \bullet P3(t)$$
$$[P4(t + \Delta t) - P4(t)]/\Delta t = \lambda_2 \bullet P2(t) + \lambda_1 \bullet P3(t)$$

Taking the limit as $\Delta t \rightarrow 0$:

$$dP1(t)/dt = -(\lambda_1 + \lambda_2) \bullet P1(t)$$
$$dP2(t)/dt = \lambda_1 \bullet P1(t) - \lambda_2 \bullet P2(t)$$
$$dP3(t)/dt = \lambda_2 \bullet P1(t) - \lambda_1 \bullet P3(t)$$
$$dP4(t)/dt = \lambda_2 \bullet P2(t) + \lambda_1 \bullet P3(t)$$

In matrix form this becomes:

$$\begin{vmatrix} dP1(t)/dt \\ dP2(t)/dt \\ dP3(t)/dt \\ dP4(t)/dt \end{vmatrix} = \begin{vmatrix} -(\lambda_1 + \lambda_2) & 0 & 0 & 0 \\ \lambda_1 & -\lambda_2 & 0 & 0 \\ \lambda_2 & 0 & -\lambda_1 & 0 \\ 0 & \lambda_2 & \lambda_1 & 0 \end{vmatrix} \bullet \begin{vmatrix} P1(t) \\ P2(t) \\ P3(t) \\ P4(t) \end{vmatrix}$$

or $\dot{P} = [A] \bullet \underline{P}$ where [A] is defined as the state transition matrix. The important thing to note here is that the analyst need only generate the states and the transitions between states as defined by the element failure rates. This information can then be inputted to the computer in a form which allows it to set up the state transition matrix and compute the state probabilities using matrix mathematics.

7.5.6.4 Markov Model Reduction Techniques

Since the Markov modeling approach can generate all the possible states of a system, the number of states can be extremely large even for a relatively small number of Markov elements. Therefore it become imperative for the analyst using the Markov modeling approach to become familiar with the reduction techniques that can be applied to reduce the number of states significantly while maintaining the accuracy of the model. As an example, if we assume a system contains 10 elements, each of which have two states (good and failed), the total number of possible states becomes:

$$\# \text{ STATES} = 2^N = 2^{10} = 1024$$

Furthermore, a system containing only 10 elements would be considered small when modeling digital electronic engine controls, for instance. Fortunately many simplification techniques exist which can be used alone, or in combination, to reduce the amount of states in the model.

One approach is to use the principle that states which represent multiple levels of failure contribute insignificantly to the overall probability of failure of the system. The model can be truncated at a certain failure level, combining all states below that level into one failed state. If for instance it is desired to truncate at the n^{th} level, all state transitions from n-1 level states would be directed to one n^{th} order failed state. Care should be taken, however, to make sure that this truncation does not have an overly conservative impact on the system.

Many elements have the same, or nearly the same, impact on system operation when failed. In this case the states which are the result of failure of these elements can be combined. As an example, assume we have a two channel engine control in dual active mode. By dual active mode we mean both channels are simultaneously in control. Let each channel have a failure rate λ. If one channel fails we have the ability to control with the other good channel. Because loss of either channel leads to the same effect (i.e., single channel control), the corresponding states can be combined. Figure 7.5-23 shows the Markov model for this system using no reduction technique and the equivalent model by combining States 2 and 3. Because the system impact was the same independent of what channel failed first, and because the channel failure rates were the same, we are able to reduce the number of states with no loss of accuracy. If this is not the case, assumptions have to be made as to what element failure caused transition to the new state so that a conservative approximation to the transitions out of the state can be made.

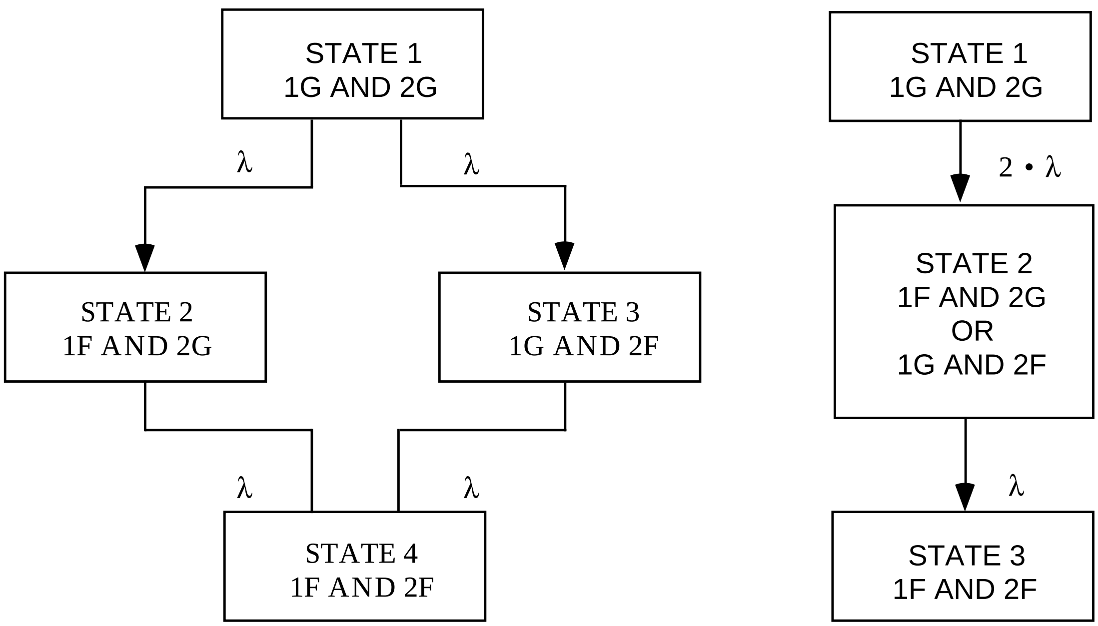


FIGURE 7.5-23: TWO CHANNEL EXAMPLE

Many times failure of a certain element causes loss of other elements in a system. An example would be loss of a power supply. In this case the analyst need not define transitions for the other lost element(s) because by definition they are also no longer part of the functioning system.

Another reduction technique involves dividing the top level event for which the probability of failure is desired into n sub-events, each of which is modeled separately. The probabilities for each sub-event are then combined to yield the probability of the top level event. If for instance the system being modeled has ten elements, we have a possible of 1024 total states. If the top level event containing these ten elements can be broken down into two sub-events, each containing five elements, the system can be described with two models each with a possible thirty-two states. If the top level event has probability P and the two sub-events have probabilities P1 and P2 respectively, the top level probability can be computed as P = f(P1,P2).

7.5.6.5 Application of Coverage to Markov Modeling

In redundant system modeling we generally consider three Markov element states - good, failed covered, and failed uncovered. Covered and uncovered markov element states are mutually exclusive meaning that an element cannot fail both covered and uncovered. System coverage is generally defined in terms of the conditional probability.

$$P[\text{detect, isolate, reconfigure} \mid \text{failure}]$$

When computing a coverage for Markov model elements we are concerned with that portion of the Markov element failure rate that is detectable and isolatable. Reconfiguration becomes a function of what resources are available at the time the failure occurs.

As an example of how coverage is used in the Markov model, we will return to the two channel dual active engineer control discussed previously. In this case if either channel fails **covered**, the other channel has the ability to take over full control. However, if either channel fails **uncovered**, system failure occurs. The Markov model for this example appears in Figure 7.5-24. Note that once state two is entered, no resources are available and both the covered and uncovered portions of the remaining channels failure rate are routed to system failure.

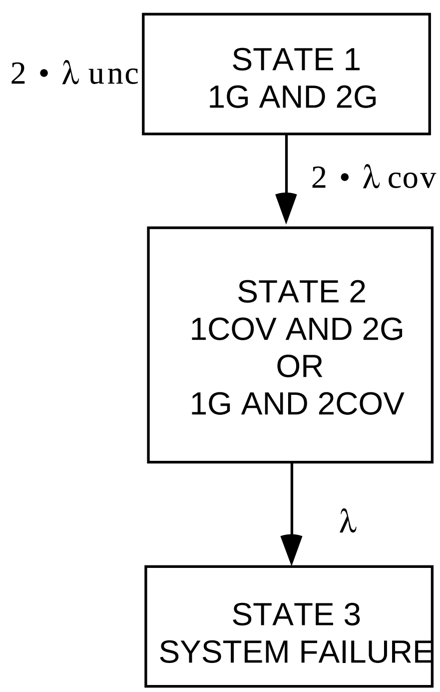


FIGURE 7.5-24: COVERAGE EXAMPLE

7.5.6.6 Markov Conclusions

Markov modeling is a powerful reliability analysis tool which allows the analyst to model complex fault tolerant systems that would otherwise be difficult to model with classical techniques. The Markov technique decreases the analysts task by reducing the problem from one of mathematical computation to one of state modeling. Many model reduction techniques exist which yield relatively simple models with insignificant impact on model accuracy.

An excellent resource document dealing with the Markov methodology is IEC 1165, Reference [25].

7.6 Environmental Design

7.6.1 Environmental Strength

A series of Engineering Design Handbooks deals explicitly, and in great detail, with environmental design problems (References [26] - [30]). Those handbooks should be consulted for specific information. This section will concentrate on some general environmental design considerations against specific environments. Many of the details on environmental prediction and specific design methods are in the previously mentioned documents.

To design inherently reliable equipment, the design engineer must take into account the environment in which the equipment is to operate, with relation to the ideal operating conditions for the elements which make up the equipment. Each item in a system has its own failure rate based upon the conditions under which it operates.

MIL-STD-210 (Climatic Extremes for Military Equipment) establishes climatic design criteria for material intended for worldwide usage. It provides design conditions for land, sea, and air in which equipment will be required to operate (or be stored). The standard breaks down climate extremes into three categories - ground, naval surface and air, and worldwide air. For these three categories, the climatic conditions for which values and factors are presented include temperature, humidity, precipitation, atmospheric pressure, and many others. MIL- STD-210 is the baseline document from which climatic environmental conditions can be derived. Operating conditions may vary considerably from climatic conditions due to changes caused by system operation, e.g., equipment heating. The designer may have to address climatic problems using special parts. Such parts may need to operate at low temperature, incorporate pre-heating arrangements, utilize temperature-tolerant lubricants, or incorporate other methods of adjusting for climatic conditions.

7.6.2 Designing for the Environment

Since the reliability achieved in actual use depends on the operating conditions that are encountered during the entire life of the equipment, it is important that such conditions are accurately identified at the beginning of the design process. Environmental factors which exert a strong influence on equipment reliability are included in Table 7.6-1, which is a checklist for environmental coverage.

TABLE 7.6-1: ENVIRONMENTAL COVERAGE CHECKLIST (TYPICAL)

NATURAL		INDUCED
Clouds	Rain	Acceleration
Fog	Salt Spray	Electromagnetic, Laser
Freezing Rain	Sand and Dust	Electrostatic, Lightning
Frost	Sleet	Explosion
Fungus	Snow	Icing
Geomagnetism	Hail	Radiation, Electromagnetic
Gravity, Low	Ice	Radiation, Nuclear
Temperature, High	Wind	Shock
Temperature, Low		Temperature, High, Aero. Heating
Humidity, High		Temperature, Low, Aero. Cooling
Humidity, Low		Turbulence
Ionized Gases		Vapor Trails
Lightning		Vibration, Mechanical
Meteoroids		Vibration, Acoustic
Pollution, Air		
Pressure, High		
Pressure, Low		
Radiation, Cosmic, Solar		
Radiation, Electromagnetic		

Concurrent (combined) environments are usually more detrimental to reliability than the effects of any single environment. Design and test criteria must consider both single and combined environments. Figure 7.6-1 illustrates the effects of combined environments (typical) in a matrix relationship. It shows the combinations where the total effect is more damaging than the cumulative effect of each environment acting independently. Concurrent environments may include a combination such as temperature, humidity, altitude, shock, and vibration. Table 7.6-2 provides reliability considerations for pairs of environmental factors.

The impact of each of the environmental factors anticipated during the life cycle of equipment on the operational and reliability characteristics of the materials and parts comprising the equipment being designed must be determined. Packaging techniques that afford the necessary protection against such degrading factors must also be identified.

1 *Combine to intensify mechanical deterioration
2 *Combine to intensify operational deterioration
3 Interdependent (one environment dependent on other
4 Coexist with no significant combined effect
5 Weakened effect (one environment weakens effects of others)
6 Incompatible
7 Unknown (unlikely combinattion or indeterminate combined effect)
☐ (Blank) Combination not considered (independent environments)
* A minus sign (-) following number indicates that intensification through combinataion is weak or doubtful

Natural: Earth and Lower Atmosphere (Clouds through Wind); Hyper and Space (Gravity, Low through Radiation, Van Allen)
Induced: Acceleration through Vibration, Acoustic

	Clouds	Fog	Freezing Rain	Frost	Fungus	Geomagnetism	Hail	Humidity	Lightning	Pollution, Air	Rain	Salt Spray	Sand and Dust
Clouds													
Fog	5												
Freezing Rain	3												
Frost	5												
Fungus													
Geomagnetism													
Hail	2,1	6											
Humidity	3	3	3	3	3								
Lightning	3	6				2-	2,1						
Pollution, Air	4	1						1					
Rain	3	2		6	3	6	1,2	3	4	4			
Salt Spray		1-	5					5			5		
Sand and Dust		1-						1-		1-		1-	
Sleet	3	7					6	3	6				
Snow	3	7						3	6		2-		
Radiation, Solar		5		5	5			1-			6	1	1-
Temperature, High	5	5	6	6	3			1				1	5
Temperature, Low		3	3	3	5			1			1-	5	1-
Wind	2-	5	7				1,2	7	1,2	3	1,2	1	1,2
Gravity, Low													
Ionized Glass					5			1					
Meteoroids													
Pressure, Low					7			1					
Radiation, Cosmic						3							
Radiation, Electromag.						3							
Radiation, Van Allen						3							
Acceleration								7					
Explosion								1-					
Icing	3	1-	2,1								1-	5	
Radiation, Nuclear													
Shock													
Temp., High Aero. Heat													
Temp., Low, Cryogenic													
Turbulence	2-	6					2,1		2,1		2		
Vapor Trails	5		6				6	3	6	7	6		
Vibration, Mechanical								1,2				1,2	1,2
Vibration, Acoustic													

	Sleet	Snow	Radiation, Solar	Temperature, High	Temperature, Low	Wind	Gravity, Low	Ionized Gases	Meteoroids	Pressure, Low	Radiation, Cosmic	Radiation, Electromag.	Radiation, Van Allen
Clouds													
Fog													
Freezing Rain													
Frost													
Fungus													
Geomagnetism													
Hail													
Humidity													
Lightning													
Pollution, Air													
Rain													
Salt Spray													
Sand and Dust													
Sleet													
Snow													
Radiation, Solar													
Temperature, High	6	6	1										
Temperature, Low	3	3	5	6									
Wind	1,2	1,2	1	1-	2,1								
Gravity, Low			4	4	4								
Ionized Glass			1	1	5		4						
Meteoroids													
Pressure, Low			1-	1	1	5	4	3					
Radiation, Cosmic			3				7	1-					
Radiation, Electromag.							7	1-			3		
Radiation, Van Allen			3				4				3		
Acceleration			7		1								
Explosion			7	7	7					1-			
Icing			5	6	3	2-							
Radiation, Nuclear			1-	1-							1	7	7
Shock				1						1-			
Temp., High Aero. Heat				6									
Temp., Low, Cryogenic				6									
Turbulence						2,1							
Vapor Trails	6	6	4	6	3	5							
Vibration, Mechanical			1	1	1					1-			
Vibration, Acoustic													

	Acceleration	Explosion	Icing	Radiation, Nuclear	Shock	Temp., High Aero. Heat	Temp., Low, Cryogenic	Turbulence	Vapor Trails	Vibration, Mechanical	Vibration, Acoustic
Clouds											
Fog											
Freezing Rain											
Frost											
Fungus											
Geomagnetism											
Hail											
Humidity											
Lightning											
Pollution, Air											
Rain											
Salt Spray											
Sand and Dust											
Sleet											
Snow											
Radiation, Solar											
Temperature, High											
Temperature, Low											
Wind											
Gravity, Low											
Ionized Glass											
Meteoroids											
Pressure, Low											
Radiation, Cosmic											
Radiation, Electromag.											
Radiation, Van Allen											
Acceleration											
Explosion											
Icing											
Radiation, Nuclear	7	7	7	7							
Shock				7							
Temp., High Aero. Heat	6	7	1								
Temp., Low, Cryogenic				7	1	6					
Turbulence			2,1			7					
Vapor Trails											
Vibration, Mechanical	1			5		1	1	1			
Vibration, Acoustic				7		1	1-				

FIGURE 7.6-1: EFFECTS OF COMBINED ENVIRONMENTS

TABLE 7.6-2: VARIOUS ENVIRONMENTAL PAIRS

HIGH TEMPERATURE AND HUMIDITY	HIGH TEMPERATURE AND LOW PRESSURE	HIGH TEMPERATURE AND SALT SPRAY
High temperature tends to increase the rate of moisture penetration. The general deterioration effects of humidity are increased by high temperatures.	Each of these environments depends on the other. For example, as pressure decreases, outgassing of constituents of materials increases, and as temperature	

TABLE 7.6-2: VARIOUS ENVIRONMENTAL PAIRS (CONT'D)

	LOW TEMPERATURE AND SAND AND DUST	LOW TEMPERATURE AND FUNGUS
LOW TEMPERATURE AND SHOCK AND VIBRATION	LOW TEMPERATURE AND ACCELERATION	LOW TEMPERATURE AND EXPLOSIVE ATMOSPHERE
Low temperature tends to intensify the effects of shock and vibration. It is, however, a consideration only at very low temperatures.	This combination produces the same effect as low temperature and shock and vibration.	Temperature has very little effect on the ignition of an explosive atmosphere. It does however, affect the air-vapor ratio which is an important consideration.
LOW TEMPERATURE AND OZONE	HUMIDITY AND LOW PRESSURE	HUMIDITY AND SALT SPRAY
Ozone effects are reduced at lower temperatures, but ozone concentration increases with lower temperatures.	Humidity increases the effects of low pressure, particularly in relation to electronic or electrical equipment. However, the actual effectiveness of this combination is determined largely by the temperature.	High humidity may dilute the salt concentration, but it has no bearing on the corrosive action of the salt.
HUMIDITY AND FUNGUS	HUMIDITY AND SAND AND DUST	HUMIDITY AND SOLAR RADIATION
Humidity helps the growth of fungus and microorganisms but adds nothing to their effects.	Sand and dust have a natural affinity for water and this combination increases deterioration.	Humidity intensifies the deteriorating effects of solar radiation on organic materials.
HUMIDITY AND VIBRATION	HUMIDITY AND SHOCK AND ACCELERATION	HUMIDITY AND EXPLOSIVE ATMOSPHERE
This combination tends to increase the rate of breakdown of electrical material.	The periods of shock and acceleration are considered too short for these environments to be affected by humidity.	Humidity has no effect on the ignition of an explosive atmosphere, but a high humidity will reduce the pressure of an explosion.
HUMIDITY AND OZONE	LOW PRESSURE AND SALT SPRAY	LOW PRESSURE AND SOLAR RADIATION
Ozone meets with moisture to form hydrogen peroxide, which has a greater deteriorating effect on plastics and elastomers than the additive effects of moisture and ozone.	This combination is not expected to occur.	This combination adds nothing to the overall effects.
	LOW PRESSURE AND FUNGUS	
	This combination adds nothing to the overall effects.	

TABLE 7.6-2: VARIOUS ENVIRONMENTAL PAIRS (CONT'D)

LOW PRESSURE AND SAND AND DUST	LOW PRESSURE AND VIBRATION	LOW PRESSURE AND SHOCK OR ACCELERATION
This combination only occurs in extreme storms during which small dust particles are carried to high altitudes.	This combination intensifies effects in all equipment categories but mostly with electronic and electrical equipment.	These combinations only become important at the hyper-environmental levels, in combination with high temperature.
LOW PRESSURE AND EXPLOSIVE ATMOSPHERE	SALT SPRAY AND FUNGUS	SALT SPRAY AND DUST
At low pressures, an electrical discharge is easier to develop, but the explosive atmosphere is harder to ignite.	This is considered an incompatible combination.	This will have a more corrosive effect than humidity and sand and dust.
SALT SPRAY AND VIBRATION	SALT SPRAY AND SHOCK OR ACCELERATION	SALT SPRAY AND EXPLOSIVE ATMOSPHERE
This will have a more corrosive effect than humidity and vibration.	These combinations will produce no added effects.	This is considered an incompatible combination.
SALT SPRAY AND OZONE	SOLAR RADIATION AND FUNGUS	SOLAR RADIATION AND SAND AND DUST
These environments have a more corrosive effect than humidity and ozone.	Because of the resulting heat from solar radiation, this combination probably produces the same combined effect as high temperature and fungus. Further, the ultraviolet in unfiltered radiation is an effective fungicide.	It is suspected that this combination will produce high temperatures.
SOLAR RADIATION AND OZONE	FUNGUS AND OZONE	SOLAR RADIATION AND SHOCK OR ACCELERATION
This combination increases the rate of oxidation of materials.	Fungus is destroyed by ozone.	These combinations produce no additional effects.
SOLAR RADIATION AND VIBRATION		SAND AND DUST AND VIBRATION
Under vibration conditions, solar radiation deteriorates plastics, elastomers, oils, etc., at a higher rate.		Vibration might possibly increase the wearing effects of sand and dust.
SHOCK AND VIBRATION	VIBRATION AND ACCELERATION	
This combination produces no added effects.	This combination produces increased effects when encountered with high temperatures and low pressures in the hyper-environmental ranges.	
SOLAR RADIATION AND EXPLOSIVE ATMOSPHERE		
This combination produces no added effects.		

In the environmental stress identification process that precedes the selection of environmental strength techniques, it is essential that the environments associated with all life intervals of the equipment be considered. These include not only the operational and maintenance environments, but also the pre-operational environments, when stresses imposed on the parts during manufacturing assembly, inspection, testing, shipping, and installation may have a significant impact on the eventual reliability of the equipment. Stresses imposed during the pre-operational phase are often overlooked. They may, however, represent a particularly harsh environment which the equipment must withstand. Often, the environments to which a system is exposed during shipping and installation are more severe than those it will encounter under normal operating conditions. It is also probable that some of the environmental strength features of a system design address conditions that are encountered in the pre-operational phase, not in the operational phases.

Environmental stresses affect parts in different ways. Table 7.6-3 illustrates the principal effects of typical environments on system parts and materials.

High temperatures impose a severe stress on most electronic items since they can cause not only catastrophic failure (such as melting of solder joints and burnout of solid-state devices), but also slow, progressive deterioration of performance levels due primarily to chemical degradation effects. It is often stated that excessive temperature is the primary cause of poor reliability in electronic equipment.

In electronic systems design, great emphasis is placed on small size and high part densities. This design philosophy generally requires a cooling system to provide a path of low thermal resistance from heat-producing elements to an ultimate heat sink of reasonably low temperature.

Solid-state parts are generally rated in terms of maximum junction temperatures. The thermal resistance from a junction to either the case or to free air is usually specified. The specification of maximum ambient temperature for which a part is suitable is generally not a sufficient method for part selection, since the surface temperatures of a particular part can be greatly influenced by heat radiation or heat conduction effects from nearby parts. These effects can lead to overheating, even though an ambient temperature rating appears not to be exceeded. It is preferable to specify thermal environment ratings such as equipment surface temperatures, thermal resistance paths associated with conduction, convection and radiation effects, and cooling provisions such as air temperature, pressure and velocity. In this manner, the true thermal state of the temperature-sensitive internal elements can be determined. Reliability improvement techniques for high temperature stress include the use of heat dissipation devices, cooling systems, thermal insulation, and heat withstanding materials.

TABLE 7.6-3: ENVIRONMENTAL EFFECTS

ENVIRONMENT	PRINCIPAL EFFECTS	TYPICAL FAILURES INDUCED
High temperature	Thermal aging: Oxidation Structural change Chemical reaction Softening, melting, and sublimination Viscosity reduction and evaporation Physical expansion	Insulation failure; Alteration of electrical properties Loss of lubrication properties. Structural failure; Increased mechanical stress; Increased wear on moving parts
Low temperature	Increased viscosity and solidification Ice formation Embrittlement Physical Contraction	Loss of lubrication properties. Alteration of electrical properties. Loss of mechanical strength; cracking, fracture structural failure; increased wear on moving parts.
High relative humidity	Moisture absorption Chemical reaction Corrosion Electrolysis	Swelling, rupture of container; physical breakdown; loss of electrical strength. Loss of mechanical strength; interference with function; loss of electrical properties; increased conductivity of insulators.
Low relative humidity	Desiccation Embrittlement Granulation	Loss of mechanical strength; Structural collapse; Alteration of electrical properties, "dusting".
High pressure	Compression	Structural collapse; Penetration of sealing; Interference with function.

TABLE 7.6-3: ENVIRONMENTAL EFFECTS (CONT'D)

ENVIRONMENT	PRINCIPAL EFFECTS	TYPICAL FAILURES INDUCED
Low pressure	Expansion Outgassing Reduced dielectric strength of air	Fracture of container; Explosive expansion. Alteration of electrical properties; Loss of mechanical strength. Insulation breakdown and arc-over; Corona and ozone formation.
Solar radiation	Actinic and physio-chemical reactions: Embrittlement	Surface deterioration; Alteration of electrical properties; Discoloration of materials; Ozone formation.
Sand and dust	Abrasion Clogging	Increased wear; Interference with function; Alteration of electrical properties.
Salt Spray	Chemical reactions: Corrosion Electrolysis	Increased wear. Loss of mechanical strength; Alteration of electrical properties; Interference with function. Surface deterioration; Structural weakening; Increased conductivity.
Wind	Force application Deposition of materials Heat loss (low velocity) Heat gain (high velocity)	Structural collapse; Interference with function; Loss of mechanical strength; Mechanical interference and clogging; Abrasion accelerated. Accelerates low-temperature effects. Accelerates high temperature effects.
Rain	Physical stress Water absorption and immersion Erosion Corrosion	Structural collapse. Increase in weight; Aids heat removal; Electrical failure; Structural weakening. Removes protective coatings; Structural weakening; Surface deterioration. Enhances chemical reactions.

TABLE 7.6-3: ENVIRONMENTAL EFFECTS (CONT'D)

ENVIRONMENT	PRINCIPAL EFFECTS	TYPICAL FAILURES INDUCED
Temperature Shock	Mechanical stress	Structural collapse or weakening; Seal damage
High-speed particles (nuclear irradiation)	Heating Transmutation and ionization	Thermal aging; Oxidation. Alteration of chemical physical, and electrical properties; Production of gases and secondary particles.
Zero gravity	Mechanical stress Absence of convection cooling	Interruption of gravity-dependent functions. Aggravation of high-temperature effects.
Ozone	Chemical reactions: Crazing. cracking Embrittlement Granulation Reduced dielectric strength of air	Rapid oxidation; Alteration of electrical properties; Loss of mechanical strength; Interference with function. Insulation breakdown and arc-over.
Explosive decompression	Severe mechanical stress	Rupture and cracking; Structural collapse.
Dissociated gases	Chemical reactions: Contamination Reduced dielectric strength	Alteration of physical and electrical properties. Insulation breakdown and arc-over.
Acceleration	Mechanical stress	Structural collapse.
Vibration	Mechanical stress Fatigue	Loss of mechanical strength; Interference with function; Increased wear. Structural collapse.
Magnetic fields	Induced magnetization	Interference with function; Alteration of electrical properties; Induced heating.

Low temperatures experienced by electronic equipment can also cause reliability problems. These problems are usually associated with mechanical elements of the system. They include mechanical stresses produced by differences in the coefficients of expansion (contraction) of metallic and nonmetallic materials, embrittlement of nonmetallic components, mechanical forces caused by freezing of entrapped moisture, stiffening of liquid constituents, etc. Typical examples include cracking of seams, binding of mechanical linkages, and excessive viscosity of lubricants. Reliability improvement techniques for low temperature stress include the use of heating devices, thermal insulation and cold-withstanding materials.

Additional stresses are produced when electronic equipment is exposed to sudden changes of temperature or rapidly changing temperature cycling conditions. These conditions generate large internal mechanical stresses in structural elements, particularly when dissimilar materials are involved. Effects of the thermal shock- induced stresses include cracking of seams, delamination, loss of hermeticity, leakage of fill gases, separation of encapsulating components from components and enclosure surface leading to the creation of voids, and distortion of support members.

A thermal shock test is generally specified to determine the integrity of solder joints since such a test creates large internal forces due to differential expansion effects. Such a test has also been found to be instrumental in creating segregation effects in solder alloys leading to the formulation of lead-rich zones which are susceptible to cracking effects.

Electronic equipment is often subjected to environmental shock and vibration both during normal use and testing. Such environments can cause physical damage to parts and structural members when resulting deflections produce mechanical stresses which exceed the allowable working stress of the constituent parts.

The natural frequencies of items are important parameters which must be considered in the design process since a resonant condition can be produced if a natural frequency is within the vibration frequency range. The resonance condition will greatly amplify the deflection of the next higher level of assembly and may increase stresses beyond the safe limit.

The vibration environment can be particularly severe for electrical connectors, since it may cause relative motion between members of the connector. This motion, in combination with other environmental stresses, can produce fret corrosion. This generates wear debris and causes large variations in contact resistance. Reliability improvement techniques for vibration stress include the use of stiffening, control of resonance, and reduced freedom of movement.

Humidity and salt-air environments degrade equipment performance since they promote corrosion effects in metallic components. They can also foster the creation of galvanic cells, particularly when dissimilar metals are in contact. Another deleterious effect of humidity and salt air atmospheres is the formation of surface films on nonmetallic parts. These films cause leakage paths and degrade the insulation and dielectric properties of these materials. Absorption of moisture by insulating materials can also cause a significant increase in volume conductivity and the dissipation factor of materials so affected. Reliability improvement techniques for humidity and salt environments include the usage of hermetic sealing, moisture-resistant material, dehumidifiers, protective coatings, protective covers, and reduced use of dissimilar metals.

Electromagnetic and nuclear radiation can disrupt performance and, in some cases, cause permanent damage to exposed equipment. It is important, therefore, that such effects be considered in determining the required environmental strength required to achieve a specified reliability goal.

Electromagnetic radiation often produces interference and noise effects within electronic circuitry which can impair the functional performance of the system. Sources of these effects include corona discharges, lightning discharges, sparking, and arcing phenomena. These may be associated with high voltage transmission lines, ignition systems, brush-type motors, and even the equipment itself. Generally, the reduction of interference effects requires incorporation of filtering and shielding features, or the specification of less susceptible components and circuitry.

Nuclear radiation can cause permanent damage by alteration of the atomic or molecular structure of dielectric and semiconductor materials. High energy radiation can also cause ionization effects which degrade the insulation levels of dielectric materials. The mitigation of nuclear radiation effects typically involves the use of materials and parts possessing a higher degree of radiation resistance, and the incorporation of shielding and hardening techniques.

Each of the environmental factors experienced by an item during its life cycle must be considered in the design process. This ensures that the design will have adequate environmental strength.

Equipment failures have three convenient classifications:

(1) Poor design or incorrect choice of materials or components.

(2) Inadequate quality control which permits deviations from design specifications.

(3) Deterioration caused by environmental effects or influences.

Obviously, the first and third classes are related. Specifically, the careful selection of design and materials can extend item reliability by reducing or eliminating adverse environmental effects. The environment is neither forgiving nor understanding; it methodically surrounds and attacks every component of a system, and when a weak point exists, equipment reliability suffers. Design and reliability engineers, therefore, must understand the environment and its potential effects, and then must select designs or materials that counteract these effects or must provide methods to alter or control the environment within acceptable limits. Selecting designs or materials that withstand the environment has the advantage of not requiring extra components that also require environmental protection and add weight and cost.

In addition to the obvious environments of temperature, humidity, shock, and vibration, the design engineer will create environments by his choice of designs and materials. A gasket or seal, for example, under elevated temperatures or reduced pressures may release corrosive or degrading volatiles into the system. Teflon may release fluorine, and polyvinylchloride (PVC) may release chlorine. Certain solid rocket fuels are degraded into a jelly-like mass when exposed to aldehydes or ammonia, either of which come from a phenolic nozzle cone. These examples illustrate that internal environments designed into the system can seriously affect reliability.

7.6.3 Temperature Protection

Heat and cold are powerful agents of chemical and physical deterioration for two very simple, basic reasons

(1) The physical properties of almost all known materials are greatly modified by changes in temperature.

(2) The rate of almost all chemical reactions is markedly influenced by the temperature of the reactants. A familiar rule-of-thumb for chemical reactions (Reference [31]) is that the rate of many reactions doubles for every rise in temperature of 10°C; this is equivalent to an activation energy of about 0.6 eV.

High temperature degradation can be minimized by passive or active techniques. Passive techniques use natural heat sinks to remove heat, while active techniques use devices such as heat pumps or refrigeration units to create heat sinks. Such design measures as compartmentation, insulation of compartment walls, and intercompartment and intrawall air flow can be applied independently or in combination. Every system component should be studied from two viewpoints:

(1) Is a substitute available that will generate less heat?

(2) Can the component be located and positioned so that its heat has minimum effect on other components?

For a steady temperature, heat must be removed at the same rate at which it is generated. Thermal systems such as conduction cooling, forced convection, blowers, direct or indirect liquid cooling, direct vaporization or evaporation cooling, and radiation cooling must be capable of handling natural and induced heat sources. Passive sinks require some means of progressive heat transfer from intermediate sinks to ultimate sinks until the desired heat extraction has been achieved. Thus, when heat sources have been identified, and heat removal elements selected, they must be integrated into an overall heat removal system, so that heat is not merely redistributed within the system. Efficiently integrated heat removal techniques can significantly improve item reliability.

Besides the out-gassing of corrosive volatiles when subjected to heat, almost all known materials will expand or contract when their temperature is changed. This expansion and contraction causes problems with fit and sealing, and produces internal stresses. Local stress concentrations due to nonuniform temperature are especially damaging, because they can be so high. A familiar example is a hot water glass that shatters when immersed in cold water. Metal structures, when subjected to cyclic heating and cooling, may ultimately collapse due to the induced stresses and fatigue caused by flexing. The thermocouple effect between the junction of two dissimilar metals causes an electric current that may induce electrolytic corrosion. Plastics, natural fibers, leather, and both natural and synthetic rubber are all particularly sensitive to temperature extremes as evidenced by their brittleness at low temperatures and high degradation rates at high temperatures. Table 7.6-4 summarizes some of the basic precautions for achieving reliability at low temperatures. An always-present danger is that in compensating for one failure mode, the change will aggravate another failure mode.

The preferred method for evaluating the thermal performance of electronic equipment (with respect to reliability) is a parts stress analysis method. It can be used to determine the maximum safe temperatures for constituent parts. The parts stress analysis method for evaluating system thermal performance is based on a determination of the maximum allowable temperature for each part. This determination is to be consistent with the equipment reliability and the failure rate allocated to that part.

A reduction in the operating temperature of components is a primary method for improving reliability. Reduction in temperature generally can be achieved by providing a thermal design which reduces heat input to minimally achievable levels and provides low thermal resistance paths from heat producing elements to an ultimate heat sink of reasonably low temperature. The thermal design is often as important as the circuit design in obtaining the necessary performance and reliability characteristics of electronic equipment. Adequate thermal design maintains equipment and parts within their permissible operating temperature limits under operating conditions. Thermal design is an engineering discipline in itself, and will not be addressed in this section. An excellent document on thermal design is MIL-HDBK-251. It provides a very comprehensive review of the aspects of thermal design. Also, Chapter 9 of Reference [32] discusses the subject in some detail.

TABLE 7.6-4: LOW TEMPERATURE PROTECTION METHODS

EFFECT	PREVENTIVE MEASURES
Differential contraction	Careful selection of materials Provision of proper clearance between moving parts Use of spring tensioners and deeper pulleys for control cables Use of heavier material for skins
Lubrication stiffening	Proper choice of lubricants: • Use greases compounded from silicones, diesters or silicone diesters thickened with lithium stearate • Eliminate liquid lubricants wherever possible
Leaks in hydraulic systems	Use of low temperature sealing and packing compounds, such as silicone rubbers
Stiffening of hydraulic system	Use of proper low temperature hydraulic fluids
Ice Damage caused by freezing of collected water	Elimination of moisture by: • Provision of vents • Ample draining facilities • Eliminating moisture pockets • Suitable heating • Sealing • Desiccation of air
Degradation of material properties and component reliability	Careful selection of materials and components with satisfactory low temperature capabilities

7.6.4 Shock and Vibration Protection

Protection against mechanical abuse is generally achieved by using suitable packaging, mounting, and structural techniques. The reliability impact of mechanical protection techniques is generally singular in that these measures do or do not afford the required protection against the identified mechanical abuse stresses. In most cases, tradeoff situations between the level of protection and reliability improvements are not as pronounced as in the case of thermal protection. The one exception may be the case of fatigue damage, where the level of protection would have a significant impact on reliability if, in fact, fatigue were a primary failure mechanism in the normal life of the equipment.

Basic structural design techniques, such as proper component location and selection of suitable materials, can aid in protecting an item against failure caused by severe environmental stresses from shock or vibration.

There are two approaches that may be taken when shock or vibration are present; either isolate the equipment or build it to withstand the shock or vibration. The problem with isolation is that effective, simultaneous control of both shock and vibration is difficult. When only one or the other is present, special mountings are often used. Protective measures against shock and vibration stresses are generally determined by an analysis of the deflections and mechanical stresses produced by these environment factors. This generally involves the determination of natural frequencies and evaluation of the mechanical stresses within component and materials produced by the shock and vibration environment. If the mechanical stresses so produced are below the allowable safe working stress of the materials involved, no direct protection methods are required. If, on the other hand, the stresses exceed the safe levels, corrective measures such as stiffening, reduction of inertia and bending moment effects, and incorporation of further support members are indicated. If such approaches do not reduce the stresses below the safe levels, further reduction is usually possible by the use of shock absorbing mounts.

One factor, however, which is not often considered, is that the vibration of two adjacent components, or separately insulated subsystems, can cause a collision between them if maximum excursions and sympathetically induced vibrations are not evaluated by the designer. Another failure mode, fatigue (the tendency for a metal to break under cyclic stressing loads considerably below its tensile strength) is an area of reliability concern due to shock or vibration. Fatigue includes low cycle fatigue, acoustic fatigue, and fatigue under combined stresses. The interaction between multiaxial fatigue and other environmental factors such as temperature extremes, temperature fluctuations, and corrosion requires careful study. Stress-strength analysis of components and parameter variation analysis are particularly suited to these effects. Destruction testing methods are also very useful in this area. For one shot devices, several efficient nondestructive evaluation (NDE) methods are available - such as X-ray, neutron radiography, and dye penetrant - which can be used to locate fatigue cracks. Developing a simple design that is reliable is much better than elaborate fixes and subsequent testing to redesign for reliability.

In some cases, even though an item is properly isolated against shock and vibration damage, repetitive forces may loosen the fastening devices. Obviously, if the fastening devices loosen enough to permit additional movement, the device will be subjected to increased forces and may fail. Many specialized self-locking fasteners are commercially available, and fastener manufacturers usually will provide valuable assistance in selecting the best fastening methods.

An isolation system can be used at the source of the shock or vibration, in addition to isolating the protected component. The best results are obtained by using both methods. Damping devices are used to reduce peak oscillations, and special stabilizers employed when unstable configurations are involved. Typical examples of dampeners are viscous hysteresis, friction, and air damping. Vibration isolators commonly are identified by their construction and material used for the resilient elements (rubber, coil spring, woven metal mesh, etc.). Shock isolators differ from vibration isolators in that shock requires stiffer springs and a higher natural frequency for the resilient element. Some of the types of isolation mounting systems are underneath, over-and-under, and inclined isolators.

A specific component may initially appear to be sufficiently durable to withstand the anticipated shock or vibration forces without requiring isolation or insulation. However, this observation can be misleading since the attitude in which a part is mounted, its location relative to other

protective coatings; (4) providing rounded edges to allow uniform coating of protective material; (5) using materials resistant to moisture effects, fungus, corrosion, etc.; (6) hermetically sealing components, gaskets and other sealing devices; (7) impregnating or encapsulating materials with moisture resistant waxes, plastics, or varnishes; and (8) separating dissimilar metals, or materials that might combine or react in the presence of moisture, or of components that might damage protective coatings. The designer also must consider possible adverse effects caused by specific methods of protection. Hermetic sealing, gaskets, protective coatings, etc., may, for example, aggravate moisture difficulties by sealing moisture inside or contributing to condensation. The gasket materials must be evaluated carefully for out-gassing of corrosive volatiles or for incompatibility with adjoining surfaces or protective coatings.

MIL-HDBK-454 provides common requirements for electronic equipment related to corrosion protection (Guideline 15), dissimilar metals (Guideline 16), and moisture pockets (Guideline 31).

7.6.6 Sand and Dust Protection

Sand and dust primarily degrade equipment by:

(1) Abrasion leading to increased wear.

(2) Friction causing both increased wear and heat.

(3) Clogging of filters, small apertures, and delicate equipment.

Thus, equipment having moving parts requires particular care when designing for sand and dust protection. Sand and dust will abrade optical surfaces, either by impact when being carried by air, or by physical abrasion when the surfaces are improperly wiped during cleaning. Dust accumulations have an affinity for moisture and, when combined, may lead to corrosion or the growth of fungus.

In relatively dry regions, such as deserts, fine particles of dust and sand are readily agitated into suspension in the air, where they may persist for many hours, sometimes reaching heights of several thousand feet. Thus, even though there is virtually no wind present, the speeds of vehicles or vehicle-transported equipment through these dust clouds can cause surface abrasion by impact, in addition to the other adverse effects of the sand or dust.

Although dust commonly is considered to be fine, dry particles of earth, it also may include minute particles of metals, combustion products, solid chemical contaminants, etc. These other forms may provide direct corrosion or fungicidal effects on equipment, since this dust may be alkaline, acidic, or microbiological.

Since most equipment requires air circulation for cooling, removing moisture, or simply functioning, the question is not whether to allow dust to enter, but, rather, how much or what size dust can be tolerated. The problem becomes one of filtering the air to remove dust particles above a specific nominal size. The nature of filters, however, is such that (for a given working filter area), as the ability of the filter to stop increasingly smaller dust particles is increased, the flow of air or other fluid through the filter is decreased. Therefore, the filter surface area either must be increased, the flow of fluid through the filter decreased, or the allowable particle size increased. Interestingly enough, a study by R.V. Pavia (Reference [33]) showed that, for aircraft engines, the amount of wear was proportional to the weight of ingested dust, but that the wear produced by 100m dust is approximately half that caused by 15m dust. The 15m dust was the most destructive of all sizes tried.

Sand and dust protection, therefore, must be planned in conjunction with protective measures against other environmental factors. It is not practical, for example, to specify a protective coating against moisture if sand and dust will be present, unless the coating is carefully chosen to resist abrasion and erosion, or is self-healing.

7.6.7 Explosion Proofing

Protection against explosion is both a safety and reliability problem. An item that randomly exhibits explosive tendencies is one that has undesirable design characteristics and spectacular failure modes. Preventing this type of functional termination, therefore, requires extreme care in design and reliability analyses.

Explosion protection planning must be directed to three categories (not necessarily mutually exclusive) of equipment:

(1) Items containing materials susceptible to explosion.

(2) Components located near enough to cause the explosive items to explode.

(3) Equipment that might be damaged or rendered temporarily inoperative by overpressure, flying debris, or heat from an explosion.

The first category includes devices containing flammable gases or liquids, suspensions of dust in the air, hypergolic materials, compounds which spontaneously decompose in certain environments, equipment containing or subjected to high or low extremes of pressure (includes implosions), or any other systems capable of creating an explosive reaction. The second category is fairly obvious and includes many variations on methods for providing an energy pulse, a catalyst, or a specific condition that might trigger an explosion. A nonexplosive component, for example, could create a corrosive atmosphere, mechanical puncture, or frictional wear on the side of a vessel containing high pressure air and thereby cause the air container to explode. The third category encompasses practically everything, including items in the first two categories, since a potentially explosive device (such as a high pressure air tank) can be damaged or made to explode by the overpressure from another explosion. Thus, some reasoning must be applied when considering devices not defined by the first two categories. From a practical standpoint, explosion protection for items in the third category ought to be directed to equipment that might possibly be near explosions. The sides of an electronic maintenance van, for example, will be subjected to overpressures from exploding enemy artillery rounds. If designed for protection against anything but a direct hit, the van would be extremely difficult to transport. Thus, mobility (and size) and protections against blast are traded off. On the other end of the compromise scale, however, is the bad effect on the reliability of internal equipment when explosion protection is minimal or nonexistent.

The possibility of an explosive atmosphere leaking or circulating into other equipment compartments must be recognized. Lead acid batteries, for example, create hydrogen gas that, if

Another form of natural electromagnetic radiation is that associated with lightning. It is estimated that lightning strikes the earth about 100 times each second, each stroke releasing large bursts of electromagnetic energy which encircle the globe. Most of this energy is concentrated at the low frequency end of the electromagnetic spectrum with the maximum power level being concentrated at about 3 kHz.

Man-made electromagnetic energy is second in importance only to solar energy. Artificial electromagnetic radiators include those in power distribution systems, a multitude of uses in communications, and specialized detection and analytical applications. The development of lasers has introduced another intense source of electromagnetic radiation and, in military application, the electromagnetic pulse (EMP) associated with nuclear weapon detonations is of considerable importance.

The EMP spectrum is similar to that created by lightning with a maximum energy appearing at about 10 kHz but distributed with smaller amplitudes throughout a broad region of the frequency spectrum. EMP energy is of considerably greater magnitude than that observed in lightning and extends over a much larger area of the earth. Despite the similarities among EMP and lightning and other strong sources of electromagnetic energy, it cannot be assumed that protective measures consistent with these other electromagnetic radiation sources will protect material from the effects of EMP. The rapid rise time of the pulse associated with a nuclear detonation and the strength of the resulting pulse are unique.

A variety of effects of electromagnetic radiation on material are known, probably a number of effects are still unrecognized, and some of the effects on humans are poorly understood. Of course, one of the most important effects of electromagnetic radiation in the environment is the electromagnetic interference (EMI) it produces in the electromagnetic spectrum. Well known examples are called radio interference and radar clutter. Another important effect in the military is the interaction of electromagnetic radiation with electroexplosive devices used as detonators. Military as well as civilian explosives are provided with detonators that often depend on heating a small bridge wire to initiate the explosion. Absorbed electromagnetic radiation can accidentally activate such fuzes.

Protection against the effects of electromagnetic radiation has become a sophisticated engineering field of electromagnetic compatibility (EMC) design. The most direct approach to protection is, in most cases, to avoid the limited region in which high radiation levels are found. When exposure cannot be avoided, shielding and filtering are important protective measures. In other cases material design changes or operating procedural changes must be instituted in order to provide protection.

7.6.9 Nuclear Radiation

Although a natural background level of nuclear radiation exists, the only nuclear radiation that is of interest to design engineers is that associated with manmade sources such as reactors, isotope power sources, and nuclear weapons. The most important of these sources is nuclear weapons, the effects of which can produce both transient and permanent damaging effects in a variety of material.

X-rays, gamma rays, and neutrons are the types of nuclear radiation of most concern. As opposed to charged nuclear particles, which also emanate from nuclear reactions, those forms of radiation listed have long ranges in the atmosphere; thus, they can irradiate and damage a variety of military material.

Among the nuclear effects that are of most concern are those called "transient radiation effects on electronics," often referred to as TREE. These transient effects are due primarily to the non-equilibrium free charged condition induced in material primarily by the ionization effects of gamma rays and X-rays. The separation of transient and permanent effects is made on the basis of the primary importance of the radiation effects. For example, a large current pulse may be produced by ionizing radiation, and this current pulse may result in permanent damage to a device by overheating. This effect is considered transient because the permanent damage results from overheating due to excess current rather than to direct radiation-induced material property change.

It is impossible to completely protect material items from nuclear radiation. The variety of effects produced by nuclear radiation for different materials and components makes protective design difficult. The procedure employed is to define a radiation hardness level in a given material item and to design and test the item to that level.

Nuclear radiation hardening is a large and complex field with a variety of specialists required to deal with different aspects of the problem. This subject is treated extensively in the Design Engineers' Nuclear Effects Manual (References [34] - [37]).

Table 7.6-5 represents a summary of environmental effects and design techniques to overcome them.

TABLE 7.6-5: ENVIRONMENTAL STRESSES IMPROVEMENT TECHNIQUES IN ELECTRONIC EQUIPMENT

ENVIRONMENTAL STRESS	EFFECTS	RELIABILITY IMPROVEMENT TECHNIQUES
High Temperature	Parameters of resistance, inductance, capacitance, power factor, dielectric constant, etc. will vary; insulation may soften; moving parts may jam due to expansion; finishes may blister; devices suffer thermal aging; oxidation and other chemical reactions are enhanced; viscosity reduction and evaporation of lubricants are problems; structural overloads may occur due to physical expansions.	Heat dissipation devices, cooling systems, thermal insulation, heat-withstanding materials.
Low Temperature	Plastics and rubber lose flexibility and become brittle; electrical constants vary; ice formation occurs when moisture is present; lubricants gel and increase viscosity; high heat losses; finishes may crack; structures may be overloaded due to physical contraction.	Heating devices, thermal insulation, cold-withstanding materials.
Thermal Shock	Materials may be instantaneously overstressed causing cracks and mechanical failure; electrical properties may be permanently altered. Crazing, delamination, ruptured seals.	Combination of techniques for high and low temperatures.
Shock	Mechanical structures may be overloaded causing weakening or collapse; items may be ripped from their mounts; mechanical functions may be impaired.	Strengthened members, reduced inertia and moments, shock absorbing mounts.
Vibration	Mechanical strength may deteriorate due to fatigue or overstress; electrical signals may be mechanically and erroneously modulated; materials and structures may be cracked, displaced, or shaken loose from mounts; mechanical functions may be impaired; finishes may be scoured by other surfaces; wear may be increased.	Stiffening, control of resonance.
Humidity	Penetrates porous substances and causes leakage paths between electrical conductors; causes oxidation which leads to corrosion; moisture causes swelling in materials such as gaskets; excessive loss of humidity causes embrittlement and granulation.	Hermetic sealing, moisture-resistant material, dehumidifiers, protective coatings.
Salt Atmosphere and Spray	Salt combined with water is a good conductor which can lower insulation resistance; causes galvanic corrosion of metals; chemical corrosion of metals is accelerated.	Nonmetal protective covers, reduced use of dissimilar metals in contact, hermetic sealing, dehumidifiers.

TABLE 7.6-5: ENVIRONMENTAL STRESSES IMPROVEMENT TECHNIQUES IN ELECTRONIC EQUIPMENT (CONT'D)

ENVIRONMENTAL STRESS	EFFECTS	RELIABILITY IMPROVEMENT TECHNIQUES
Electromagnetic Radiation	Causes spurious and erroneous signals from electrical and electronic equipment and components; may cause complete disruption of normal electrical and electronic equipment such as communication and measuring systems.	Shielding, material selection, part type selection.
Nuclear/Cosmic Radiation	Causes heating and thermal aging; can alter chemical, physical and electrical properties of materials; can produce gases and secondary radiation; can cause oxidation and discoloration of surfaces; damages electrical and electronic components especially semiconductors.	Shielding, component selection, nuclear hardening.
Sand and Dust	Finely finished surfaces are scratched and abraded; friction between surfaces may be increased; lubricants can be contaminated; clogging of orifices, etc.; materials may be worn, cracked, or chipped; abrasion, contaminates insulations, corona paths.	Air-filtering, hermetic sealing.
Low Pressure (High Altitude)	Structures such as containers, tanks, etc. are overstressed and can be exploded or fractured; seals may leak; air bubbles in materials may explode causing damage; internal heating may increase due to lack of cooling medium; insulations may suffer arcing and breakdown; ozone may be formed, outgasing is more likely.	Increased mechanical strength of containers, pressurization, alternate liquids (low volatility), improved insulation, improved heat transfer methods.

7.6.10 Avionics Integrity Program (AVIP)

Attention is increasingly being given to potential wear-out mechanisms associated with electronic equipments used in modern aircraft. Fatigue induced failures are recognized as a major portion of complex aircraft electronics system failures. Both vibration and temperature cycling are major contributors to the fatigue phenomenon. Of these two factors, temperature cycling by itself usually makes the more significant contribution, but the combined effect of the two factors acting in concert can be much greater than either one in isolation. Many of the metals and plastics used in complex avionics electronic systems have a high thermal coefficient of expansion (TCE) and also a high modulus of elasticity.

This combination of TCE mismatch and high modulus of elasticity can lead to high localized stress within various circuit elements which is exacerbated by any vibration contribution as the equipment is exposed to the full range of operational temperatures, and the shock and vibration effects incident to high performance aircraft operation. Some of the greatest thermal-expansion problem areas are in the electronic-component lead wires, solder joints, and printed-circuit-board materials. A great deal of attention is also being focused on the field of leadless chip carrier components and other surface-mounted devices with respect to preventing thermal-creep strain in the solder joints. A large number of different materials, with various different TCE's, are involved in the manufacture and assembly of these types of devices.

The Air Force Avionics Integrity Program (AVIP) as detailed in MIL-HDBK-87244, "Avionics Integrity Program," is specifically designed to address these types of problems. MIL-HDBK-87244 is a guidance handbook that emphasizes reliability by design including linkages to related systems engineering areas and experience from recent programs, program studies, related initiatives, and the latest concepts in integrated product development (IPD). AVIP is a logical and disciplined systems engineering approach to requirements definition, development, and production of avionics and other electronics products. It defines life, usage, environment and supportability requirements and process tasks to achieve required performance over the life of the electronics. AVIP employs basic physics, chemistry, and engineering principles to ensure an understanding of the influence of the usage and environments on materials and parts. It focuses on key production and process characteristics and control of variability of materials, parts and processes.

Incorporation of the AVIP philosophy into an integrated engineering and manufacturing process supports the following:

a. Understanding and defining:
   - product life requirements
   - how and where the equipment will be operated and maintained and the associated environments
   - user supportability and constraints

b. Understanding:
   - materials, processes and technologies to include properties, life limits and variabilities
   - the stresses imposed by the life cycle usage and environments

c. Establishing product and process design criteria tailored for the specific application

d. Identifying key product characteristics, design parameters, and production process characteristics and controlling their impact on cost, performance and supportability

e. Performing iterative analyses, simulations and trade studies to facilitate a balanced design solution

f. Conducting incremental developmental and qualification testing to verify analyses and design solutions

While all TCE and vibration induced stress concentrations cannot be eliminated in a typical electronic box, they can be minimized by the proper selection of parts and materials, and by the optimization of fabrication techniques, and geometrics. It is virtually impossible to analyze every material interface, every notch, hole, rivet, bend and screw in every section of an electronic box. Time and money will usually permit the examination and analysis of only major structural members in the system. Therefore, it is necessary to recognize and identify the most probable primary and secondary stress points during the preliminary design phase and to adequately address at least these concerns before the design effort culminates in the final manufactured product.

Each event and situation in the life cycle of an item can be related to environmental factors. These events and situations in the pre-operational, operational, and maintenance environments can be related to stresses, which the equipment must withstand to perform reliably.

7.6.10.1 MIL-STD-1670: Environmental Criteria and Guidelines for Air Launched Weapons

This standard:

(1) provides guidelines for determining the environmental conditions to which air-launched weapons will be subjected during the factory-to-target sequence (acceptance-to-end-of-useful-life profile).

(2) describes the tasks involved in applying the essential environmental design criteria in all phases of weapon development.

(3) provides the developer with background data on which to base environmental design and test requirements.

Table 7.6-6 provides a checklist for typical system use conditions. This checklist helps the designer or analyst to determine if environments have been adequately considered in the design for events and situations of an item's life cycle.

Table 7.6-7 shows somes effects of natural and induced environments during the various phases of the lifetime of an item. Table 7.6-8 rates the importance of the environmental factors for the various regions of the environment.

Starting with program initiation, the standard defines the requirements necessary for the development of information leading to full-scale development. Usage information needed for delineation and examination of all probable environments that could affect reliability or operational capability of an air-launched weapon includes the aircraft profile (launch-to-landing subphases), combat use tactics, store mix, etc., of the same nature as items shown in Table 7.6-6. For reference, Figure 1 through 28 of MIL-STD-1670 demonstrate a method of presenting

TABLE 7.6-6: SYSTEM USE CONDITIONS CHECKLIST (TYPICAL)

HANDLING/TRANSFER	TRANSPORTATION
- CONUS - Oversea Global Locality - Shore Station - NWS - Depot - Commercial Rework - Truck Transport - Rail Transport - Air Transport - Marine Transport - Carrier Onboard Delivery (COD) Aviation spares airlift - Underway Replenishment (UNREP) Vertical (Rotary Wing Aircraft) Cargo aircraft Ram tensioned highline (RTHL) High line transfer UNREP Ship - Launch Platform Aircraft carrier Expeditionary airlift Short Airfield for Tactical Support (SATS) Non-aviation ship (AGC, AK, CA, DE, DLGN,...) - Operational A/C handling, weapons handling Shipboard tie-down Land based tie-down Land based apron tie down Towing, Spotting Handling equipment Maintenance test Maintenance shop Avionics maintenance van A/C elevator vertical transit A/C cyclic turnaround Hangar/flight deck Mobile maintenance facility Flight deck-to-storage, storage-to-flight deck	- CONUS - Oversea Global Locality - Truck Transport Flatbed truck, exposed Van, Truck Trailer Containerized - Rail Transport Boxcar Flatcar Containerized - Air Transport Turboprop Propeller Jet - Marine Transport Ammunition Ship (AE) Fast Combat Support Ship (AOE) Cargo Ship (AK) Other auxiliary ship (AKL,...) Ship Hold Ship deck exposure - NWS - Shore station - Depot - Commercial rework - Packaging

TABLE 7.6-6: SYSTEM USE CONDITIONS CHECKLIST (TYPICAL) (CONT'D)

STORAGE	OPERATIONAL
- CONUS	- Natural environment
- Oversea global locality	- Induced environment
- Shore station	- Combined environment
- NWS	- Catapult launch
- Depot	- Arrested landing
- Commercial rework	- Store separation
- Igloo magazine	- Weapon release
- Uninsulated building	- Weapon delivery
- Roofed Structure - no sidewalls	- Weapon exhaust impingement
- Dump storage, exposed	- Weapon to weapon
- Dump storage, revetment	- Weapon to A/C
- Railroad siding	- A/C to weapon
- Store item	- A/C taxi
- Weapons item	- Jet exhaust backflow
- Explosives item	- Helicopter in-flight refueling (HIFR)
- Aircraft carrier	- Probe/drogue refueling
- Expeditionary airfield	- Buddy tanker
- SATS	- Jet blast (other aircraft)
- Non-aviation ship	- Jet blast (VTOL)
- Long term	- Mission mix
- Short term	- Store mix
- Interim	- Combat tactics
- Maintenance shop	- Operational deployment
- Avionics maintenance van	- A/C/Weapons maneuvers
- Mobile maintenance facility	- Equipment location
- Containerization	- Flight line operations
- Packaging	- Chance of environment encounter
	- Launch platform

TABLE 7.6-7: ENVIRONMENTAL ANALYSIS (INDUCED ENVIRONMENT)

Mission Regime	Storage	Transportation	Standby (Idle)	Standby	Use	Mainten-ance
Acceleration	NA	NA	NA	ox	ox	NA
Acoustic Vibration	NA	NA	ox	ox	ox	NA
Countermeasures	NA	NA	NA	NA	o x	NA
Enemy Action	x	x	x	ox	ox	NA
Explosive Atmosphere	NA	NA	NA	NA	NA	NA
Flutter	NA	NA	NA	NA	xo	NA
Ionized Gases	NA	NA	NA	NA	x	NA
Magnetic Fields	NA	NA	NA	o	o	o
Moisture	x	NA	x	ox	ox	ox
Nuclear Radiation	NA	NA	NA	x	ox	ox
Pressure	NA	NA	NA	NA	ox	NA
Shock	NA	x	NA	x	ox	x
Temperature	NA	NA	ox	ox	ox	NA
Temperature Shock	NA	NA	o x	ox	ox	NA
Vibration	NA	x	NA	x	x	x

Effects	Operational Effects	Mechanical/Physical Effects
o - Operational x - Mechanical/Physical xo - Either or both NA - Not Applicable	Function, mission, etc. influenced rather than direct physical alternation of item. Example: reduced visibility caused by fog	Direct physical alteration of item. Examples: corrosion, fracture, puncture, melting

TABLE 7.6-7: ENVIRONMENTAL ANALYSIS
(NATURAL ENVIRONMENT) (CONT'D)

Mission Regime	Storage	Transportation	Standby (Idle)	Standby	Use	Mainten-ance
Aridity	x	x	x	x	x	x
Asteroids	NA	NA	NA	NA	NA	NA
Birds	o	NA	NA	NA	xo	xo
Clouds	NA	NA	NA	o	o	NA
Cosmic Radiation	NA	NA	NA	NA	x	NA
Density, Air	NA	NA	NA	NA	o	NA
Dust, Interplanetary	NA	NA	NA	NA	NA	NA
Dust, Lunar	NA	NA	NA	NA	NA	NA
Dust, Terrestrial	xo	x	x	o	NA	x
Electricity, Atmospheric	NA	NA	NA	NA	xo	NA
Fog	x	NA	x	o	NA	o
Frost	x	NA	x	o	NA	x
Fungi	x	NA	NA	NA	NA	x
Geomagnetism	NA	NA	NA	NA	o	NA
Gravity	NA	NA	NA	NA	o	NA
Heat	x	x	x	xo	xo	x
Humidity	x	x	x	xo	xo	x
Icing	x	x	xo	xo	xo	xo
Ionized Gases	NA	NA	NA	NA	xo	NA
Insects	xo	xo	xo	xo	xo	xo
Lightning	x	x	x	xo	xo	xo
Meteoroids	NA	NA	NA	NA	NA	NA
Ozone	NA	NA	NA	NA	x	NA
Pollution, Air	x	x	x	NA	xo	NA
Pressure, Air	NA	NA	NA	o	xo	o
Rain	x	x	x	x	x	x

TABLE 7.6-8: ASSOCIATION OF FACTOR IMPORTANCE WITH REGION OF ENVIRONMENT

Region of the Environment	Terrain	Temperature	Humidity	Pressure	Solar radiation	Rain	Solar precipation	Fog	Wind	Salt	Ozone	Macrobiological organism	Microbiological organism	Atmospheric pollutants	Sand and Dust	Vibration	Shock	Acceleration	Acoustics	Electromagnetic radiation	Nuclear radiation
Storage	O	A	A	C	O	O	O	O	O	C	C	B	B	C	C	C	B	O	O	O	O
Transportation																					
Highway	A	B	B	O	C	B	B	B	C	O	O	O	O	O	C	A	A	C	O	O	O
Rail	A	B	B	O	C	C	B	C	C	O	O	O	O	O	C	A	A	C	O	O	O
Ship	O	C	B	O	C	C	O	B	C	B	O	C	O	O	O	B	C	C	O	O	O
Air	O	B	O	C	O	C	C	A	B	O	O	O	O	O	C	B	B	B	O	O	O
Operational Use																					
Cold regions	A	A	A	O	B	A	A	B	B	C	O	C	O	C	C	B	B	O	C	B	C
Hot-wet	A	A	A	C	B	A	O	O	B	B	O	C	A	O	O	B	B	O	C	B	C
Hot-dry	A	A	O	O	A	O	O	O	B	B	O	O	O	O	A	B	B	O	C	B	C
Temperature	A	A	A	C	B	A	B	B	B	B	C	C	B	C	B	B	B	O	C	B	O
Indoor Use	O	B	B	O	O	O	O	O	O	O	O	C	C	C	B	C	O	O	O	B	O
Operational Storage	O	A	A	O	B	B	B	O	C	B	C	B	B	C	C	O	C	O	O	C	C

A - Major Importance B - Important C - Minor O - Absent environmental criteria. The standard illustrates the major events, corresponding environments, and weapon status in a factory-to-target sequence. The air-launched weapon must perform as required in this sequence subsequent to, or while being subjected to, the established environments.

For more detailed information on environments, see References [26] - [30].

7.7 Human Performance Reliability

This section contains copyright-protected material for which permission has been granted for publication in this handbook.

7.7.1 Introduction

A short, but informative history of human performance reliability is given by Dougherty and Fragola [38]. Lee et. al. [39] developed an extensive, useful literature survey on the subject. The survey is sorted into the following categories:

(1) Human-Operator Reliability Prediction
(2) Human Reliability in Maintenance Work
(3) Data on Human Reliability Estimates
(4) Human-Machine System Effectiveness
(5) Allocation of Human-Machine Reliability
(6) Human Operator Models in Control Loop Systems
(7) Literature Survey and Overview
(8) Miscellany

The survey includes a convenient comparison of hardware and human reliability, see Table 7.7-1.

Another major comparative work is the survey of human reliability models performed by Meister [40]. Although somewhat dated now, the work provides excellent detailed narratives of the models extant in 1971 and, to a large extent, still applicable. Each of the many models are described and then evaluated with respect to comprehensiveness, applicability, and timing. Model characteristics are described in terms of objectivity and structure. Ten analytic methods for operability prediction, six simulation methods for operability prediction, and three maintainability prediction methods are described.

In a profile of the state of the art almost 20 years after Meister's work, Apostolakis et al. [41] reviewed human performance reliability analysis techniques, primarily with respect to those used in the nuclear power industry. Some of the evaluations tend to be pessimistic regarding the utility and validity of the available models. The reader certainly is advised to consider the views they provide when considering a specific prediction technique.

TABLE 7.7-1: COMPARISON BETWEEN HARDWARE AND HUMAN RELIABILITY [39]

	HARDWARE RELIABILITY	HUMAN RELIABILITY	HUMAN RELIABILITY
Function	Hardware Reliability	Discrete Task	Continuous Task
System Definition	A set of components which perform their intended functions.	A task which consists of several human behavioral units.	Continuous control task such as vigilance, tracking, and stabilizing
System Configuration	Functional relationships of components	Relationships of behavior units for a given task (task taxonomy)	Not necessary to define functional relationships between task units.
System failure analysis	Fault-tree analysis	Human error categorization; derivation of mutually exclusive and exhaustive set of human errors for a given task.	Binary error logic for continuous system response.
Nature of failure	- Mostly binary failure logic - Multi-dimensionality of failure - Common-cause failure	- Sometimes hard to apply binary error logic to human action - Multi-dimensionality of error - Common cause error - Error correction	Same as discrete task
Cause of failure	Most hardware failures are explained by the laws of physics and chemistry.	No well-codified laws which are generally accepted as explanations of human errors.	Same as discrete task
System reliability evaluation	- With probabilistic treatments of failure logic and statistical independence assumption between components, mathematical models are derived. - In cases of network reliability and phased mission reliability, which require statistical dependency between components, it is hard to evaluate exact system reliability.	Very difficult because of problems in depicting the functional relationships between human behavioral units.	With probabilistic treatments of binary error logic for system response stochastic models are derived.
Data	The data for most types of machines is relatively large and robust compared to human reliability.	- No trustworthy and useful data base exists for human behavior units. - Largely depends on the judgment of experts.	Same as discrete task.

A brief survey of current industrial practices was conducted by LaSala [42]. In the survey, aerospace industries were queried with regard to the techniques used in human reliability prediction, design approaches, design validation approaches, allocation approaches, and needed tools. The survey revealed that the greatest need was for front-end tools.

The facts that up to 70% of operational failures, Reference [43], and 40% of in-plant rework are due to human error demand aggressive consideration of operator and maintainer reliability for operational systems and aggressive consideration of assembler and maintenance technician reliability in in-plant operations.

7.7.2 Reliability, Maintainability, and Availability Parameters for Human - Machine Systems

A careful analysis of human-machine systems recognizes that both humans and machine elements can fail, and that human errors can have varying effects on a system. In some cases, human errors result from an individual's action during operation, while others are a consequence of system design or manner of use. Some human errors cause system failure or increase the risk of such failure while others merely create delays in reaching objectives. Thus, as with other system elements, the human elements exert a strong influence on the design and ultimate reliability of all human-machine systems.

The human interacts in a complicated manner with the non-human portions of the system. A tendency that must be resisted is to segregate human and machine functions. Watson and Hebenstreit [44] effectively characterized the interplay of human and machine in complex systems, as shown in Figure 7.7-1. In reality, effective system design recognizes that the "human-in-the-loop" cannot be segregated from other system functions.

Human errors take many forms and are due to many causes. There are types of human errors that are not caused specifically by design, although good design practices can reduce the occurrence of these errors. These are Reference [45]:

(1) Slips - attentional failures
  (a) Intrusion
  (b) Omission
  (c) Reversal
  (d) Misordering
  (e) Mistiming
(2) Lapses - memory failures
  (a) Omission of planned items
  (b) Place-losing
  (c) Forgetting intentions

(3) Mistakes - rule- and knowledge-based
   (a) Misapplication of good rules
   (b) Application of bad rules
   (c) Many types of knowledge-based mistake

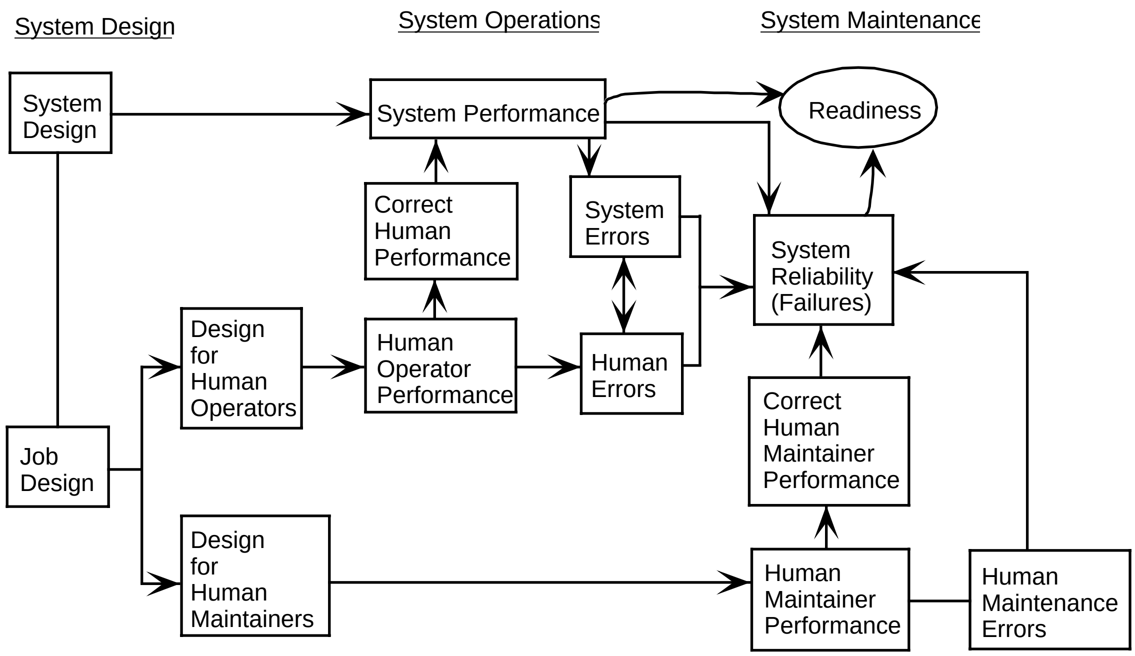
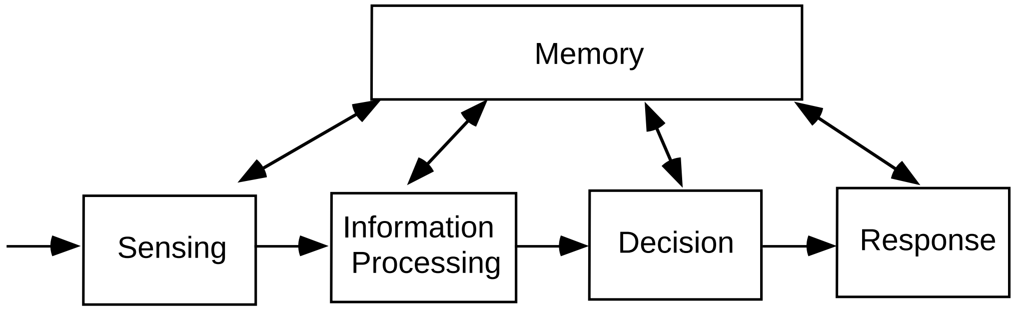


FIGURE 7.7-1: THE HUMAN IN SYSTEM RELIABILITY AND MAINTAINABILITY [44]

Closely related to the selection of reliability, maintainability, and availability models for human-machine systems is the subject of models of human performance. Although many models exist, for reliability purposes, the one that is most easily used is the "cognitive model" shown in Figure 7.7-2. The cognitive model considers a human function as four basic subfunctions, assisted by memory.

The reliability of the human function is affected by several types of factors as shown in Figure 7.7-3.



Examples: Radar operator, electric power monitor

FIGURE 7.7-2: THE COGNITIVE HUMAN MODEL

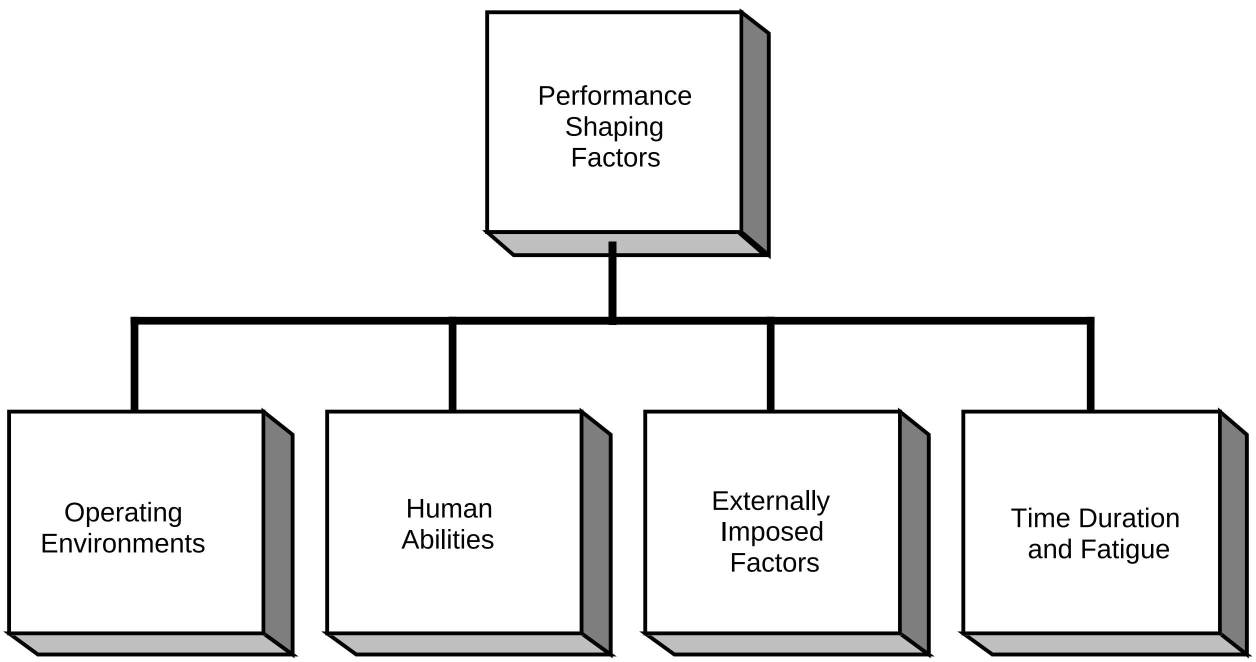


FIGURE 7.7-3: FACTORS THAT AFFECT HUMAN FUNCTION RELIABILITY

Of the factors shown in Figure 7.7-3, operating environments are, perhaps the easiest to understand. Some of the more commonly known environmental factors or combinations of factors are:

(1) Temperature-humidity
(2) Pressure-oxygen concentration
(3) Longitudinal and transverse vibration
(4) Ambient noise

For each of these, there is a region where human performance is not degraded, a region where human performance ceases, and a region of transition between the previous two (see Figure 7.7-4). Sources such as MIL-STD-1472E, "Human Engineering Design Criteria for Military Systems, Equipment, and Facilities" provide this information Although specific reliability data have not been published for these regions, inferences can be made regarding the impact on the human performance reliability. In the case of ambient noise, message complexity, vocabulary content, distance between speaker and listener, and background noise levels and frequencies affect human performance reliability.

Human abilities pertain to the ability of the human to detect input signals, analyze their meaning, make decisions, and then perform the proper response. Typically, inputs consist of visual, sound, or touch-related signals. There are minimum levels for the detectability of each and levels at which damage is done to the human. There also are transition zones from the threshold of detection to physical damage.

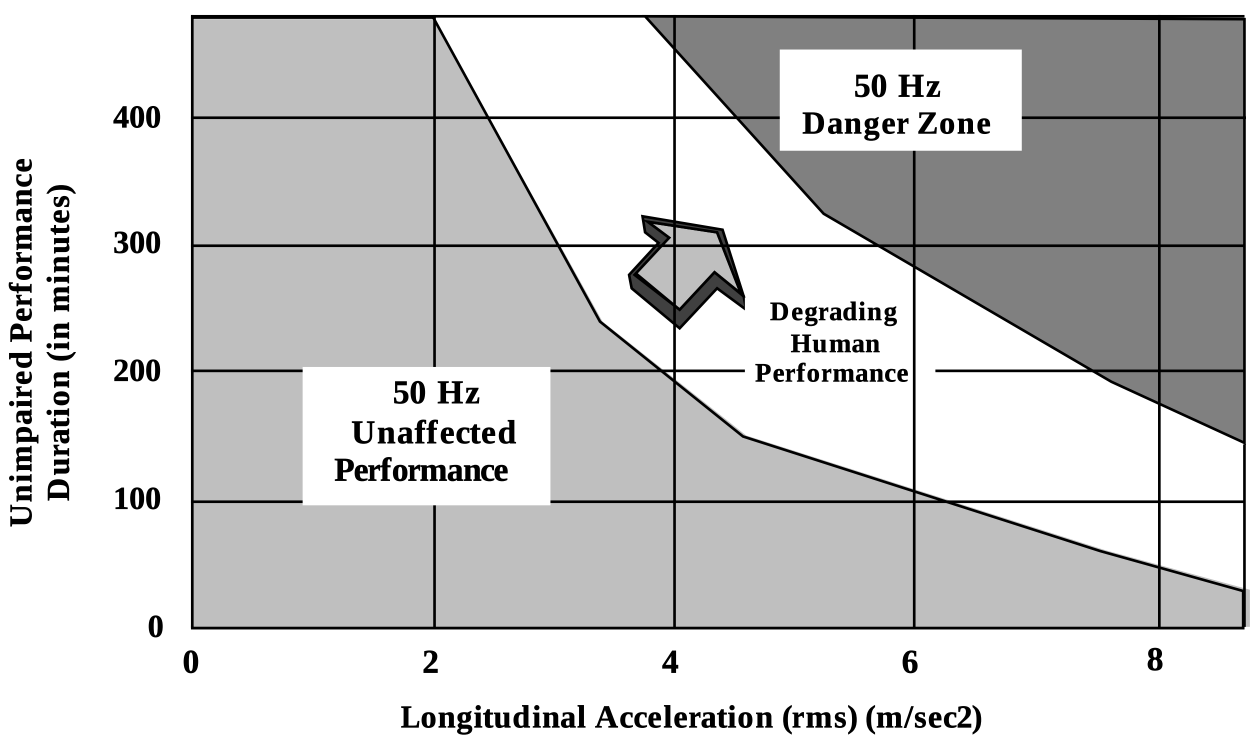


FIGURE 7.7-4: ZONES OF HUMAN PERFORMANCE FOR LONGITUDINAL VIBRATION (ADAPTED FROM MIL-STD-1472)

Externally imposed factors consist of workplace layout, assignments, group interactions and similar factors. Specific, reliability oriented data for these have not been tabulated, although studies have suggested changes in performance due to these factors.

Time duration and fatigue are important factors that frequently are neglected. For most tasks, performing 30 minutes without a break is the recommended limit because longer durations result in degradation of human performance. Also, there must be a balance between the allowed time to perform a task and the actual time to perform the task, otherwise human errors will occur.

Much of the system development process depends on quantitative measures. Consequently, for human-machine systems, it is necessary to define a set of parameters that includes the human as well as the hardware. Fortunately, it is possible to construct a set of analogues to conventional reliability, maintainability, and availability measures [46]. Two examples follow.

$$\text{Human Performance Reliability} = \frac{\text{No. Human Task Successes}}{\text{No. Human Task Attempts}}$$

$$\text{Human Availability} = 1 - \frac{\text{Unmanned Station Hours}}{\text{Total Hours}}$$

These parameters can be used in simulations and can be used in probability compounding models as well. Like all reliability and maintainability parameters, they should not be construed as ends in themselves but rather vehicles for obtaining good system designs.

7.7.3 Allocating System Reliability to Human Elements

The allocation of reliability and maintainability requirements is the first step in the man-machine system development process beyond the receipt of the customer requirements. This section discusses qualitative allocation and two forms of quantitative allocation: an application of the AGREE method and dynamic programming. Qualitative allocation pertains to the earliest stages of system functional analysis and the evaluation of potential design solutions. Although, in practice, quantitative allocation rarely is performed, the consequence of not performing a quantitative allocation is the inadequate design of human tasks and an increase in the likelihood of human error - in some cases to very significant and dangerous levels.

7.7.3.1 Qualitative Allocation

One of the early stages of system engineering is the identification and non-quantitative allocation of system functions. This step commonly is known as "functional analysis." "Functions" are discrete actions for the attainment of specific objectives. Generally, the products of functional analysis are functional flow diagrams that are structured in a hierarchical manner [47]. A simple example is shown in Figure 7.7-5.

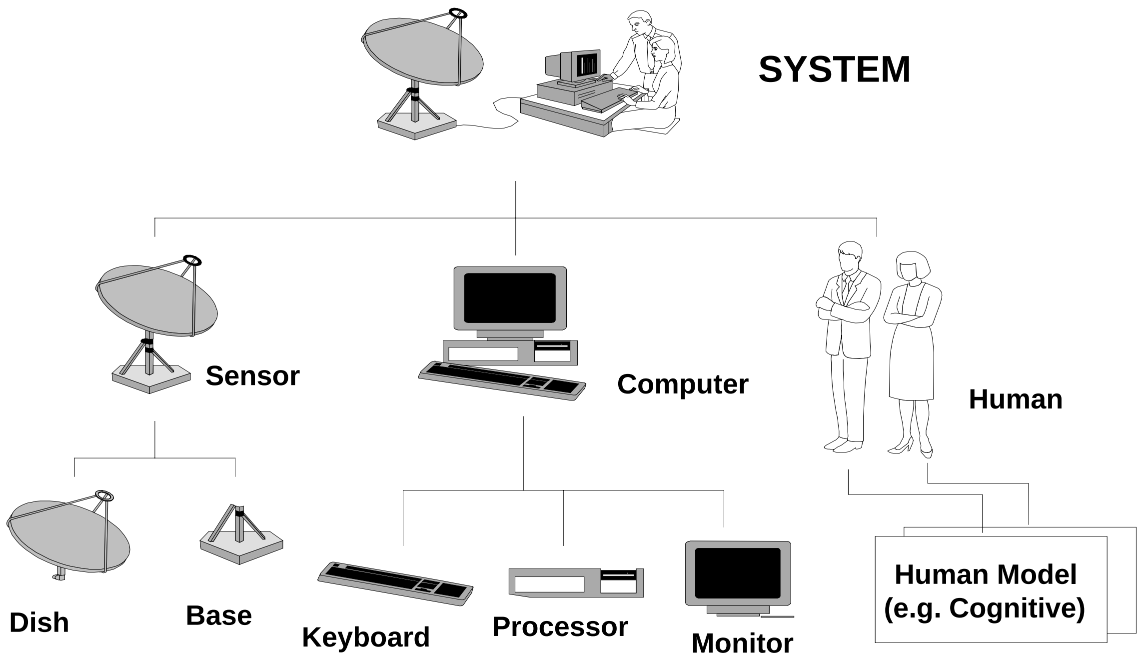


FIGURE 7.7-5: HIERARCHICAL STRUCTURE OF FUNCTIONAL ANALYSIS (EXAMPLE)

At an appropriate level in the functional analysis, it must be decided whether a function will be performed by human or machine. This can be a relatively high level, e.g. first tier, or at a detailed level such as the third or lower tier. For man-machine systems, the functional analysis can include operation and maintenance functions presented as separate flows or as a combined flow. Examples are given in reference [47].

Qualitative allocation is simply the selection of which functions are best performed by the human and which are best performed by the machine. Table 7.7-2 identifies the functions at which humans and machines excel. In general, the human is better at handling a variety of different information-processing tasks, adapting to new tasks and environments, devising new procedures, and resolving unexpected contingencies. The greatest limitations of the human are the rate of data processing and the amount of immediate retention.

TABLE 7.7-2: HUMAN-MACHINE COMPARATIVE CAPABILITIES

HUMAN SUPERIORITY	MACHINE SUPERIORITY
1. Originality (ability to arrive at new, different problem solutions)	1. Precise, repetitive operations
2. Reprogramming rapidly (as in acquiring new procedures)	2. Reacting with minimum lag (in microseconds, not milliseconds)
3. Recognizing certain types of impending failures quickly (by sensing changes in mechanical and acoustic vibrations)	3. Storing and recalling large amounts of data
4. Detecting signals (as radar scope returns) in high-noise environments	4. Being sensitive to stimuli (machines sense energy in bands beyond human's sensitivity spectrum)
5. Performing and operating though task-overloaded	5. Monitoring functions (even under stress conditions)
6. Providing a logical description of events (to amplify, clarify, negate other data)	6. Exerting large amounts of force
7. Reasoning inductively (in diagnosing a general condition from specific symptoms)	7. Reasoning deductively (in identifying a specific item as belonging to a larger class)
8. Handling unexpected occurrences (as in evaluating alternate risks and selecting the optimal alternate or corrective action)	___
9. Utilizing equipment beyond its limits as necessary (i.e. advantageously using equipment factors for safety)	___

From An Introduction to the Assurance of Human Performance in Space Systems, SP-6506, NASA, 1968.

7.7.3.2 Quantitative Allocation

The first of the quantitative methods, and the simplest, for allocating man-machine reliability is an adaptation of the AGREE allocation method. This method, described in Reference [48], was developed for electronic equipments and was based on unit complexity and importance.

Unit complexity is described in terms of modules, where a module is a single functional unit. Unit importance is defined as the probability that the system will fall if the unit fails. A importance value of one implies that the unit is essential for successful system operation. A value of zero means that the unit has no impact on system performance.

The AGREE allocation is expressed in terms of allocated reliability $R(t_j)$.

$$R(t_j) = 1 - \frac{1 - [R^*(T)]^{n_j / N}}{E_j}$$

where:

$R^*(T)$ = system reliability requirement
n_j = number of modules in unit j,
E_j = importance factor of unit j,
t_j = number of hours unit j will be required to operate in T system hours

N = total number of modules in the system.

Although the AGREE report discusses the allocation for redundant situations, the quality of the approximations for those cases is inadequate. Hence, serial system configurations should be considered the primary application of the AGREE allocation method.

To apply the AGREE method to man-machine systems, the system must be broken down into discrete functions and be depicted in a serial reliability block manner. The first order assignment of functions to man or machine can be made using the qualitative approach described in Section 7.7.3.1. In a similar manner to the machine portions, the human functions must be decomposed into discrete elements, e.g. the portions of the cognitive model. These elements determine function complexity. Function operating hours and importance then are determined. An example of a non-unity importance factor that is applicable to a man or a machine in a target (or malfunction) detection function might be that if the man or machine fails to perform properly, 25% of the targets (or malfunctions) may be lost. Hence E_j would be 0.25. The allocation formulas are used to determine the allocated failure rate or reliability as required.

In most practical situations, the problem is to design a man-machine system with the highest achievable reliability subject to practical considerations such as competing mission requirements, limitations on personnel, and limitations on cost. In an analytic sense, the problem is represented as a function (reliability) to be maximized subject to a set of constraints. Allocations with assigned constraints generally are solvable by means of dynamic programming because they become very complicated very quickly.

Dynamic programming is a mathematical technique for optimizing a sequence of decisions by breaking the sequence into a sequence of simpler problems. For each simpler problem, or stage, an optimal feasible solution is determined. The selected set of solutions is the optimal solution for the entire problem. It is used frequently in capital allocation and transportation problems. Blanchard and Fabricky [49] and Taha [50] have excellent treatments of the subject.

Mission reliability and operational readiness are common parameters for optimization. Other parameters could be used. For example, availability may be used in place of operational readiness depending on the planned use of the system. Bazovski [51] provided an excellent discussion for distinguishing between readiness and availability. The parameters provide a direct link between high level effectiveness requirements and component oriented design parameters such as MTBF, MTTR, and both numbers and skill levels of personnel.

To apply mission reliability and operational readiness to the allocation process, first identify critical functions and their reliability and maintainability design parameters and use the parameters to write expressions for the mission reliability and operational readiness of the man-machine system. One mission reliability constraint and one operational readiness equation must be written for each mission.

Cost and personnel constraints must also be constructed. Personnel can be specified in many ways. The most convenient way is to specify numbers of men at a discrete skill level. Acquisition cost, support cost and life cycle cost can all be used as constraints. Regardless of the type of cost selected, the system cost must be defined in terms of human and hardware contributions and in terms of basic design parameters.

Example

A multi-mission equipment in which each mission, i, has a probability of p_i being required. The reliability function associated with each mission is r_i. The r functions are constructed to include the human element. For example, an operational sequence diagram, which is roughly equivalent to a reliability block diagram, can be merged with a functional reliability block diagram to provide a mission reliability function. Human functions are constructed in terms of reliability and maintainability parameters.

With the above preparation, the allocation problem can be written as the following optimization problem:

Maximize R_{op} (operational reliability):

$$R_{op} = \sum_{i=1}^{n} p_i r_i$$

subject to:

$$R_m \geq P_m$$
$$P_m \geq X_m$$
$$N \leq v$$
$$C \leq c$$

where:

R_m	=	mission reliability
P	=	availability
N	=	number of personnel
C	=	cost
m	=	a specific mission

There will be one set of constraint equations for each mission. This leads to exactly the form of optimization problem that is solved by dynamic programming. A simplified flow of dynamic programming is shown in Figure 7.7-6.

7.7.4 Sources of Human Performance Reliability Data

One of the major issues in man-machine reliability is the location and availability of human performance reliability data. The tabulated data take the form of time measurements, including reaction time and task performance time, error and error frequency data. The most commonly used data are error data. Task performance data play an important role where task performance within a specified time is a criterion for mission success: e.g., restoration of full power within 30 minutes. Most of the data come from controlled laboratory studies; an increasing amount come from simulators; very little come from field use. The laboratory data have the liability of being derived from artificial situations that do not represent the real world. Consequently, although they have the requisite degree of precision, they have limited utility. Simulator data are better because they represent the real world more accurately, especially as simulators improve. However, they are collected generally for special applications and, hence, have limited application. Laboratory and simulator data vary from what would be expected in the real world because of the subjects' awareness that they are being tested. The most realistic data, field data, generally are not collected in a systematic way. Consequently, the physical and psychological environment in which the data were generated usually are not recorded or analyzed.

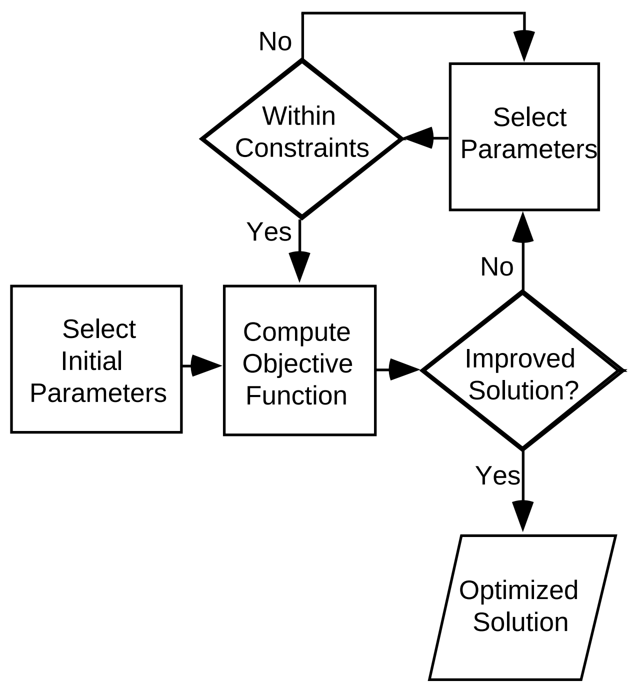


FIGURE 7.7-6: SIMPLIFIED DYNAMIC PROGRAMMING

An alternative to the use of "hard data" is the use of expert opinion. Techniques such as Delphi methods or paired comparison are used. The primary use of expert opinion occurs when hard data are modified special situations.

Early data sources consisted primarily of human factors data collections; e.g. Human Engineering Guide to Equipment Design [52], MIL-STD-1472E, “Human Engineering Design Criteria for Military System, Equipment, and Facilities.” More recent data sources are the following: “Handbook Of Perception and Human Performance” [53], “Engineering Data Compendium:

Human Perception and Performance” [54], the Compendium on a compact disc, and the Crew System Ergonomics Information Analysis Center.

The second vehicle for obtaining human performance reliability data is the "data bank." Table 7.7-3 shows the current major data banks [55]. Table 7.7-4 summarizes the data incorporated into each of the data banks.

Other sources of human performance reliability are described by references [56] and [57]. More detailed descriptions of many of the data sources and data banks described herein are given by Booher [55].

TABLE 7.7-3: DATA BANKS AND THEIR AFFILIATIONS [55]

	DATA BANK	ORGANIZATION
MICRO	Human Performance Evaluation System (HPES)	Nuclear Power
	Nuclear Computerized Library for Assessing Reactor Reliability (NUCLARR)	Nuclear Power
	Crew Station Ergonomics Information Analysis Center (CSERIAC)	Department of Defense
MACRO	Training and Performance Data Center (TPDC)	Department of Defense
	Materiel Readiness Support Activity (MRSA) MANPRINT Data Base	Department of the Army
	Project "A" Data Base	Department of the Army

TABLE 7.7-4: DATA CATEGORIES OF NATIONAL DATA BANKS [55]

Data Categories	HPES	NUCLARR	CSERIAC	TPDC	MRSA MANPRINT	PROJECT "A"
Human Performance	X	X	X	X		X
Human Performance (Error)	X	X	X			
Hardware Reliability		X			X	
Human Factors Engineering			X			
Manpower			X	X	X	
Personnel			X	X	X	X
Training				X	X	X
System Safety			X		X	
Health Hazards			X		X	

There is a recognized need for a human performance data bank that applies to both the military and commercial sectors. Until a broadly supported effort such as this is implemented, there will be both considerable subjectivity and many limitations in the use of available human performance reliability data.

7.7.5 Tools for Designing Man-Machine Systems

This section explores the various tools that are used to design reliable man-machine systems. The tools are many and varied in their approaches. An overview of them is provided by Figure 7.7-7.

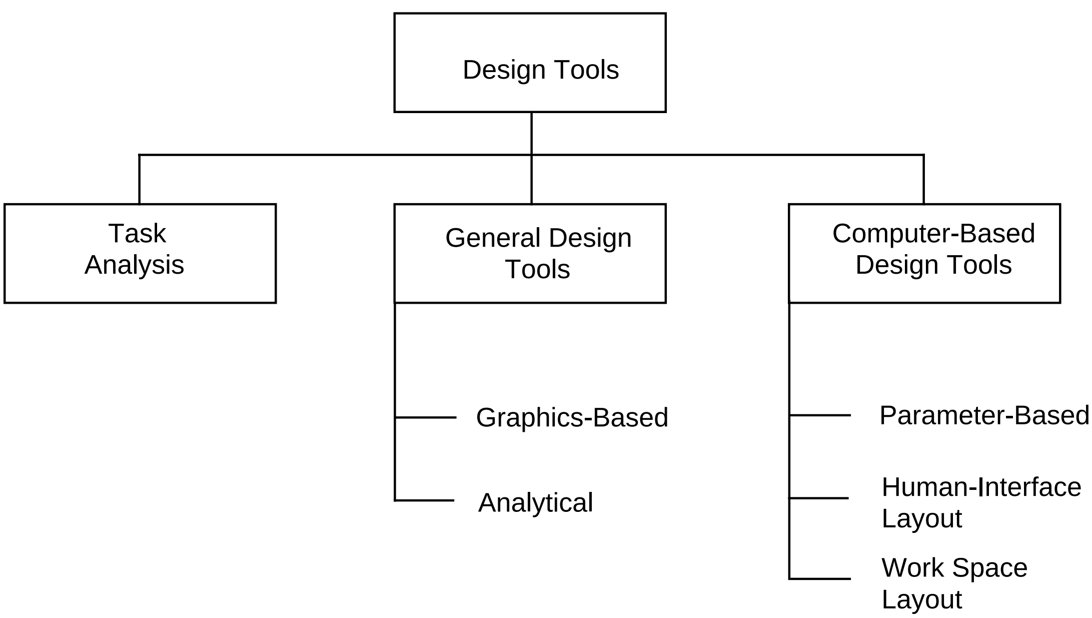


FIGURE 7.7-7: TOOLS FOR DESIGNING HUMAN-MACHINE SYSTEMS

7.7.5.1 Task Analysis

Task analysis is a critical part of a human factors program. However, it is usually not done as part of reliability program. Maintenance task analysis is usually done in a maintainability program. Task analysis focuses on the following:

(1) Input information to human
(2) Evaluation processes
(3) Action to be taken
(4) Environments and constraints
(5) Tools and job aids
(6) Manpower
(7) Communications

7.7.5.2 General Design Tools

There are many general design tools that apply to the design of human-machine interfaces. One of the most useful is the Operational Sequence Diagram (OSD). The features of the OSD are:

(1) Shows all participants
(2) Displays functional flow - all functions
(3) Represents types of operations with standard symbols
(4) Represents approximate time line
(5) Employs rules for representing flows
(6) Represents a certain level of system or process indenture

Goal, success, and fault trees are other useful tools. The general format of these is shown in Figure 7.7-8. Operator action trees and event trees are horizontally-oriented trees that show possible branches in action and the consequences. The Failure Mode, Effects, and Criticality Analysis can be adapted for use in the analysis of human-machine reliability by incorporating human error modes and evaluating their consequences.

The treatment of the role of traditional human factors is brief. The brevity should not be construed as a reflection of the importance of the subject. Many excellent volumes, some of which are referenced, have been written on the subject and any attempt to replicate even portions of them here would serve little purpose. Human factors provides many of the basic design disciplines that enable reliable systems to be designed. Too often, this fact is not appreciated and the human factors experts are given a secondary priority in system development (unless safety is a critical factor). Human factors experts need to be involved in the earliest stages of system development, especially in the function allocations. The role of human factors in the achievement of system reliability often is clouded by the lack of sensitivity to the human by reliability engineers. The key to understanding that role is the recognition that functionally, human behavior can be modeled as stimulus-input chain: internal-response: output- response. Complex behavior is a combination of many of these chains. Human errors occur when:

(1) A physical change in the environment is not perceived as a stimulus.

(2) Several stimuli cannot be discriminated by an operator.

(3) A stimulus is perceived, but its meaning is not understood.

(4) The stimulus is correctly understood, but the correct output-response is unknown.

(5) The correct output-response is known, but it is beyond the operator's physical capabilities.

(6) The correct output-response is within the operator's capabilities, but the response is performed incorrectly or out of sequence.

The implications for equipment design are: in order for an operator to respond adequately, the stimulus must be perceivable and it must demand a response which the operator is capable of producing. Hence equipment and task characteristics must be tailored to the capabilities and limitations of the operator. To accomplish this, the design of equipment must take into account limitations on body size, body weight, and reaction times to environmental stimuli. The operator must receive some verification or feedback from his actions.

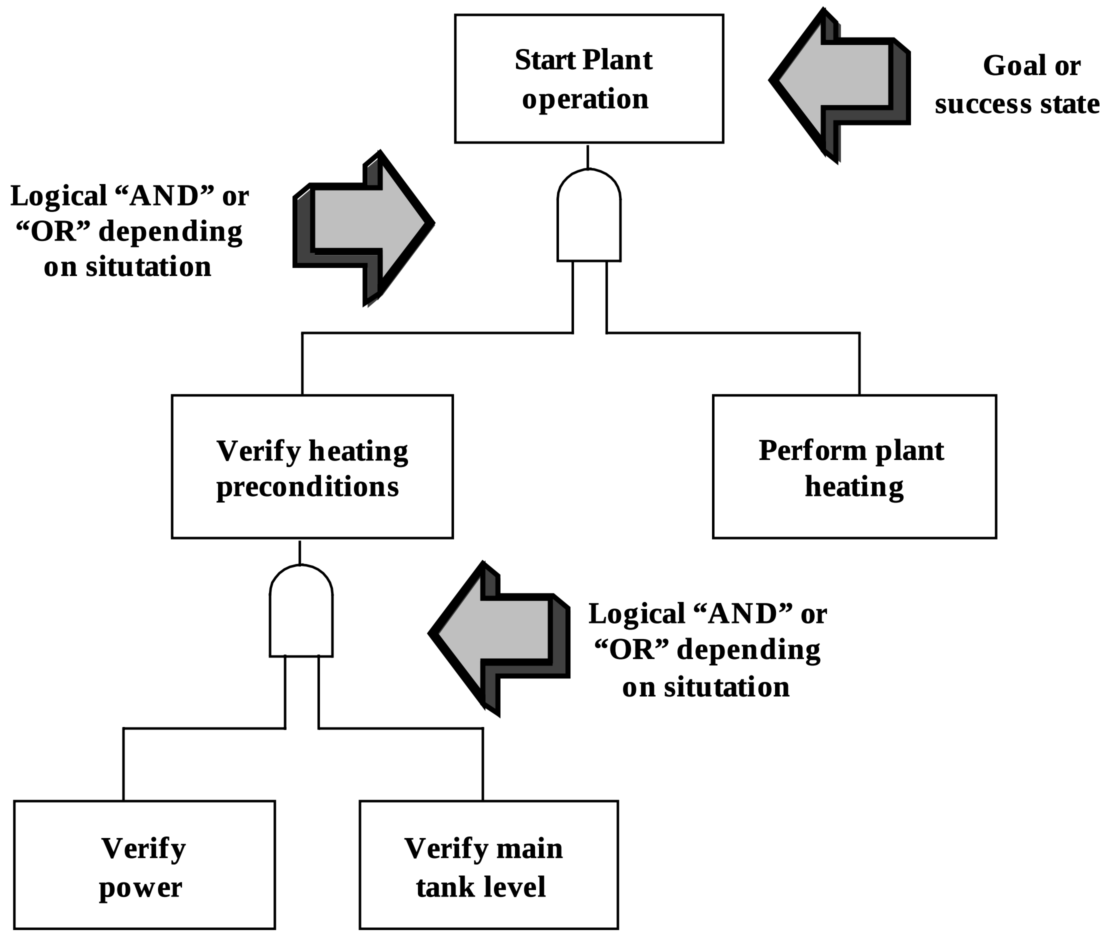


FIGURE 7.7-8: GOAL-SUCCESS TREE

7.7.5.3 Computer-Based Design Tools

There are many computer-based design tools and more are emerging rapidly. Some of the available ones are summarized in the following paragraphs. Almost all of the ones described here are proprietary. They are described here without either endorsement or criticism.

The computer-based design tools fall into three basic groups: parametric, interface design, and work space design. Some of the tools are:

(1) Parametric design
  (a) REHMS-D™
  (b) GOMS
(2) Interface design - VAPS™
  (a) VAPS™

(b) Computer Interface Toolkits (e.g. Visual Basic™)

(3) Work space design
   (a) SAMMIE™
   (b) CREW CHIEF
   (c) JACK™
   (d) SAFEWORK™

7.7.5.3.1 Parametric Design Tools

REHMS-D uses reliability as a metric for selection of human interface and task parameters. It includes two levels of parameter sensitivity analysis, provides on-line help and safety warnings, and derives plans for testing human interfaces. REHMS-D is based on the cognitive model and offers options for configuring inputs to the human and responses by the human. It addresses the effects of the following environmental factors: illumination, atmospheric conditions, temperature-humidity, pressure-oxygen, ambient noise, and vibration. GOMS develops four aspects of human tasks: goals, operators, methods, and selection. The "goals" are a user defined set of things to do or obtain to achieve a task. The "operators" are the actions available to achieve the goals. The methods are the sequence of operations and subgoals for achieving the task - the "how to." Finally, the "selection" is a means for choosing a method when more than one method applies - usually rule-based. GOMS uses time-to-perform for its metric.

7.7.5.3.2 Interface Design Tools

VAPS is a work station-based tools that draws objects, specifies data-driven animation, connects objects to application data and specifies response behavior, and communicates with external sources to interact.

7.7.5.3.3 Work Space Design Tools

CREW CHIEF is a 3-dimensional modeling system developed by the US Air Force Human Resources Laboratory, Wright-Patterson AFB, OH. It is oriented toward the computer graphic simulation of an aircraft maintenance technician and interfaces readily with existing commercial CAD systems. CREW CHIEF reduces the incidence of design problems by allowing the designer to perform maintainability analyses and correct design defects while the system is in the early design stage. It does this by providing a 3-dimensional modeling system that creates a computerized man-model.

SAMMIE (System for Aiding Man/Machine Interaction Evaluation) is a human factors, 3-D design system that includes a sophisticated, computerized man-model with built-in reach and sight capabilities. The man-model is constructed from logically related links or joint. Virtually any size or shape person can be represented through specific dimensional changes or statistical profiles of population groups.

SAFEWORK creates virtual humans of various percentiles. It contains access to population statistics. SAFEWORK mannequin movement includes fully articulated hand and spine models, virtual viewing, collision detection, and scene animation.

JACK includes 3D interactive environment for controlling articulated figures, a detailed human model, realistic behavior controls, anthropomorphic scaling, task animation and evaluation, view analysis, automatic reach and grasp, and collision detection and avoidance.

Design Evaluation for Personnel, Training, and Human Factors (DEPTH) analyzes maintenance activity using Transom Technologies Transom Jack™ human models; controls human model movements through standard mouse, body tracking equipment, or automatic simulation; handles a variety of populations, dress modes, and tools; and reports on accessibility, visibility, and strength.

7.7.6 Reliability Prediction for Human-Machine Systems

A great majority of the work published on human reliability has been concerned with human performance reliability prediction. Earlier work focused on probability compounding techniques. Some of these were linked with available data sources (never in really great abundance); other compounding techniques used specially collected data. With the proliferation of computers, digital simulation models were developed and used. More recently, stochastic models have been proposed. An excellent comparison of prediction techniques is given in reference [58]. Figure 7.7-9 shows the categories of the many human performance reliability prediction techniques that have been published.

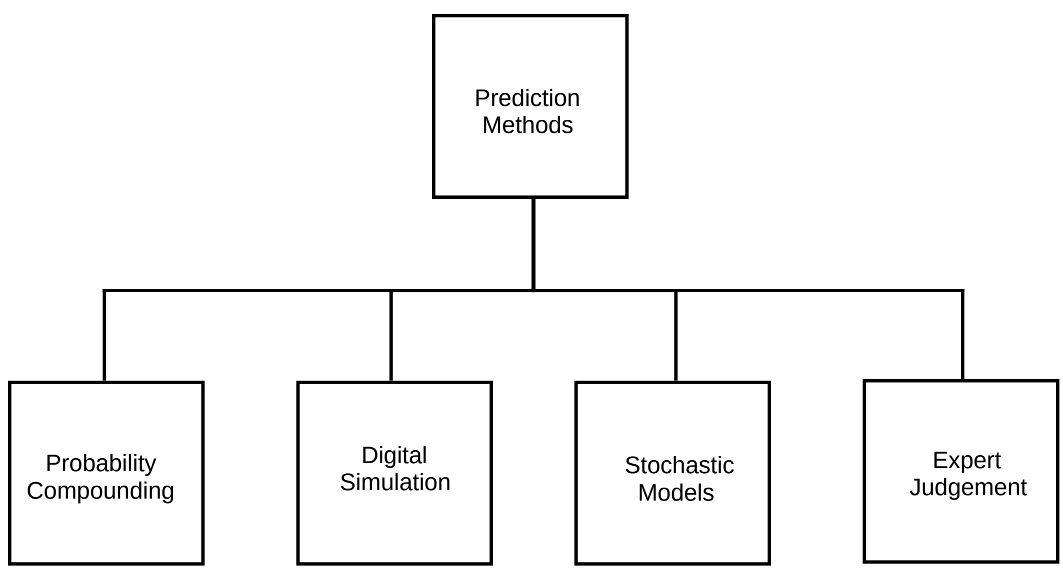


FIGURE 7.7-9: CATEGORIES OF HUMAN PERFORMANCE RELIABILITY PREDICTION METHODS

Although there are a great many models for prediction - over 20 years worth of work - there is no consensus on human reliability prediction technology or a human reliability parameter database [59]. Dougherty [60] noted much the same situation. His expectation is that there will be a recognition that there is a need for a second generation of human reliability models.

Swain [61] notes the following inadequacies in human reliability analysis:

(1) Inadequate data

(2) Stop-gap models and expert judgment are used in place of "hard" data

(3) Lack of agreement on expert judgment methods

(4) Inadequate calibration of simulator data

(5) Inadequate proof of accuracy in human reliability analyses

Increased use of higher mental functions is required by inadequate design of displays, controls, and their interactions.

The emphasis here is on the lack of data to support the human reliability analysis rather than the methodology itself. Swain does identify inadequate implementation of human factors disciplines as a root cause of the lack of data on favorable human performance situations.

7.7.6.1 Probability Compounding

There are a considerable number of probability compounding models for estimating human performance reliability in man-machine systems. Meister [40] provides excellent summaries of them. Selected techniques are summarized below.

Technique for Human Error Rate Prediction (THERP) [62], [63] has been the best known and most frequently applied technique for human reliability prediction. It is a method for predicting human error rates and for evaluating the degradation to a man-machine system likely to be caused by human errors in association with factors such as equipment reliability, procedures, and other factors. The THERP technique has been influenced strongly by hardware reliability techniques.

THERP involves five steps:

(1) Define the system or subsystem failure which is to be evaluated. This involves describing the system goals and functions and the consequences of not achieving them. It also requires identifying mission, personnel, and hardware/software characteristics.

(2) Identify and list all the human operations performed and their relationships to the system tasks and functions. This requires an analysis of all operator and maintainer tasks.

(3) Predict error rates for each human operation or group of operations. Errors likely to be made in each task or subtask must be identified. Errors that are not important in terms of system operation are ignored. This step includes estimating the likelihood of each error occurring and the likelihood of an error not being detected.

(4) Determine the effect of human errors on the system, including the consequences of the error not being detected. This requires the development of event trees. The left limbs of the event trees are success paths; the right limbs are failure paths. Probabilities are assigned to each path. The tree reflects the effects of task dependence. The relative effects of performance-shaping factors, e.g. stress and experience, are estimated.

(5) Recommend changes as necessary to reduce the system or subsystem failure rate as a consequence of the estimated effects of the recommended changes. The recommendations can be developed through the use of sensitivity analyses, in which factors and values are varied and effects monitored. THERP makes no assumptions about the dependence or independence of personnel behaviors. The data are taken from available sources.

One of the key aspects of THERP is the determination of the probability that an error or class of errors will result in a system failure. This probability is assigned a value F_i. Branching trees are constructed to determine the paths to system success and failure (Figure 7.7-10). The probability that an error will occur is given by P_i. F_iP_i is the joint probability that an error will occur and that the error will lead to system failure. $1-F_iP_i$ is the probability that an operation will be performed which does not lead to system failure. The probability that a class of errors will lead to system failure is given by:

$$Q_i = (1 - F_i P_i)^{n_i}$$

where n_i is the number of independent operations. The total system or subsystem failure rate is given by:

$$Q_T = 1 - \left[\prod_{k=1}^{n} (1 - Q_k) \right]$$

where Q_T is the probability that one or more failure conditions will result from errors in at least one of the n failure classes.

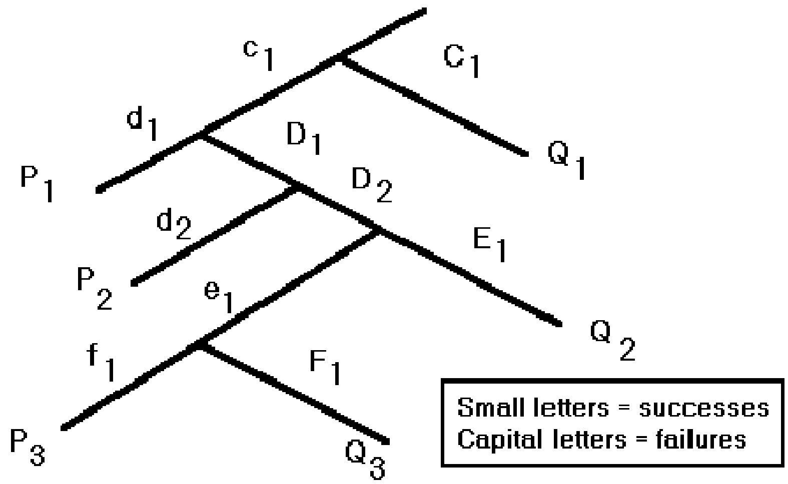


FIGURE 7.7-10: THERP PROBABILITY TREE [62]

THERP can be used for design analysis, manpower selection, prediction of system effectiveness, and determination of training requirements. For design analysis, it allows the comparison of alternative system configurations in terms of effect on operator capability. It also allows comparative analysis of initial system configuration options and reconfiguration if deficiencies are identified. For manpower selection, THERP allows the determination of the types, numbers, and skill levels of the personnel required to operate the system. For system effectiveness, THERP allows an assessment of whether quantitative requirements will be met. The determination of training requirements is more implicit than explicit. Unacceptable task performance error rates suggest the need for training to improve proficiency. Hence, THERP can suggest the need for training rather than specific training topics.

THERP can be applied to all types of equipments, tasks, and behaviors. With the aid of standard human engineering techniques, it can be used for design analysis. Finally, THERP can be applied to the early stages of system design as well as the later stages.

Constraints on its application are that it is applicable to situations where discrete task descriptions can be developed, error probability data must be available, the effects of performance-shaping factors must be known, and that time must be available to analyze and categorize all potential errors in a task. THERP is regarded as the standard tool for estimating human error probabilities in routine tasks. It uses performance shaping factors (PSFs) to make judgments about particular situations. However, experience has shown that in some cases, it was difficult to accommodate all of the PSFs that were considered important [64]. In many cases, THERP gave lower error probabilities than other methods. One evaluation of THERP [65] notes that THERP has the advantage of simplicity but does not account for the influence of time. Fragola [38] describes extensions to THERP, particularly with respect to nuclear power applications. Another evaluation [66] notes that when applied to severe accident applications, several problems were noted. In this case, the task information provided in NRC data forms typically is more detailed than required by THERP. Matching the NRC task data to THERP operator actions was found to be subjective and a source of error. Also the THERP error data received criticism with respect to its being an adaptation of non-nuclear data to nuclear applications. Krois et. al. note that other data bases are available to be used in THERP and obviate this last criticism.

Dougherty and Fragola [38] have introduced the time-reliability correlation (TRC) system. This approach uses simulator training results to create a family of time-reliability correlations, which are adjusted with either the Success Likelihood Index or other expert judgment methods to account for special conditions. TRC is the relationship between human performance reliability and time. Data from simulators suggest that the lognormal distribution is sufficient for modeling TRCs. Interpolation between the s-confidence bounds can be accomplished through the use of a Success Likelihood Index (SLI). The SLI is derived in the following manner:

(1) Choose the influences appropriate to the event and the situation.

(2) Rank the influences as multiples of the least important for a given situation, which is set at "10."

(3) Sum the rankings of all influences and normalize the rankings to this sum.

(4) Assess the impact of each influence from best (1) to worst (10).

(5) Compute the "dot product" of the ranking and the quality vectors. This is the SLI.

(6) Apply the SLI. Mathematically, the SLI is expressed by:

$$SLI = \sum_{1}^{N} \left(\frac{r_i}{R}\right) q_i$$

where:

$$R = \sum_{1}^{N} r_i$$

and r_i is the rank of the influence i and q_i is the quality of the influence i.

Dougherty and Fragola focus on a lognormal TRC based on simulator data. This is in consonance with the modified Human Cognitive Reliability.

Human Cognitive Reliability (HCR) was developed by Hannaman et al. [67] for calculating the operator probability of non-response to a cognitive processing task as a function of time. The type of cognitive processing may be rule based, skill based, or knowledge based. For task j, the probability of non-response P(t) is given by:

$$P(t) = \exp\text{-}(X_j)^{b_j}$$

where:

$$X_j = \frac{(t/\,T_{med}) - C_{gj}}{C_{nj}}$$

and T_{med} = median time to perform the task corrected by a shaping factor K_j

b_j = shape parameter

C_{gj} = time delay factor as a fraction of T_{med} for type j cognitive processing

C_{nj} = scale parameter as a fraction of T_{med} for type j cognitive processing

The dependency between tasks was not considered. The model was derived from a three parameter Weibull distribution. The model was used to develop normalized curves corresponding to rule based, skill based, and knowledge based cognitive processing. In applying the HCR approach to operator task analysis, a table records the following for each subtask:

(1) Dominant human behavior
(2) Stress level
(3) Operator experience
(4) Quality of operator/system interface
(5) Median time assumed

The HCR approach has been modified [68] to use the log-normal distribution instead of the Weibull. The modified approach has the acronym HCR/ORE and is supported by simulator data. Guassardo [65] notes that the HCR must be used carefully because variations in applications can lead to very different results. The model does allow some accounting for time effects on error probability but is complicated by the fact that the correlation only can be used once when subtasks have the same influence parameters. In this case, there is an ambiguity regarding whether or not to cluster subtasks or to convolve individual subtask correlations. When examining consistency among teams using HCR, Poucet [64] noted that the results have greater variability than THERP methods. The method was very sensitive to assumptions about median time and the behavior type of the action Very good median response time data must be available in order to apply HCR. Poucet also notes that some of the teams in his comparative study combined the use of THERP and HCR. THERP was used for manual or routine tasks; HCR was used for cognitive tasks.

7.7.6.2 Stochastic Models

This approach to modeling the impact of the human in man-machine systems employs Markov models to analyze parallel, k-of-n, and standby redundant configurations with human errors and common failures. The human error is defined as a failure to perform a prescribed task (or the performance of a prohibited action), which could result in damage to equipment and property or disruption of scheduled operations. The errors are categorized as being "critical" or "non-critical." A critical error causes system failure. Common cause failures are cases where multiple units fail due to a single cause.

Five models are described by Dhillon [69] [70]. Each addresses a different redundant configuration. The models assume that:

(1) Units fail independently

(2) Failure rates for hardware, human error, and common cause failures are constant

(3) Repair rates are constant

(4) A repaired system is as good as new

(5) Switchover mechanisms are perfect for standby configurations

(6) System units are statistically identical

The first model represents a two independent and identical unit parallel system, which can fail because of human error or hardware failure. A Markov model is constructed and an expression for system availability A and mean-time-to-repair (MTTR) is obtained. An expression for mean-time-to-failure (MTTF) also is derived. All the expressions are complicated functions of the state transition probabilities (failure rates, error rates, and repair rates).

The second model is a special case of the first when the non-critical human error rate is zero. The non-critical human errors are omitted from the system transition diagram, which becomes much simplified. Expressions are derived for A, MTTR, MTTF, and variance of time to failure (TTF).

The third model represents a 2-out-of-3 unit system with critical human errors and common cause failures. All system units are identical. A system reliability function and an expression for MTTF are derived. It is noted that repair helps to increase MTTF and human errors decrease it, as expected.

The fourth model is a 3-out-of-four system with critical human errors and common cause failures. MTTF and TTF variance expressions are derived.

The fifth model represents a standby system with critical human errors and common cause failures. Again, MTTF and TTF variance are calculated.

7.7.6.3 Digital Simulation

Digital simulation provides an inexpensive means for evaluating the impact of operator and maintainer performance in man-machine systems without the cost or complexity of physical experiments. It allows for the identification of problem areas before the actual system has been constructed. It can provide the answers to the following questions [71]:

(1) What are the quantitative personnel requirements?

(2) What are the qualitative personnel requirements? Where, during the system utilization, are the operators most overloaded? Underloaded?

(3) How will cross-training improve system effectiveness?

(4) Are the system operators able to complete all of their required tasks within the time allotted?

(5) Where in the task sequence are operators or teams likely to fail most often? Least often?

(6) In which states of the system is the human subsystem and its components least reliable and why?

(7) How will task restructuring or task allocation affect system effectiveness?

(8) How much will performance degrade when the systems operators are fatigued or stressed?

(9) How will various environmental factors (e.g. heat, light, terrain) affect total man-machine system performance?

(10) To what extent will system effectiveness improve or degrade if more or less proficient operators are assigned?

(11) How do group factors such as morale and cohesion affect system performance?

Simulations can be used in the conceptual as well as developmental stages of system evolution. They provide vehicles for tradeoff analyses.

At the time this section was written, the Maintenance Personnel Performance Simulation (MAPPS) [72] is the only supported member of the family of Siegel-Wolf simulation models. The basic features of the model are shown in Table 7.7-5.

TABLE 7.7-5: MAPPS SCOPE

FEATURE	MODEL LIMIT
Maximum number of tasks	200
Number of maintainers	2-8
Types of maintainers	5
Number of subtasks	100
Types of subtasks	28
Maximum task duration (days)	2
Number of shifts	1-10
Protective clothing types	3
Types of ability	2

The model establishes a team of up to eight members who begin a maintenance task at time t=0 under a set of initial conditions established by the user. For each maintenance task, subtasks are identified with data and shift termination information. Subtasks may be repeated because of inadequate performance, group decisions and looping back. MAPPS selects the maintainers to be assigned to each task or subtask and then processes the input data and current system state data to arrive at estimates of task performance. MAPPS is written in FORTRAN IV H (Enhanced) for the IBM 3033 system. MAPPS output can be characterized in terms of type of information and degree of detail. Output can be provided by subtask(the most detailed), by iteration, and by run. For a subtask, the model will provide results such as: degree of success; probability of success; start and end times; duration; time and communication stresses; effects of accessibility, fatigue and heat; and required ability.

The Cognitive Environment Simulation (CES) [73] is an artificial intelligence approach that simulates predictions about operator action by simulating the processes by which intentions are formed. It enables the analyst to represent of state of knowledge regarding a particular situation and then observe the consequences in terms of human intended actions. It is an application of the proprietary EAGOL artificial intelligence problem solving system developed by Seer Systems. EAGOL has the ability to reason in multiple fault situations and to reason in situations that evolve over time. The specific CES application of EAGOL is for emergency situations in nuclear power plants. Note that CES is not intended to be a "micro" view of human cognitive processing. Applying CES consists of matching CES resources to those of the power plant under study. Input data consists of a time series of plant state data that would be available to operator personnel. The data are processed into a virtual display board which reports the status of the plant, recognizes undesirable situations, and generates proposed rectifications to those situations. The output is a series of intentions to act and resolve the undesirable situation. CES contains three types of activities: monitoring, explanation building, and response management. The CES user can vary the demands placed on CES and the resources available to solve problems. CES is used in association with the Cognitive Reliability Assessment Technique (CREATE) for use in the probabilistic risk analysis of nuclear power plants.

7.7.6.4 Expert Judgment Techniques

These are a collection of techniques that address the lack of "hard data" or firm interpretations of data through the use experts.

The Success-Likelihood Index Methodology (SLIM, SLIM-MAUD) [74] [75] examines performance shaping factors (PSFs) and establishes both ratings and weight for each PSF. SLIM-MAUD is a personal computer implementation of SLIM. The MAUD acronym refers to "Multi-Attribute Utility Decomposition." MAUD is a proprietary stand alone software package that aids the user in assessing alternatives. PSFs that can be considered are:

(1) Situational characteristics
(2) Job and task instructions
(3) Task characteristics
(4) Equipment characteristics
(5) Psychological stressors
(6) Physiological stressors
(7) Internal factors (training, experience, skill)

SLIM ranks the most important PSFs. The products of the rating and the normalized weight for each task are added to obtain the SLI (described earlier). The SLI is related to the task success probability through the following calibration equation:

$$\ln P(\text{success}) = a^*\text{SLI} + b$$

where a and b are empirical constants. Rosa et al. [76] notes that SLIM-MAUD requires that tasks be sorted into subsets of 4 to 10 tasks that are similarly affected by a proposed set of PSFs. The weighting and ranking of tasks is accomplished by a group of experts, usually four in number, who are led by a facilitator. Analysis is conducted with the aid of the MAUD computer program.

Rosa et al. noted many positive characteristics of SLIM-MAUD, including face validity, practicality, estimates with acceptable levels of reliability, ease of use and understanding, and ability to identify which PSFs have the most effect on the SLIs. Guassardo [65] notes that SLIM results are highly dependent on the boundary conditions used to find the calibration equation coefficients. Poucet [64] indicates that the SLIM results were highly dependent on the calibration reference points. The use of SLIM is recommended only if good reference data are available. Poucet also notes that SLIM does not address relationships among PSFs when such exist.

7.7.7 Verification of Human Performance Reliability

For human-machine systems, the purposes of verification testing can be described as follows [77].

(1) Demonstrate the conformance of the product to human engineering design criteria

(2) Confirm compliance with specific performance requirements

(3) Secure quantitative measures of human-machine performance characteristics that are functions of human-machine interaction

(4) Determine whether or not undesirable characteristics have been introduced

The demonstration of human performance reliability (in maintenance situations) may overlap with maintainability demonstrations or testing. The same set of tasks may be considered but with different criteria. For example, in a maintainability demonstration, the principle concern is the time required to complete a task. If the same task is employed in a human performance reliability context, the important criteria are not only correct completion of the task but also completion of the task within a time constraint. The references provide additional details on structuring tests to estimate maintenance technician reliability.

For estimates of reliability in an operator type situation, data must be accumulated either by use of a simulator or by an expanded reliability demonstration that includes the operator as well as the equipment. In either case, the data will resemble actual field results only to the extent that the test scenario and the performance of the test subjects resemble the actual field conditions.

7.8 Failure Mode and Effects Analysis (FMEA)

7.8.1 Introduction

Failure Mode and Effects Analysis is a reliability procedure which documents all possible failures in a system design within specified ground rules. It determines, by failure mode analysis, the effect of each failure on system operation and identifies single failure points, that are critical to mission success or crew safety. It may also rank each failure according to the criticality category of failure effect and probability occurrence. This procedure is the result of two steps: the Failure Mode and Effect Analysis (FMEA) and the Criticality Analysis (CA).

In performing the analysis, each failure studied is considered to be the only failure in the system, i.e., a single failure analysis. The FMEA can be accomplished without a CA, but a CA requires that the FMEA has previously identified critical failure modes for items in the system design. When both steps are done, the total process is called a Failure Mode, Effects and Criticality Analysis (FMECA). The procedures for performing both the FMEA and the CA are found in Reference [78] and Reference [79]. At the time of this update Reference [78], MIL-STD-1629, was scheduled to be cancelled and replaced by a non-government standard by June 1997. However, it is not known at this time what that new document will be.

FMEA utilizes inductive logic in a "bottoms up" approach. Beginning at the lowest level of the system hierarchy, (e.g., component part), and from a knowledge of the failure modes of each part, the analyst traces up through the system hierarchy to determine the effect that each failure mode will have on system performance. This differs from fault tree analysis (discussed in the next section) which utilizes deductive logic in a "top down" approach. In fault tree analysis, the analyst assumes a system failure and traces down through the system hierarchy to determine the event, or series of events, that could cause such a failure.

The FMEA provides:

(1) A method of selecting a design with a high probability of operational success and crew safety.

(2) A documented method of uniform style for assessing failure modes and their effect on operational success of the system.

(3) Early visibility of system interface problems.

(4) A list of possible failures which can be ranked according to their category of effect and probability of occurrence.

(5) Identification of single failure points critical to mission success or to crew safety.

(6) Early criteria for test planning.

(7) Quantitative and uniformly formatted data input to the reliability prediction, assessment, and safety models.

(8) A basis for design and location of performance monitoring and fault sensing devices and other built-in automatic test equipment.

(9) A tool which serves as an aid in the evaluation of proposed design, operational, or procedural changes and their impact on mission success or crew safety.

Items (5) and (8) are the two most important functions performed by an FMEA.

The FMEA is normally accomplished before a reliability prediction is made to provide basic information. It should be initiated as an integral part of the early design process and should be periodically updated to reflect design changes. Admittedly, during the early stages, one usually does not have detailed knowledge of the component parts to be used in each equipment. However, one usually has knowledge of the "black boxes" which make up the system. Thus, at this stage, an FMEA might start at the "black box" level and be expanded as more detailed knowledge becomes available. This analysis may also be used to provide a model for analyzing already-built systems. An FMEA is a major consideration in design reviews.

The principles of FMEA are straightforward and easy to grasp. The practice of FMEA is tedious, time consuming and very profitable. It is best done in conjunction with Cause-Consequence and Fault Tree Analysis. The bookkeeping aspects, namely, the keeping track of each item and its place in the hierarchy, are very important because mistakes are easily made.

The Cause-Consequence chart shows the logical relationships between causes (events which are analyzed in no more detail) and consequences (events which are of concern only in themselves, not as they in turn affect other events). The chart usually is represented with consequences at the top and causes at the bottom; and the words Top and Bottom have come into common use to describe those portions of the chart. A Failure Modes and Effects Analysis (FMEA) deals largely with the bottom part of the chart. A fault tree is a part of a Cause-Consequence chart. It consists of only one consequence and all its associated branches. The Cause-Consequence chart is created by superimposing the separately created fault trees. The Cause-Consequence chart can be used to organize one's knowledge about any set of causes and their consequences; its use is not limited to hardware oriented systems.

The FMEA consists of two phases which provide a documented analysis for all critical components of a system. First, however, definitions of failure at the system, subsystem, and sometimes even part level must be established.

Phase 1 is performed in parallel with the start of detailed design and updated periodically throughout the development program as dictated by design changes. Phase 2 is performed before, or concurrent with, the release of detail drawings.

The Phase 1 analysis consists of the following steps:

(1) Constructing a symbolic logic block diagram, such as a reliability block diagram or a Cause-Consequence chart.

(2) Performing a failure effect analysis, taking into account modes of failure such as:
(a) Open circuits
(b) Short circuits
(c) Dielectric breakdowns
(d) Wear
(e) Part-parameter shifts

(3) Proper system and item identification.

(4) Preparation of a critical items list.

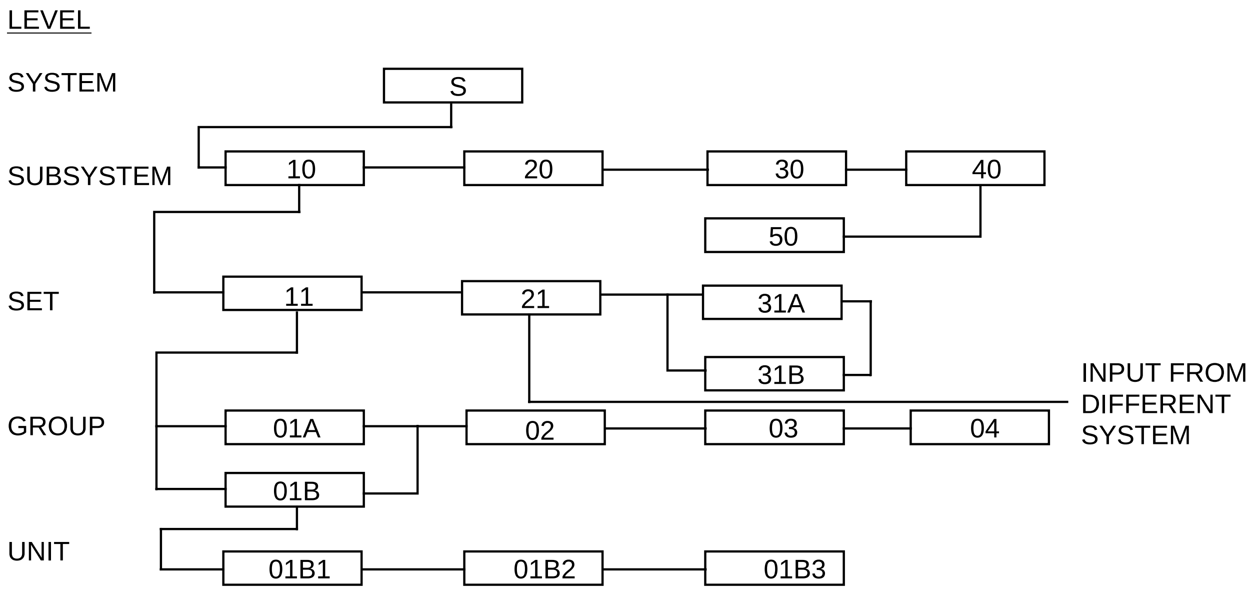
During Phase 2, the results of Phase 1 are revised and updated as required by design changes. In addition, all items in the system are analyzed to determine their criticality with respect to the system.

7.8.2 Phase 1

During this phase the following detailed steps are performed:

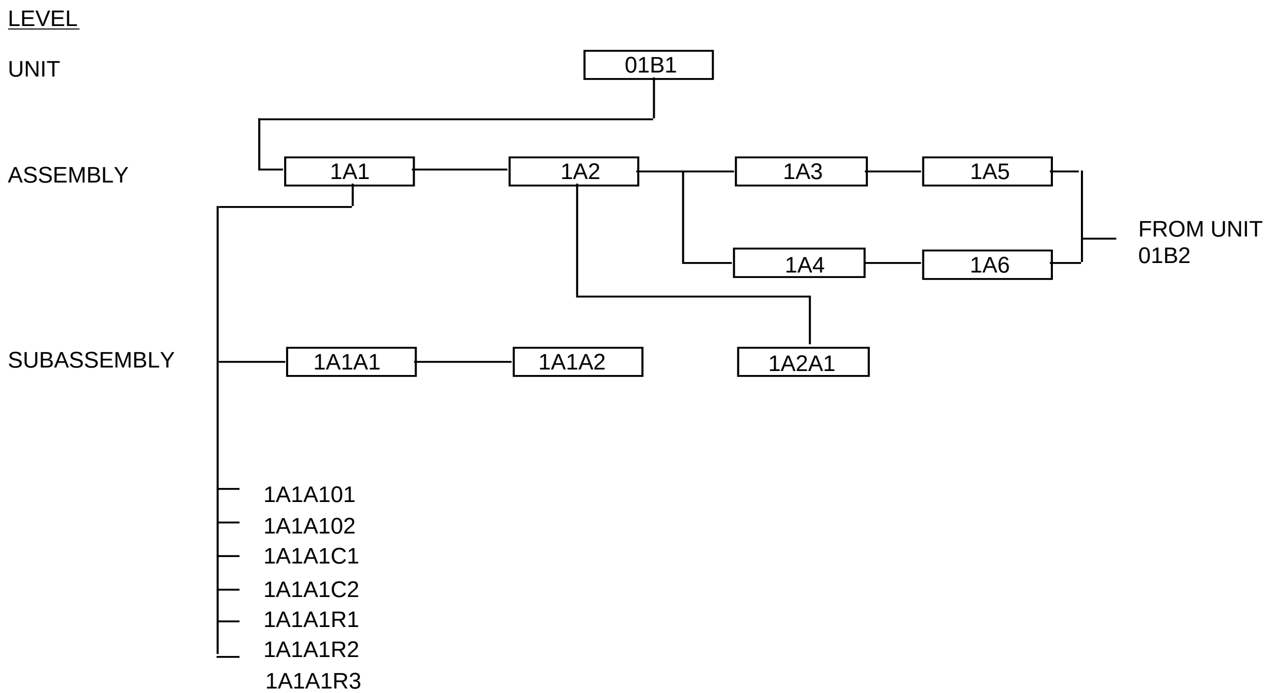
(1) A Symbolic Logic Block Diagram is constructed. This diagram is developed for the entire system to indicate the functional dependencies among the elements of the system and to define and identify its subsystems. It is not a functional schematic or a signal flow diagram, but a model for use in the early analysis to point out weaknesses. Figures 7.8-1 and 7.8-2 show typical symbolic logic diagrams. Figure 7.8-1 illustrates the functional dependency among the subsystems, sets, groups, and units that make up the system. Figure 7.8-2 illustrates the functional dependencies among assemblies, subassemblies, and parts that make up one of the units in Figure 7.8-1.

(2) A failure effect analysis is performed for each block in the symbolic logic block diagram, indicating the effect of each item failure on the performance of the next higher level on the block diagram. Table 7.8-1 shows a typical group of failure modes for various electronic and mechanical parts. The failure mode ratios are estimates and should be revised on the basis of the user's experience. However, they can be used as a guide in performing a detailed failure effect analysis.



Notes:
(1) The system depends on subsystems 10, 20, 30 and 40
(2) Subsystem 10 depends on sets 11, 21, 31A, and 31B
(3) Set 11 depends on groups 01A, 01B, 02, 03, and 04
(4) Group 01B depends on units 01B1, 01B2, and 01B3
(5) Sets 31A and 31B are redundant
(6) Groups 01A and 01B are redundant
(7) Subsystem 40 depends on subsystem 50
(8) Set 21 depends upon an input from another system

FIGURE 7.8-1: TYPICAL SYSTEM SYMBOLIC LOGIC BLOCK DIAGRAM



Notes:

(1) Unit 01B1 depends on assemblies 1A1, 1A2 AND either '1A3 and 1A5' OR '1A4 and 1A6'
(2) Assembly 1A1 depends on subassemblies 1A1A1 AND 1A1A2
(3) Assembly 1A2 depends on subassembly 1A2A1
(4) Subassembly 1A1A1 depends on all parts contained therein

FIGURE 7.8-2: TYPICAL UNIT SYMBOLIC LOGIC BLOCK DIAGRAM

TABLE 7.8-1: FAILURE MODE DISTRIBUTION OF PARTS[6]

DEVICE TYPE	FAILURE MODE	MODE PROBABILITY (α)
Accumulator	Leaking Seized Worn Contaminated	.47 .23 .20 .10
Actuator	Spurious Position Change Binding Leaking Seized	.36 .27 .22 .15
Alarm	False Indication Failure to Operate Spurious Operation Degraded Alarm	.48 .29 .18 .05
Antenna	No Transmission Signal Leakage Spurious Transmission	.54 .21 .25
Battery, Lithium	Degraded Output Startup Delay Short Open	.78 .14 .06 .02
Battery, Lead Acid	Degraded Output Short Intermittent Output	.70 .20 .10
Battery, Ni-Cd	Degraded Output No Output	.72 .28
Bearing	Binding/Sticking Excessive Play Contaminated	.50 .43 .07
Belt	Excessive Wear Broken	.75 .25
Brake	Excessive Wear Leaking Scored Corroded Loose	.56 .23 .11 .05 .05
Bushing	Excessive Wear Loose Cracked	.85 .11 .04
Cable	Short Excessive Wear Open	.45 .36 .19
Capacitor, Aluminum, Electrolytic	Short Open Electrolyte Leak Decrease in Capacitance	.53 .35 .10 .02

6 Reliability Analysis Center, "Failure Mode/Mechanism Distributions" (FMD-91)

TABLE 7.8-1: FAILURE MODE DISTRIBUTION OF PARTS (CONTD)

DEVICE TYPE	FAILURE MODE	MODE PROBABILITY (α)
Capacitor, Ceramic	Short Change in Value Open	.49 .29 .22
Capacitor, Mica/Glass	Short Change in Value Open	.72 .15 .13
Capacitor, Paper	Short Open	.63 .37
Capacitor, Plastic	Open Short Change in Value	.42 .40 .18
Capacitor, Tantalum	Short Open Change in Value	.57 .32 .11
Capacitor, Tantalum, Electrolytic	Short Open Change in Value	.69 .17 .14
Capacitor, Variable, Piston	Change in Value Short Open	.60 .30 .10
Circuit Breaker	Opens Without Stimuli Does Not Open	.51 .49
Clutch	Binding/Sticking Slippage No Movement	.56 .24 .20
Coil	Short Open Change in Value	.42 .42 .16
Connector/Connection	Open Poor Contact/Intermittent Short	.61 .23 .16
Counter Assembly	Inaccurate Count Seized	.91 .09
Diode, General	Short Open Parameter Change	.49 .36 .15
Diode, Rectifier	Short Open Parameter Change	.51 .29 .20

TABLE 7.8-1: FAILURE MODE DISTRIBUTION OF PARTS (CONTD)

DEVICE TYPE	FAILURE MODE	MODE PROBABILITY (α)
Diode, SCR	Short Open	.98 .02
Diode, Small Signal	Parameter Change Open Short	.58 .24 .18
Diode, Thyristor	Failed Off Short Open Failed On	.45 .40 .10 .05
Diode, Triac	Failed Off Failed On	.90 .10
Diode, Zener, Voltage Reference	Parameter Change Open Short	.69 .18 .13
Diode, Zener, Voltage Regulator	Open Parameter Change Short	.45 .35 .20
Electric Motor, AC	Winding Failure Bearing Failure Fails to Run, After Start Fails to Start	.31 .28 .23 .18
Fuse	Fails to Open Slow to Open Premature Open	.49 .43 .08
Gear	Excessive Wear Binding/Sticking	.54 .46
Generator	Degraded Output No Output Fails to Run, After Start Loss of Control	.60 .22 .09 .09
Hybrid Device	Open Circuit Degraded Output Short Circuit No Output	.51 .26 .17 .06
Injector	Corroded Deformed Cracked/Fractured	.87 .08 .05

TABLE 7.8-1: FAILURE MODE DISTRIBUTION OF PARTS (CONT'D)

DEVICE TYPE	FAILURE MODE	MODE PROBABILITY (α)
Keyboard Assembly	Spring Failure Contact Failure Connection Failure Lock-up	.32 .30 .30 .08
Lamp/Light	No Illumination Loss of Illumination	.67 .33
Liquid Crystal Display	Dim Rows Blank Display Flickering Rows Missing Elements	.39 .22 .20 .19
Mechanical Filter	Leaking Clogged	.67 .33
Meter	Faulty Indication Unable to Adjust Open No Indication	.51 .23 .14 .12
Microcircuit, Digital, Bipolar	Output Stuck High Output Stuck Low Input Open Output Open	.28 .28 .22 .22
Microcircuit, Digital, MOS	Input Open Output Open Supply Open Output Stuck Low Output Stuck High	.36 .36 .12 .09 .08
Microcircuit, Interface	Output Stuck Low Output Open Input Open Supply Open	.58 .16 .16 .10
Microcircuit, Linear	Improper Output No Output	.77 .23
Microcircuit, Memory, Bipolar	Slow Transfer of Data Data Bit Loss	.79 .21
Microcircuit, Memory, MOS	Data Bit Loss Short Open Slow Transfer of Data	.34 .26 .23 .17

TABLE 7.8-1: FAILURE MODE DISTRIBUTION OF PARTS (CONT'D)

DEVICE TYPE	FAILURE MODE	MODE PROBABILITY (α)
Microwave Amplifier	No Output Limited Voltage Gain	.90 .10
Microwave, Connector	High Insertion Loss Open	.80 .20
Microwave Detector	Power Loss No Output	.90 .10
Microwave, Diode	Open Parameter Change Short	.60 .28 .12
Microwave Filter	Center Frequency Drift No Output	.80 .20
Microwave Mixer	Power Decrease Loss of Intermediate Frequency	.90 .10
Microwave Modulator	Power Loss No Output	.90 .10
Microwave Oscillator	No Output Untuned Frequency Reduced Power	.80 .10 .10
Microwave VCO	No Output Untuned Frequency Reduced Power	.80 .15 .05
Optoelectronic LED	Open Short	.70 .30
Optoelectronic Sensor	Short Open	.50 .50
Power Supply	No Output Incorrect Output	.52 .48
Printed Wiring Assembly	Open Short	.76 .24
Pump, Centrifugal	No Output Degraded Output	.67 .33
Pump, Hydraulic	Leaking Improper Flow No Flow	.82 .12 .06
Relay	Fails to Trip Spurious Trip Short	.55 .26 .19

TABLE 7.8-1: FAILURE MODE DISTRIBUTION OF PARTS (CONT'D)

DEVICE TYPE	FAILURE MODE	MODE PROBABILITY (α)
Resistor, Composition	Parameter Change Open Short	.66 .31 .03
Resistor, Film	Open Parameter Change Short	.59 .36 .05
Resistor, Wirewound	Open Parameter Change Short	.65 .26 .09
Resistor, Network	Open Short	.92 .08
Resistor, Variable	Open Erratic Output Short	.53 .40 .07
Rotary Switch	Improper Output Contact Failure	.53 .47
Software	Design Changes Design Errors User Error Documentation Error	.46 .41 .07 .06
Solenoid	Short Slow Movement Open	.52 .43 .05
Switch, Push-button	Open Sticking Short	.60 .33 .07
Switch, Thermal	Parameter Change Open No Control Short	.63 .27 .08 .02
Switch, Toggle	Open Sticking Short	.65 .19 .16
Synchro	Winding Failure Bearing Failure Brush Failure	.45 .33 .22

TABLE 7.8-1: FAILURE MODE DISTRIBUTION OF PARTS (CONT'D)

DEVICE TYPE	FAILURE MODE	MODE PROBABILITY (α)
Transducer	Out of Tolerance False Response Open Short	.68 .15 .12 .05
Transformer	Open Short Parameter Change	.42 .42 .16
Transistor, Bipolar	Short Open	.73 .27
Transistor, FET	Short Output Low Parameter Change Open Output High	.51 .22 .17 .05 .05
Transistor, GaAs FET	Open Short Parameter Change	.61 .26 .13
Transistor, R.F.	Parameter Change Short Open	.50 .40 .10
Tube, Traveling Wave	Reduced Output Power High Helix Current Gun Failure Open Helix	.71 .11 .09 .09
Valve, Hydraulic	Leaking Stuck Closed Stuck Open	.77 .12 .11
Valve, Pneumatic	Leaking Stuck Open Stuck Closed Spurious Opening Spurious Closing	.28 .20 .20 .16 .16
Valve, Relief	Premature Open Leaking	.77 .23

In order to accurately address the failure modes of a given LSI microcircuit each of these factors must be accounted for. As an example, if the IC chip is packaged in an hermetic cavity package there is a possibility that one wire may break and short to an adjacent wire. If this same chip were encapsulated in a plastic package, this short could not occur, since the wire is constrained by the potting material. However, the potting material can have other detrimental effects on an IC chip.

Figure 7.8-3 illustrates a useful form for conducting a failure effect analysis. (See also Figure 7.8-2 for an example of its use.) For each component in the system, appropriate information is entered in each column. Column descriptions are given in Table 7.8-2.

(1) ITEM	(2) CODE	(3) FUNCTION	(4) FAILURE MODE	(5) FAILURE EFFECT	(6) LOSS PROBABILITY, β

FIGURE 7.8-3: FAILURE EFFECTS ANALYSIS FORM

TABLE 7.8-2: COLUMN DESCRIPTIONS FOR FIGURE 7.8-3

COLUMN	NOMENCLATURE	DESCRIPTION
1	Item	Item name
2	Code	Item identification or circuit designation code
3	Function	Concise statement of the item's function
4	Failure Mode	Concise statement of the mode(s) of item failure
5	Failure Effect	Explanation of the effect of each failure mode on the performance of the next higher level in the symbolic logic block diagram
6	Loss Probability	Numerical value indicating the probability of system loss if the item fails in the mode indicated

A numerical reference for all items in the symbolic logic block diagram must be provided by using a standard coding system, such as that specified in MIL-STD-1629. All items below the set and group levels are identified using the scheme illustrated in Table 7.8-2. Items at and above the group and set levels are not subject to this standard nomenclature scheme. These items can be assigned a simple code such as that illustrated in Figure 7.8-1. In this illustration, the system is assigned a letter; and the subsystems, sets, and groups are assigned numbers in a specifically ordered sequence. As an example, the code S-23-01 designates the first group of the third set in the second subsystem of system S (Note, this code is limited to subsystems with less than 10 sets). The exact coding system used is not as important as making sure that each block in the diagram has its own number. Identical items (same drawing numbers) in different systems, or in the same system but used in different applications, should not be assigned the same code number.

(1) During the failure effects analysis, a number of changes to the block diagrams may be required. Therefore, to minimize the number of changes in the coding system, it is recommended that the failure effects analysis be completed before assignment of code numbers is finalized.

(2) Based on the failure effects analysis, a list of critical items should be prepared. This list will contain those items whose failure results in a possible loss, probable loss, or certain loss of the next higher level in the symbolic logic block diagram. All items that can cause system loss should be identified clearly in the list.

7.8.3 Phase 2

This phase is implemented by performing the following steps:

(1) The symbolic logic block diagram, failure effects analysis, coding, and critical items list are reviewed and brought up- to-date.

(2) Criticality is assigned, based on the item applicable failure mode, the system loss probability, the failure mode frequency ratio, and the item unreliability. The analysis of criticality is essentially quantitative, based on a qualitative failure effects analysis.

Criticality CR_{ij} defined by the equation

$$(CR)_{ij} = \alpha_{ij} \beta_{ij} \lambda_i \qquad (7.21)$$

where:

α_{ij} = failure mode frequency ratio of item i for the failure mode j (see Table 7.8-1 for an example), i.e., the ratio of failures of the type being considered to all failures of the item.

β_{ij} = loss probability of item i for failure mode j (i.e., the probability of system failure if the item fails). A suggested scale is Certain Loss = 1.00, Probable Loss ranges from 0.1 to 1.0, Possible Loss ranges from 0 to 0.10, No Effect - 0.0

λ_i = failure rate of item i

$(CR)_{ij}$ = system failure rate due to item i's failing in its mode j

The system criticality is given by Eq. (7.22)

$$(CR)_s = \sum_i \sum_j (CR)_{ij} \tag{7.22}$$

where:

$(CR)_s$ = system criticality (failure rate)

Σ_j = sum over all failure modes of item i

Σ_i = sum over all items

A form useful for conducting the criticality analysis is given in Figure 7.8-5. This form is a modification of Figure 7.8-3 to include the failure mode frequency ratio and the failure rate. The example in the next section and Figures 7.8-4 and 7.8-5 illustrate the procedure.

The CR value of the preamplifier unit is 5.739 per 10^6 hr. This number can be interpreted as the predicted total number of system failures per hour due to preamplifier failures, e.g., 5.739 x 10^{-6}. Whether or not this number is excessive, and thus calls for corrective action, depends upon the requirements for the system and the criticalities for other units in the system. If the number is excessive, it can be reduced by any of the following actions:

(1) Lowering the failure rates of parts in the system by derating.

(2) Decreasing the failure mode frequency ratio through selection of other parts.

(3) Decreasing the loss probability by changing the system or preamplifier design.

(4) Redesign using various techniques such as redundancy, additional cooling, or switching.

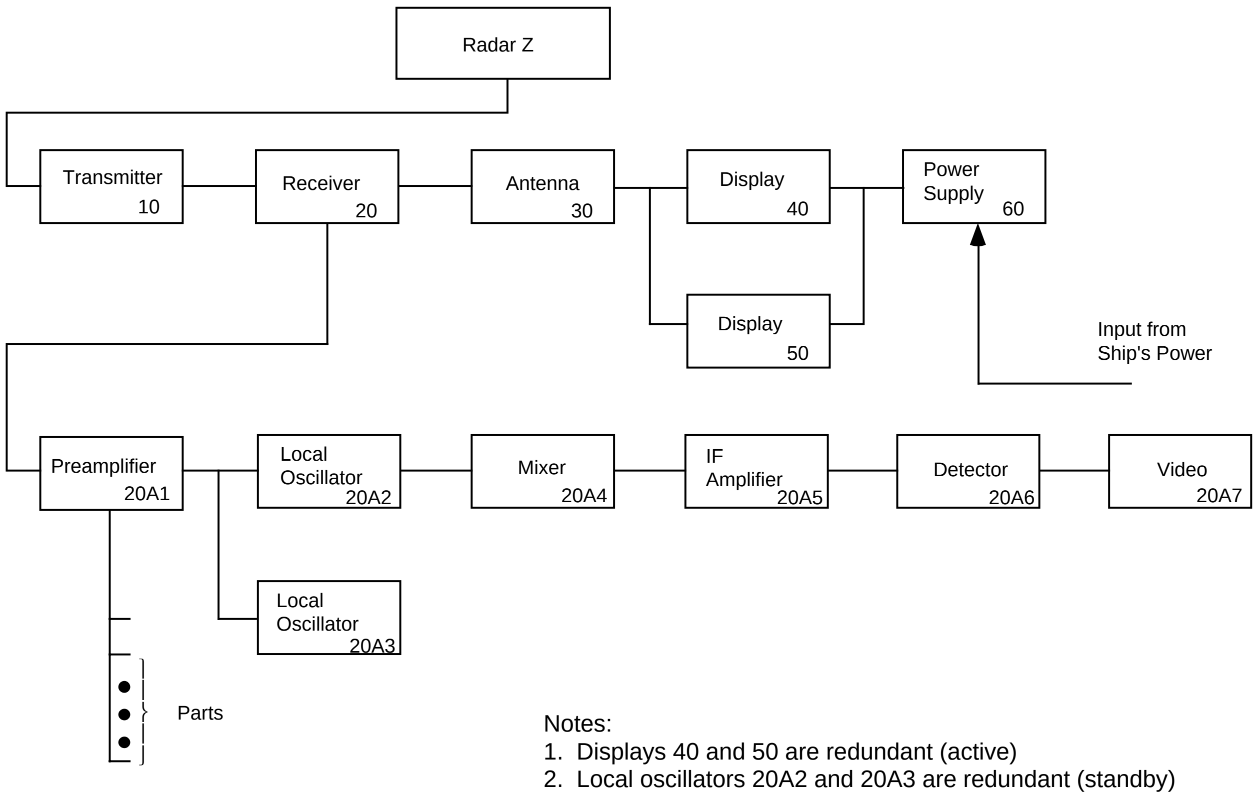


FIGURE 7.8-4: SYMBOLIC LOGIC DIAGRAM OF RADAR EXAMPLE

7.8.4 Example

The detail design of a radar system required the use of FMEA to determine the effect of item failures on the system. The FMEA analysis must be performed at this time prior to freezing the design. Perform an FMEA analysis as follows:

PROCEDURE	EXAMPLE
(1) Develop a symbolic logic block diagram of the radar system. The units making up the receiver subsystem are shown in detail. In an actual analysis, symbolic diagrams must be constructed for all other sub-systems.	See Figure 7.8-4
(2) Fill in the work sheets for all units in the receiver subsystem. Repeat this procedure for all subsystems.	See Figure 7.8.-5
(3) Qualitatively estimate the values of loss probability for each part.	An analysis indicates that for this system the following values of β are applicable: 1.0, 0.1, and 0.0.
(4) Determine the failure mode frequency ratio for each failure mode of every part.	The resistor is fixed, film (Fig. 7.8-5); from Table 7.8-1, it has two failure modes: open = 0.59 and drift = 0.36.
(5) Tabulate failure rates for each component.	λ (20A1R1) = 1.5 per 10^6 hr.
(6) Compute the CR value for each failure mode of each part by Eq. (7.21).	CR(open) = 0.59 x 1.00 x 1.5 per 10^6 hr = 0.885 per 10^6 hr CR (short) = 0.05 x 1.00 x 1.5 per 10^6 hr = 0.075 per 10^6 hr CR(parameter change) = 0.36 x 10^6 x 0.10 x 1.5 per 10^6 hr = 0.054 per 10^6 hr
(7) Compute the total CR for the unit (CR), by Eq. (7.22).	The total CR for the preamplifier unit is 5.739 per 10^6 hr (See Figure 7.8-5).

CRITICALITY WORK SHEET SYSTEM Radar (Z) SUBSYSTEM Receiver 20 UNIT Preamplifier 20A1 Parts

(1) Item	(2) Code	(3) Function	(4) Failure Effect	(5) Failure Effect	(6) Loss Probability (β)	(7) Failure Mode Frequency Ratio (α)	(8) Failure Rate (Per Million Hours (λ)	(9) Criticality (CR)	(10) Comments
Resistor	R1	Voltage Divider	Open	No Output	1.00	0.59	1.50	0.885	Film Resistor
Resistor	R1	Voltage Divider	Short	No Output	1.00	0.05	1.50	0.075	Film Resistor
Resistor	R1	Voltage Divider	Parameter Change	Wrong Output	0.10	0.36	1.50	0.054	Film Resistor
Resistor	R2	Voltage Divider	Open	No Output	1.00	0.59	1.50	0.885	Film Resistor
Resistor	R2	Voltage Divider	Short	No Output	1.00	0.05	1.50	0.075	Film Resistor
Resistor	R2	Voltage Divider	Parameter Change	Wrong Output	0.10	0.36	1.50	0.054	Film Resistor
Capacitor	C3	Decoupling	Open	No Effect	0.00	0.32	0.22	0.000	Tubular Tantalum
Capacitor	C3	Decoupling	Short	No Output	1.00	0.57	0.22	0.125	Tubular Tantalum
Capacitor	C3	Decoupling	Parameter Change	No Effect	0.00	0.11	0.22	0.000	Tubular Tantalum
Diode	CR3	Voltage Divider	Open	No Output	1.00	0.24	1.00	0.240	Small Signal
Diode	CR3	Voltage Divider	Short	No Output	1.00	0.18	1.00	0.180	Small Signal
Diode	CR3	Voltage Divider	Parameter Change	Wrong Output	0.10	0.58	1.00	0.058	Small Signal
Transistor	Q4	Amplifier	Open	No Output	1.00	0.10	3.00	0.300	R.F.
Transistor	Q4	Amplifier	Short	No Output	1.00	0.40	3.00	1.200	R.F.
Transistor	Q4	Amplifier	Parameter Change	Wrong Output	0.10	0.50	3.00	0.150	R.F.
Transformer	T5	Coupling	Open	No Output	1.00	0.42	3.00	1.260	
Transformer	T5	Coupling	Shorted	Wrong Output	0.10	0.42	3.00	0.126	
Transformer	T5	Coupling	Parameter Change	Wrong Output	0.10	0.15	3.00	0.045	
Resistor	R6	Bias	Open	No Output	1.00	0.31	0.006	0.002	Composition
Resistor	R6	Bias	Short	Wrong Output	0.10	0.03	0.006	0.000	Composition
Resistor	R6	Bias	Parameter Change	No Effect	0.00	0.66	0.006	0.000	Composition
Capacitor	C7	Bypass	Open	No Effect	0.00	0.35	0.48	0.000	Aluminum
Capacitor	C7	Bypass	Short	Wrong Output	0.10	0.53	0.48	0.025	Electrolytic
Capacitor	C7	Bypass	Parameter Change	No Effect	0.00	0.02	0.48	0.000	Capacitor
Capacitor	C7	Bypass	Electrolyte Leakage	No Effect	0.00	0.10	0.48	0.000	

CRITICALITY TOTAL FOR UNIT 5.739

FIGURE 7.8-5: DETERMINATION OF PREAMPLIFIER CRITICALITY

7.8.5 Risk Priority Number

An alternate evaluation approach to Criticality Analysis is that of the calculation of a Risk Priority Number (RPN). The risk priority number provides a qualitative numerical estimate of the design risk. This number is then used to rank order the various concerns and failure modes associated with a given design as previously identified in the FMEA. RPN is defined as the product of three independently assessed factors: Severity (S), Occurrence (O) and Detection (D).

RPN = (S) x (O) x (D)

This technique was originally developed for use by the automotive industry, but it may also be effectively tailored to many other types of applications. A more detailed description of this technique may be found in Reference [80].

A description, and one detailed example, of each of these three independently assessed factors (S), (O), and (D) follows.

SEVERITY (S) is an assessment of the seriousness of the effect of the potential failure mode to the next higher component, subsystem, system or to the customer if it were to occur. Severity is typically estimated on a scale of “1” to “10”. One such method of ranking is illustrated in Table 7.8-3. This table could be appropriately tailored for other non-automotive applications. Severity applies only to the effect of the failure.

OCCURRENCE (O) is the likelihood that a specific cause/mechanism will occur. The likelihood of occurrence ranking number is an index number rather than a probability. Removing or controlling one or more of the causes/mechanisms of the failure mode through a design change is the only way a reduction in occurrence ranking can be effected.

The likelihood of occurrence of potential failure cause/mechanism is typically estimated on a scale of “1” to “10”. One such method of ranking is illustrated in Table 7.8-4.

TABLE 7.8-3: SEVERITY CLASSIFICATION

EFFECT	CRITERIA: SEVERITY OF EFFECT	RANKING
Hazardous - without warning	Very high severity ranking when a potential failure mode affects safe vehicle operation and/or involves noncompliance with government regulation without warning.	10
Hazardous - with warning	Very high severity ranking when a potential failure mode affects safe vehicle operation and/or involves noncompliance with government regulation with warning.	9
Very High	Vehicle / item inoperable, with loss of primary function.	8
High	Vehicle / item operable, but at reduced level of performance. Customer dissatisfied.	7
Moderate	Vehicle / item operable, but Comfort / Convenience item(s) inoperable. Customer experiences discomfort.	6
Low	Vehicle / item operable, but Comfort / Convenience item(s) operate at a reduced level of performance. Customer experiences some dissatisfaction.	5
Very Low	Fit & Finish / Squeak & Rattle item does not conform. Defect noticed by most customers.	4
Minor	Fit & Finish / Squeak & Rattle item does not conform. Defect noticed by average customer.	3
Very Minor	Fit & Finish / Squeak & Rattle item does not conform. Defect noticed by discriminating customer.	2
None	No Effect	1

TABLE 7.8-4: OCCURRENCE RANKING

PROBABILITY OF FAILURE	POSSIBLE FAILURE RATES	RANKING
Very High: Failure is almost inevitable	≥ 1 in 2 1 in 3	10 9
High: Repeated failures	1 in 8 1 in 20	8 7
Moderate: Occasional failures	1 in 80 1 in 400 1 in 2000	6 5 4
Low: Relatively few failures	1 in 15,000 1 in 150,000	3 2
Remote: Failure is unlikely	≤ 1 in 1,500,000	1

In determining this estimate, questions such as the following should be considered:

(1) What is the service history/field experience with similar components or subsystems?

(2) Is this component carried over from, or similar to, a previously used component or subsystem?

(3) How significant are the changes from a previously used component or subsystem?

(4) Is the component radically different from a previously used component?

(5) Is the component new?

(6) Has the component application changed?

(7) What, if any, are the environmental changes?

(8) Has an engineering analysis been made to estimate the expected comparable occurrence rate for this application?

A consistent occurrence ranking system should be used to ensure continuity. The “Design Life Possible Failure Rates” shown in Table 7.8-4 are based upon the number of failures which are anticipated during the design life of the component, subsystem, or system. The occurrence ranking number is related to the rating scale and does not reflect the actual likelihood of occurrence.

CURRENT DESIGN CONTROLS: This is an additional parameter of concern beyond those previously addressed in the FMEA. Current Design Controls are defined as prevention, design verification/validation, or other activities which will assure the design adequacy for the failure mode and/or cause/mechanism under consideration. Current controls (e.g., road testing, design reviews, fail-safe analysis, mathematical studies, rig/lab testing, feasibility reviews, prototype tests, fleet testing etc.) are those that have been or are being used with the same or similar designs.

There are three types of Design Controls/Features to consider; those that: (1) Prevent the cause/mechanism or failure mode/effect from occurring, or reduce their rate of occurrence, (2) detect the cause/mechanism and lead to corrective actions, and (3) detect the failure mode.

The preferred approach is to first use type (1) controls if possible; second, use the type (2) controls; and third, use the type (3) controls. The initial detection ranking will be based upon the type (2) or type (3) current controls, provided the prototypes and models being used are representative of design intent.

DETECTION (D) is an assessment of the ability of the proposed type (2) current design controls, to detect a potential cause/mechanism (design weakness), or the ability of the proposed type (3) current design controls to detect the subsequent failure mode, before the component, subsystem, or system is released for production. In order to achieve a lower detection ranking, generally the planned design control (e.g. preventative, validation, and/or verification activities) has to be improved.

The detection of potential failure cause/mechanism is typically estimated on a scale of “1” to “10”. One such method of ranking is illustrated in Table 7.8-5.

TABLE 7.8-5: DETECTION RANKING

DETECTION	CRITERIA: LIKELIHOOD OF DETECTION BY DESIGN CONTROL	RANKING
Absolute Uncertainty	Design Control will not and/or can not detect a potential cause/ mechanism and subsequent failure mode; or there is no Design Control.	10
Very Remote	Very remote chance the Design Control will detect a potential cause/ mechanism and subsequent failure mode.	9
Remote	Remote chance the Design Control will detect a potential cause/ mechanism and subsequent failure mode.	8
Very Low	Very low chance the Design Control will detect a potential cause/ mechanism and subsequent failure mode.	7
Low	Low chance the Design Control will detect a potential cause/ mechanism and subsequent failure mode.	6
Moderate	Moderate chance the Design Control will detect a potential cause/ mechanism and subsequent failure mode.	5
Moderately High	Moderately high chance the Design Control will detect a potential cause/mechanism and subsequent failure mode.	4
High	High chance the Design Control will detect a potential cause/ mechanism and subsequent failure mode.	3
Very High	Very high chance the Design Control will detect a potential cause/ mechanism and subsequent failure mode.	2
Almost Certain	Design Control will almost certainly detect a potential cause/ mechanism and subsequent failure mode.	1

7.8.5.1 Instituting Corrective Action

When the failure modes have been rank ordered by RPN (the product of S, O and D), corrective action should be first directed at the highest ranked concerns and critical items. The intent of any recommended action is to reduce any one or all of the occurrence, severity and/or detection rankings. An increase in design validation/verification actions will result in a reduction in the detection ranking only. A reduction in the occurrence ranking can be effected only by removing or controlling one or more of the causes/mechanisms of the failure mode through a design revision. Only a design revision can bring about a reduction in the severity ranking. Regardless of the resultant RPN, special attention should be given when severity is high.

After the corrective action(s) have been identified, estimate and record the resulting severity, occurrence, and detection rankings and recalculate the RPN.

7.8.6 Computer Aided FMEA

As with most other reliability analyses the computer can be quite helpful in performing an FMEA, since a large number of computations and a significant amount of record keeping are required for systems of reasonable size.

In the failure effects portion of the analysis the computer is very helpful for functional evaluation, using performance models. Given that the computer program contains the design equations relating system outputs to various design parameters, each item is allowed to fail in each of its modes, and the effect on the system is computed.

Several commercial programs are available for performing an FMECA in accordance with MIL-STD-1629A.

7.8.7 FMEA Summary

The FMEA does not replace the need for sound engineering judgment at the design level. This system analysis is, however, practical in determining many of the significant details which may not otherwise be determined by separate, individual studies. Like other design tools, the FMEA has limitations such as those discussed below.

(1) It is not a substitute for good design. If used for its intended purpose it can be an aid to better design.

(2) It will not solve item problems which may exist as a limitation to effective system design. It should define and focus attention on such problems and indicate the need for a design solution.

(3) It will not, in itself, guarantee a system design. It is nothing more than a logical way of establishing "bookkeeping" which can be systematically analyzed for design reliability.

7.9 Fault Tree Analysis

The "fault tree" analysis (FTA) technique is a method for block diagramming constituent lower level elements. It determines, in a logical way, which failure modes at one level produce critical failures at a higher level in the system. The technique is useful in safety analysis where the discipline of block diagramming helps prevent an oversight in the basic FMEA discussed in the previous subsection.

As was previously mentioned, FMEA is considered a "bottoms up" analysis, whereas an FTA is considered a "top down" analysis. FMEAs and FTAs are compatible methods of risk analysis, with the choice of method dependent on the nature of the risk to be evaluated. There are some differences, however, because FTA is a top down analysis there is a higher probability of misinterpretation at the lowest level. On the other hand, FMEA starts at the lowest level, therefore will probably result in a better method of risk analysis (assuming lowest level data is available). Also, FMEA considers only single failures while FTA considers multiple failures. In general, FTA requires a greater skill level than FMEA.

Fault tree methods of analysis are particularly useful in functional paths of high complexity in which the outcome of one or more combinations of noncritical events may produce an undesirable critical event. Typical candidates for fault tree analysis are functional paths or interfaces which could have critical impact on flight safety, munitions handling safety, safety of operating and maintenance personnel, and probability of error free command in automated systems in which a multiplicity of redundant and overlapping outputs may be involved. The fault tree provides a concise and orderly description of the various combinations of possible occurrences within the system which can result in a predetermined critical output event. However, performance of the fault tree analysis does require considerable engineering time and even then the quality of results is only as good as the validity of input data and accuracy of the fault tree logic.

Fault tree methods can be applied beginning in the early design phase, and progressively refined and updated to track the probability of an undesirable event as the design evolves. Initial fault tree diagrams might represent functional blocks (e.g., units, equipments, etc.), becoming more definitive at lower levels as the design materializes in the form of specific parts and materials. Results of the analysis are useful in the following applications:

(1) Allocation of critical failure mode probabilities among lower levels of the system breakdown.

(2) Comparison of alternative design configurations from a safety point of view.

(3) Identification of critical fault paths and design weaknesses for corrective action.

(4) Evaluation of alternative corrective action approaches.

(5) Development of operational, test, and maintenance procedures to recognize and accommodate unavoidable critical failure modes.

Symbols commonly used in diagramming a fault tree analysis are shown in Figure 7.9-1. The basic relationships between functional reliability (success) block diagrams and the equivalent fault tree diagrams, using some of these symbols, are illustrated in Figures 7.9-2 and 7.9-3.

Success of the simple two element series system comprised of blocks A and B is given by $R = AB$; and the probability of system failure (i.e., unsuccessful or unsafe performance) is given by $\overline{R} = (1 - R) = 1 - AB$. When individual element unreliability ($\overline{R}_i$) is less than 0.1, the following approximations may be used to simplify computations in the fault tree logic diagram, with little (10%) error:

$$\overline{R} = 1 - AB = 1 - (1 - \overline{A})(1 - \overline{B})$$
$$= \overline{A} + \overline{B} - \overline{AB} \approx \overline{A} + \overline{B}$$

The two element block diagrams of Figure 7.9-2 is reconfigured as a simple parallel redundant system in Figure 7.9-3 to illustrate the treatment of parallel redundant elements in the fault tree logic diagram. Note that "AND" gates for the combination of successes (R_s) become "OR" gates for the combination of failures ($\overline{R}_s$); and "OR" gates for R_s become "AND" gates for $\overline{R}_s$. This is illustrated in the series parallel network of Figure 7.9-3.

The fault tree analysis of critical failure modes should proceed as illustrated in the following steps.

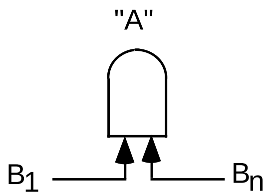
Step 1: Develop Function Reliability Block Diagram: Develop reliability block diagram for the system/equipment functional paths in which the critical failure mode is to be circumvented or eliminated. Define the critical failure mode in terms of the system level mal-performance symptom to be avoided. For example, the hypothetical firing circuit of Figure 7.9-4 is designed to ignite a proposed rocket motor in the following sequence:

(1) Shorting switch S_1 is opened to enable launcher release and firing.

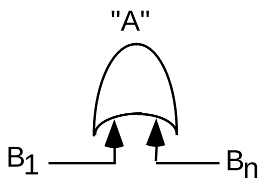
(2) Firing switch S_2 is closed by the pilot to apply power to relay R_1

(3) Relay R_1 activates the guidance and control (G&C) section.

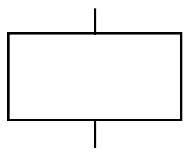
(4) Relay R_2 is activated by signal from the G&C section, closing the igniter firing circuit which starts the rocket motor.



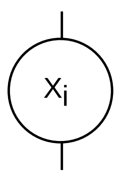
A logical "AND" gate - "A" exists if and only if all of B_1, B_2 . . . . B_n exist simultaneously.



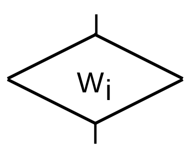
A logical inclusive "OR" gate - "A" exists if any B_1, B_2, . . . B_n or any combination thereof



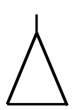
An event--usually the output of (or input to) and "AND" or an "OR" gate



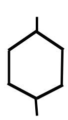
A failure rate of malfunction event-in terms of a specific circuit or component, represented by the symbol X with a numerical subscript



An event not developed further because of lack of information or because of lack of sufficient consequence. Represented by the symbol W with a numerical subscript



A connecting symbol to another part of the fault tree within the same major branch



An "inhibit" gate, used to describe the relationship between one fault and another. The input fault directly produces the output fault if the indicated conditions is satisfied

FIGURE 7.9-1: FAULT TREE ANALYSIS SYMBOLS

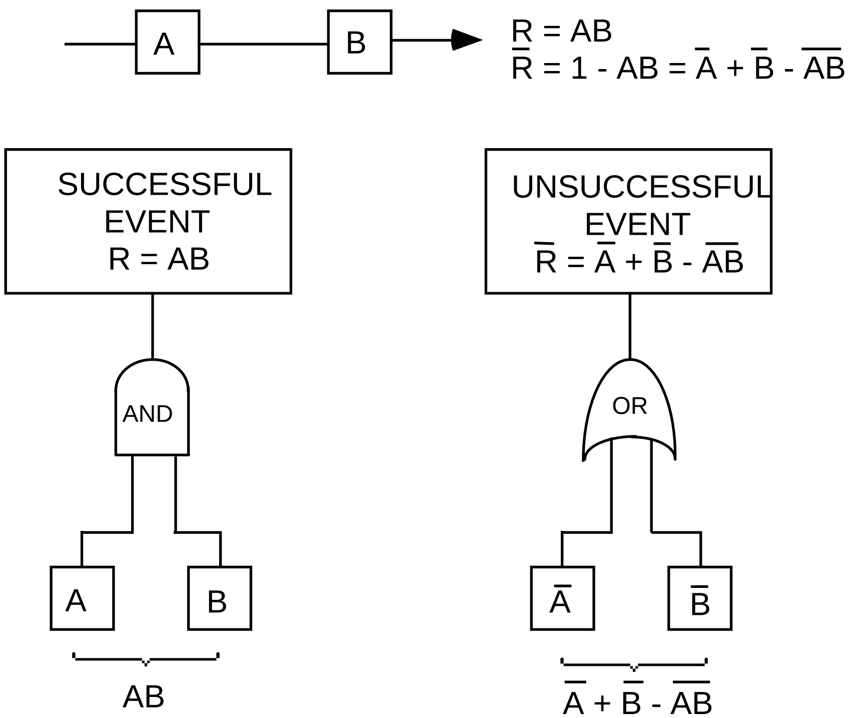


FIGURE 7.9-2: TRANSFORMATION OF TWO-ELEMENT SERIES RELIABILITY BLOCK DIAGRAM TO "FAULT TREE" LOGIC DIAGRAMS

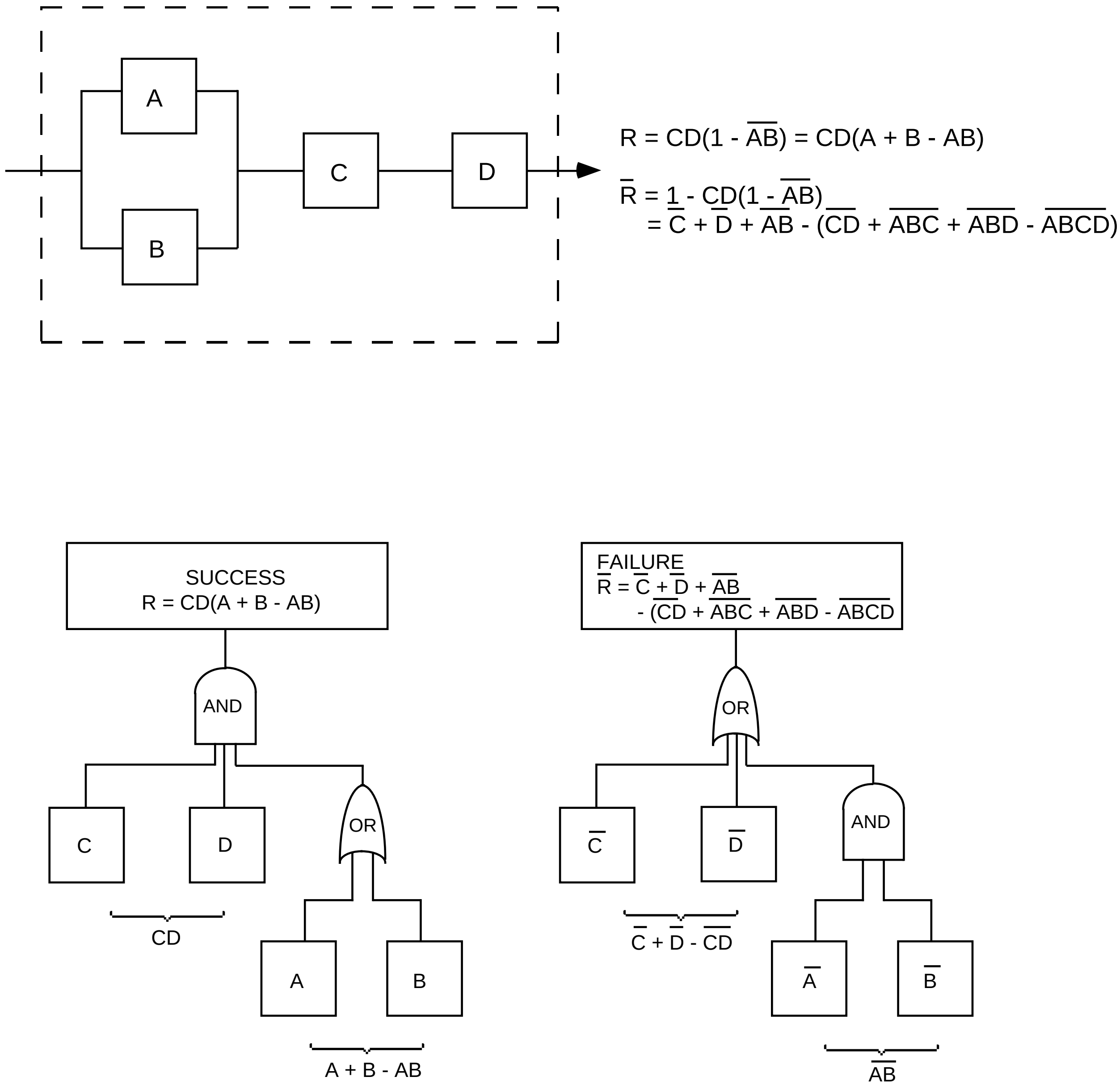


FIGURE 7.9-3: TRANSFORMATION OF SERIES/PARALLEL BLOCK DIAGRAM TO EQUIVALENT FAULT TREE LOGIC DIAGRAM

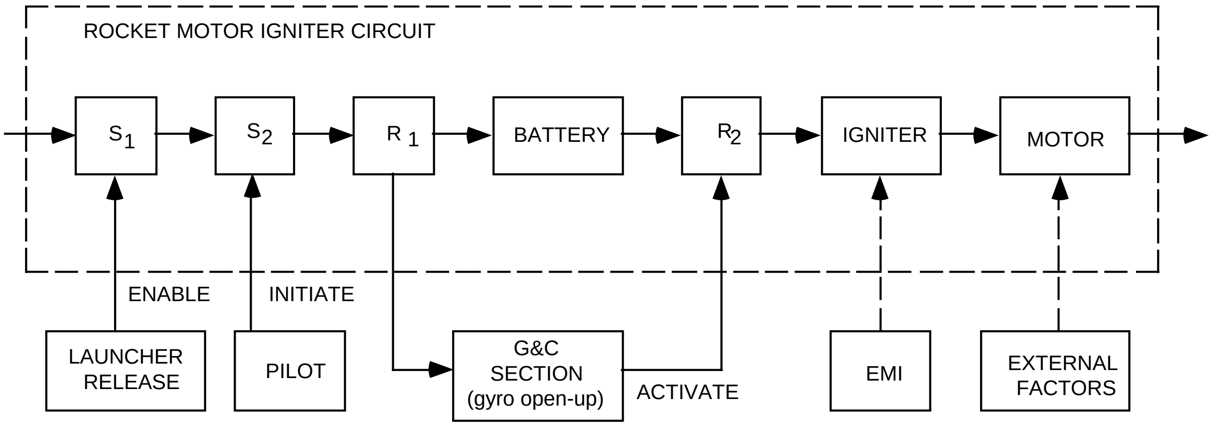


FIGURE 7.9-4: RELIABILITY BLOCK DIAGRAM OF HYPOTHETICAL ROCKET MOTOR FIRING CIRCUIT

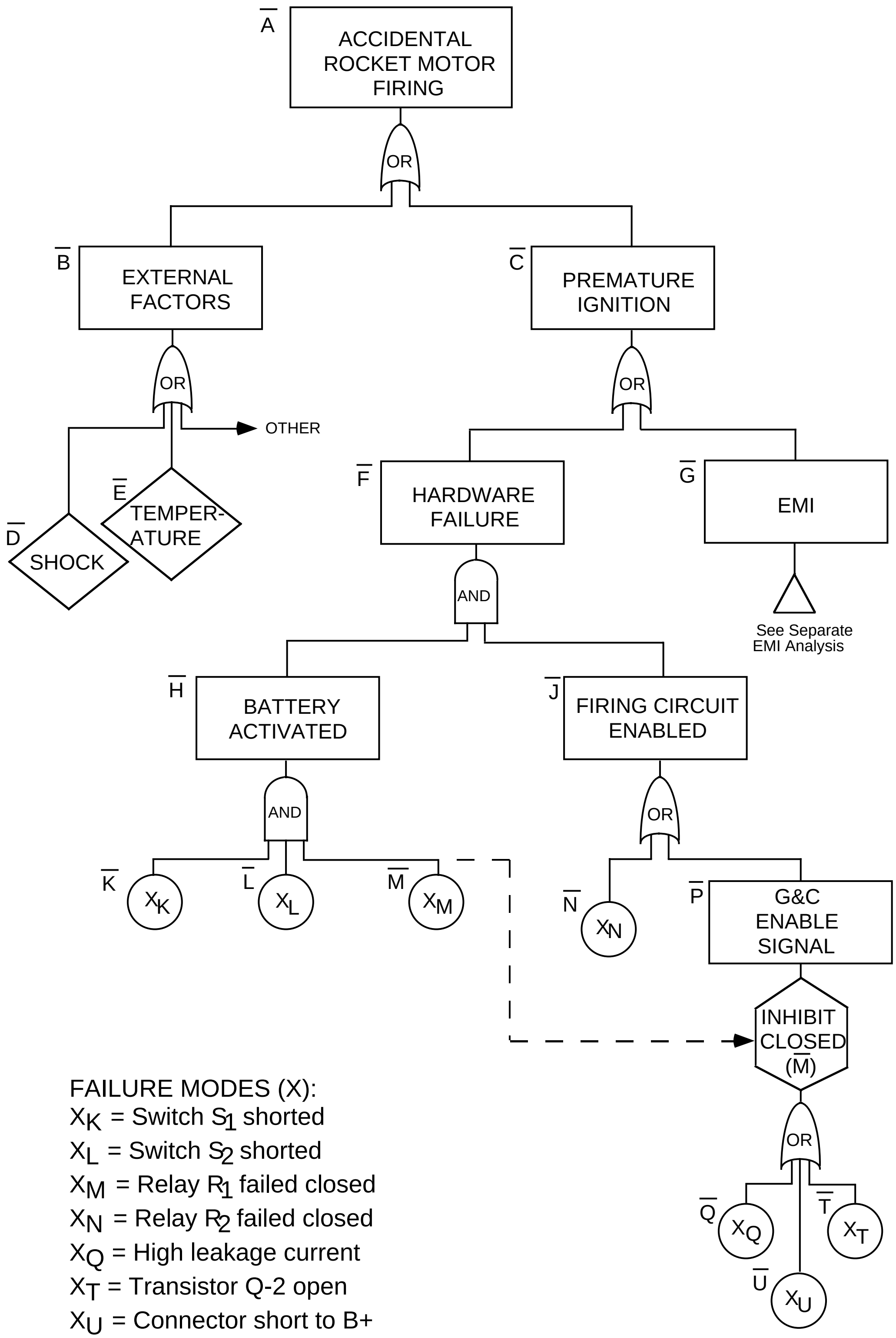
The rocket motor can be inadvertently fired by premature ignition due to electronic failure, electromagnetic interference (EMI), or by external factors such as shock, elevated temperature, etc. These are the events to be studied in the fault tree analysis.

Step 2: Construct the Fault Tree: Develop the fault tree logic diagram relating all possible sequences of events whose occurrence would produce the undesired events identified in Step 1, e.g., inadvertent firing of the missile rocket motor. The fault tree should depict the paths that lead to each succeeding higher level in the functional configuration. Figure 7.9-5 illustrates the construction of one branch of the fault tree for the ignition circuit.

In constructing the fault tree for each functional path or interface within the reliability model, consideration must be given to the time sequencing of events and functions during the specified mission profile. Very often the operational sequence involves one or more changes in hardware configuration, functional paths, critical interfaces, or application stresses. When such conditions are found to apply, it is necessary to develop a separate fault tree for each operating mode, function, or mission event in the mission sequence.

Step 3: Develop Failure Probability Model: Develop the mathematical model of the fault tree for manual (or computer) computation of the probability of critical event occurrence on the basis of failure modes identified in the diagram. For example, the undesired system level critical failure mode identified in Figure 7.9-5 is "accidental rocket motor firing," given by the top level model as follows:

$$\overline{A} = \overline{B} + \overline{C} - \overline{BC}$$



FAILURE MODES (X):
X_K = Switch S_1 shorted
X_L = Switch S_2 shorted
X_M = Relay R_1 failed closed
X_N = Relay R_2 failed closed
X_Q = High leakage current
X_T = Transistor Q-2 open
X_U = Connector short to B+

FIGURE 7.9-5: FAULT TREE FOR SIMPLIFIED ROCKET MOTOR FIRING CIRCUIT

As indicated in the figure, C represents the probability of accidental rocket motor firing due to premature ignition via the firing circuit either due to hardware failure (F) or electromagnetic interference (G), i.e.:

$$\overline{C} = \overline{F} + \overline{G} - \overline{FG}$$

Considering hardware failures only, the probability of premature ignition due to hardware failure is given by:

$$\overline{F} = \overline{HJ}$$

where:

$$\overline{H} = \overline{KLM}$$

$$\overline{J} = \overline{N} + \overline{P} - \overline{NP}$$

$$\overline{P} = \overline{Q} + \overline{T} + \overline{U} = (\overline{QT} + \overline{QU} + \overline{TU} - \overline{QTU})$$

Step 4: Determine Failure Probabilities or Identified Failure Modes: Determine probability of occurrence (i.e., probability of failure) in each event or failure mode identified in the model. Compute safety parameters at the system level by applying the failure data in the models derived in Step 3.

Assume, for example, the following failure probabilities in the premature ignition branch of the fault tree:

$$\overline{K} = 50 \times 10^{-3}$$

$$\overline{L} = 100 \times 10^{-3}$$

$$\overline{M} = 40 \times 10^{-3}$$

$$\overline{N} = 5 \times 10^{-3}$$

$$\overline{Q} = 2 \times 10^{-3}$$

$$\overline{T} = 1 \times 10^{-3}$$

$$\overline{U} = 0.5 \times 10^{-3}$$

Using the bottom up approach, combine these data in the failure probability models developed in Step 3, and estimate the system level probability as follows:

$$
\begin{aligned}
\overline{P} &= \overline{Q} + \overline{T} + \overline{U} - (\overline{QT} + \overline{QU} + \overline{TU} - \overline{QTU}) \\
&= (2 + 1 + 0.5)10^{-3} - [(2 + 1 + 0.5)10^{-6} - (1)10^{-9}] \\
&\approx 3.5 \times 10^{-3}
\end{aligned}
$$

Higher order (product) terms in the model can be dropped in the P model since the values of individual terms are much less than 0.10.

Combining $\overline{P}$ with $\overline{N}$ to find $\overline{J}$ yields:

$$
\begin{aligned}
\overline{J} &= \overline{N} + \overline{P} - \overline{NP} \\
&= 5 \times 10^{-3} + 3.5 \times 10^{-3} - 17.5 \times 10^{-6} \\
&\approx 8.5 \times 10^{-3}
\end{aligned}
$$

This is the probability of accidental firing circuit operation conditional on relay R_1 having failed in the closed position (i.e., M) in the battery branch of the fault tree. In the battery branch, the battery can be accidentally activated only if switches S_1 and S_2 fail in the short mode, and if relay R_1 fails in the closed position, given by:

$$
\begin{aligned}
\overline{H} &= \overline{KLM} \\
&= (50 \times 10^{-3})(100 \times 10^{-3})(40 \times 10^{-3}) \\
&= 200 \times 10^{-6}
\end{aligned}
$$

Probability of premature ignition because of hardware failure is then estimated from:

$$
\begin{aligned}
\overline{F} &= \overline{HJ} = (200 \times 10^{-6})(8.5 \times 10^{-3}) \\
&= 1.70 \times 10^{-6}
\end{aligned}
$$

Assume that the EMI analysis discloses a probability of accidental ignition ($\overline{G}$ = 5 x 10^{-6}) due to exposure to specified level of RF radiation in the operating environment. The probability of premature ignition to either cause (hardware failure or EMI exposure) is given by:

$$\overline{C} = \overline{F} + \overline{G} - \overline{FG}$$
$$\approx (1.70 \times 10^{-6}) + (5 \times 10^{-6}) - (1.70 \times 10^{-6})(5 \times 10^{-6})$$
$$\approx 6.70 \times 10^{-6}$$

Assume that failure data accrued during rocket motor qualification tests indicates $\overline{D} = 2.5 \times 10^{-6}$ and $\overline{E} = 12.5 \times 10^{-6}$ under specified conditions and levels of exposure. Under these circumstances,

$$\overline{B} = \overline{D} + \overline{E} - \overline{DE}$$
$$= (2.5 \times 10^{-6}) + (12.5 \times 10\text{-}6) - (2.5 \times 10^{-6})(12.5 \times 10^{-6})$$
$$\overline{B} = 15 \times 10^{-6}$$

Probability of accidental rocket motor firing during the handling and loading sequence is then:

$$\overline{A} = \overline{B} + \overline{C} - \overline{BC}$$
$$\approx (15 \times 10^{-6}) + (6.70 \times 10^{-6}) - (15 \times 10^{-6})(6.75 \times 10^{-6})$$
$$\approx 21.7 \times 10^{-6}$$

That is, approximately 22 premature rocket motor firings per million missile load/launch attempts.

Failure rate values for most standard electronic and electromechanical parts are available in MIL-HDBK-217. The most recent document for failure rate values for mechanical parts is Reference [14]. Failure rate data for new parts and more recently developed "high reliability" parts may not be available in these sources, however. In such cases, it becomes necessary to draw on vendor certified data or special tests.

In the absence of complete and validated failure rate/failure mode data for all inputs, a preliminary fault tree analysis can be performed using conservative estimates of failure rates in the critical failure modes. This preliminary analysis will identify those input values which have little effect, as well as those having a critical effect on system performance. The latter can then be investigated in depth by testing.

Evaluation of the fault tree model may reveal that the conservatively estimated values are sufficient to satisfy the performance goal. Other values will warrant further study. In some cases, it may even be more expedient to change the design than to validate a data value.

Step 5: Identify Critical Fault Paths: When the probability of an unsafe failure mode at the system level exceeds specification tolerances, identify the critical paths which contribute most significantly to the problem. For example, both paths in the preceding analysis contribute about equally to the total problem because of environmental sensitivity - ignition circuit to EMI, and propellant insulation to high ambient temperature.

7.9.1 Discussions of FTA Methods

There are basically three methods for solving fault trees: (1) direct simulation (Reference [81]), (2) Monte Carlo (Reference [82]), and (3) direct analysis (Reference [83]).

Direct simulation basically uses Boolean logic hardware (similar to that in digital computers) in a one-to-one correspondence with the fault tree Boolean logic to form an analog circuit. This method usually is prohibitively expensive. A hybrid method obtains parts of the solution using the analog technique and parts from a digital calculation, in an effort to be cost competitive. Because of the expense involved, this method rarely is used.

Monte Carlo methods are perhaps the most simplest in principle but in practice can be expensive. Since Monte Carlo is not practical without the use of a digital computer, it is discussed in that framework. The most easily understood Monte Carlo technique is called "direct simulation." The term "simulation" frequently is used in conjunction with Monte Carlo methods, because Monte Carlo is a form of mathematical simulation. (This simulation should not be confused with direct analog simulation.) Probability data are provided as input, and the simulation program represents the fault tree on a computer to provide quantitative results. In this manner, thousands or millions of trials can be simulated. A typical simulation program involves the following steps.

(1) Assign failure data to input fault events within the tree and, if desired, repair data.

(2) Represent the fault tree on a computer to provide quantitative results for the overall system performance, subsystem performance, and the basic input event performance.

(3) List the failure that leads to the undesired event and identify minimal cut sets contributing to the failure.

(4) Compute and rank basic input failure and availability performance results.

In performing these steps, the computer program simulates the fault tree and, using the input data, randomly selects the various parameter data from assigned statistical distributions; and then tests whether or not the TOP event occurred within the specified time period. Each test is a trial, and a sufficient number of trials is run to obtain the desired quantitative resolution. Each time the TOP event occurs, the contributing effects of input events and the logical gates causing the specified TOP event are stored and listed as computer output. The output provides a detailed perspective of the system under simulated operating conditions and provides a quantitative basis to support objective decisions.

A number of computer programs have been developed for fault tree analysis. References [83] - [85] provide additional information on fault tree analysis.

In practice, the methods used for fault tree analysis will depend on which ones are available for the computer being used. It will rarely, if ever, be worthwhile generating a computer program especially for a particular problem.

7.10 Sneak Circuit Analysis (SCA)

7.10.1 Definition of Sneak Circuit

A sneak circuit is an unexpected path or logic flow within a system which, under certain conditions, can initiate an undesired function or inhibit a desired function. The path may consist of hardware, software, operator actions, or combinations of these elements. Sneak circuits are not the result of hardware failure but are latent conditions, inadvertently designed into the system or coded into the software program, which can cause it to malfunction under certain conditions. Categories of sneak circuits are:

(1) Sneak paths which cause current, energy, or logical sequence to flow along an unexpected path or in an unintended direction.

(2) Sneak timing in which events occur in an unexpected or conflicting sequence.

(3) Sneak indications which cause an ambiguous or false display of system operating conditions, and thus may result in an undesired action taken by an operator.

(4) Sneak labels which incorrectly or imprecisely label system functions, e.g., system inputs, controls, displays, buses, etc., and thus may mislead an operator into applying an incorrect stimulus to the system.

Figure 7.10-1 depicts a simple sneak circuit example. With the ignition off, the radio turned to the on position, the brake pedal depressed, and the hazard switch engaged, the radio will power on with the flash of the brake lights.

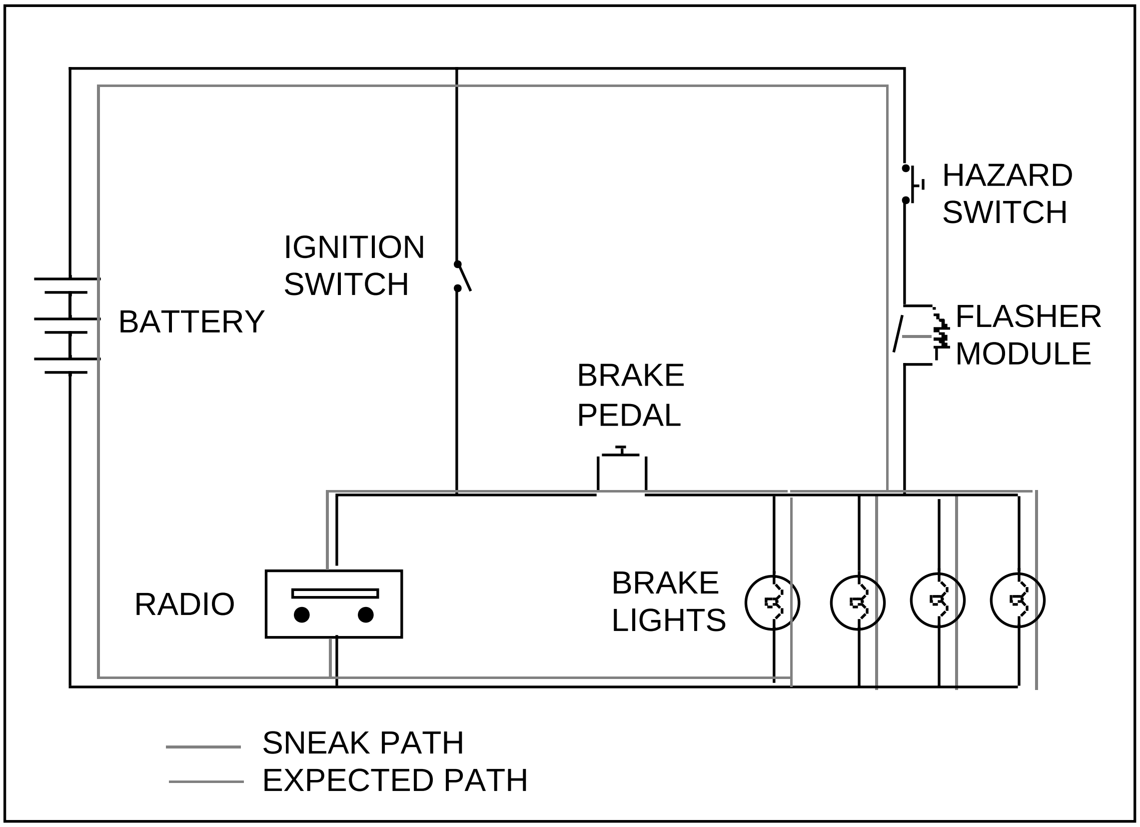


FIGURE 7.10-1: AUTOMOTIVE SNEAK CIRCUIT

7.10.2 SCA: Definition and Traditional Techniques

Sneak circuit analysis is the term that has been applied to a group of analytical techniques which are intended to methodically identify sneak circuits in systems. SCA techniques may be either manual or computer assisted, depending on system complexity. Current SCA techniques which have proven useful in identifying sneak circuits in systems include:

(1) Sneak Path Analysis: A methodical investigation of all possible electrical paths in a hardware system. Sneak path analysis is a technique used for detecting sneak circuits in hardware systems, primarily power distribution, control, switching networks, and analog circuits. The technique is based on known topological similarities of sneak circuits in these types of hardware systems.

(2) Digital Sneak Circuit Analysis: An analysis of digital hardware networks for sneak conditions, operating modes, timing races, logical errors, and inconsistencies. Depending on system complexity, digital SCA may involve the use of sneak path analysis techniques, manual or graphical analysis, computerized logic simulators or computer aided design (CAD) circuit analysis.

(3) Software Sneak Path Analysis: An adaptation of sneak path analysis to computer program logical flows. The technique is used to analyze software logical flows by comparing their topologies to those with known sneak path conditions in them.

7.10.3 New SCA Techniques

SCA is a powerful analytical tool. Historically, however, SCA has been expensive and performed late in the design cycle after all of the design documentation was virtually complete. Thus, any subsequent design changes resulting from the SCA were difficult to make and costly to implement. Therefore, the use of SCA was usually limited to only items and functions which were critical to safety or mission success or where other techniques were not proven to be effective.

This situation, however, has begun to change. Some Air Force publications shed considerable new light on SCA techniques. These publications are:

(1) *Sneak Circuit Analysis for the Common Man,* RADC-TR-89-223, Reference [86]

(2) *Integration of Sneak Circuit Analysis with Design,* RADC-TR-90-109, Reference [87]

(3) *Automated Sneak Circuit Analysis Technique (SCAT)*, Reference [88]

(4) *SCAT: Sneak Circuit Analysis Tool, Version 3.0, RL-TR-95-232*, Reference [89]

SCAT is an interactive "Expert System" design tool to assist the designer in identifying and eliminating both sneak circuits and design concerns early in the design phase. In contrast to normal sneak circuit analyses, SCAT analyses are performed at the assembly level, rather than at the system level. Thus SCAT is not considered to be a replacement for a complete Sneak Circuit Analysis. However, since SCAT is used much earlier in the design phase, it may result in the elimination of many (but not all) potential sneak circuits and decrease the later need for a complete sneak circuit analysis.

Specifically, the referenced publications identify some Sneak Circuit Design Rules, Functional Guidelines, and Device Guidelines that can be applied much earlier in the design phase. This new approach helps significantly to demystify the SCA techniques and enables the Sneak Circuit Analysis to become a much more cost effective reliability design tool.

Because the technology of hardware and software is rapidly evolving, new SCA techniques will undoubtedly evolve as well. SCA will also find applications in non-electrical/electronic systems where analogous situations of energy flow, logic timing, etc. are encountered.

7.10.4 Examples of Categories of SNEAK Circuits

The broad categories of sneak circuits were described in section 7.10.1. Following are some specific examples of each of the categories.

Sneak Path. A sneak path is one which allows current or energy to flow along an unsuspected path or in an unintended direction. There are two distinct subsets of this category. They are:

Sneak Path, Enable occurs when the sneak path initiates an undesired function or result under certain conditions, but not all conditions. An example of this class is shown in Figure 7.10-2.

The electrical power regulator output circuits shown in Figure 7.10-2 represent a portion of the power distribution system in an air vehicle instrument. The sneak path is identified by the arrows along the connection between terminal E6 and pin A of connector J16. This sneak path connects the +4VDC output of regulator VR1 to the +12VDC output of regulator VR2. This path would permit excessive current to flow from the +12VDC output into the +4VDC loads. The result could be failure of either or both regulators (VR1, VR2) and possible catastrophic burnout of the +4VDC loads. Any of these failures would result in the loss of the instrument. If immediate failure did not occur, out-of-tolerance operation of the +4VDC loads would occur due to the 3-times normal voltage being applied. The recommended correction was to remove the wire connection between terminal E6 and pin A of connector J16.

Sneak Path, Inhibit occurs when the sneak path prevents a desired function or results under certain conditions, but not all conditions. An example of this is shown in Figure 7.10-3.

The circuit shown in Figure 7.10-3 was used in a satellite to provide isolation of the power circuits in a double redundant subsystem. The technique removes both power and power ground from the nonoperating backup circuit. The sneak paths which bypass the Q3 grounding switches are identified in Figure 7.10-3 by the arrows placed along each path. When the hardware was wired as shown, total isolation no longer existed and the design intent was violated. The recommended correction was to remove the wire in cable W62 connecting pin 27 of connector P12 to terminal E5 of the single point ground (SPG). When wired as recommended, the power ground switching can be performed by either channel's Q3 and the SPG at E4.

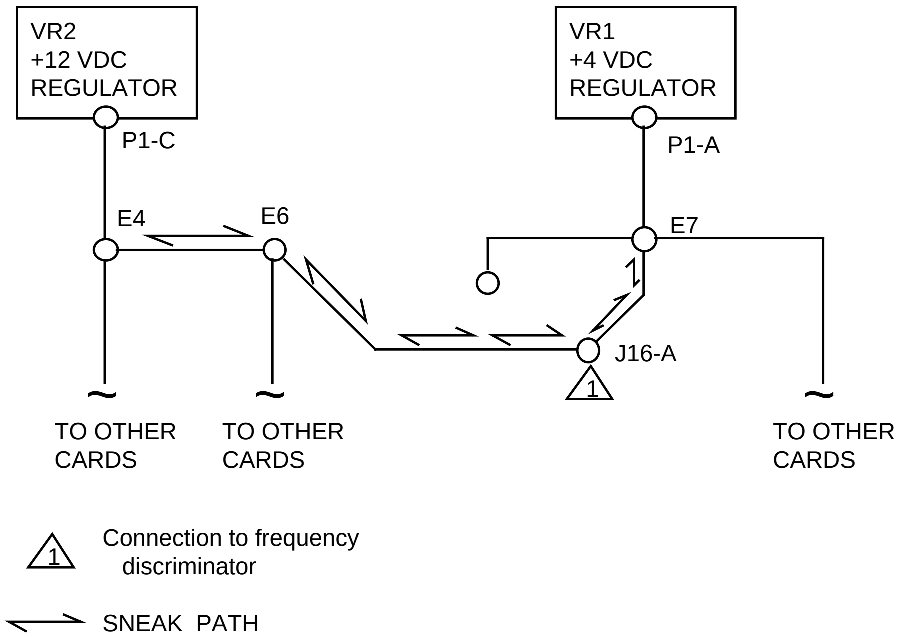


FIGURE 7.10-2: SNEAK PATH ENABLE

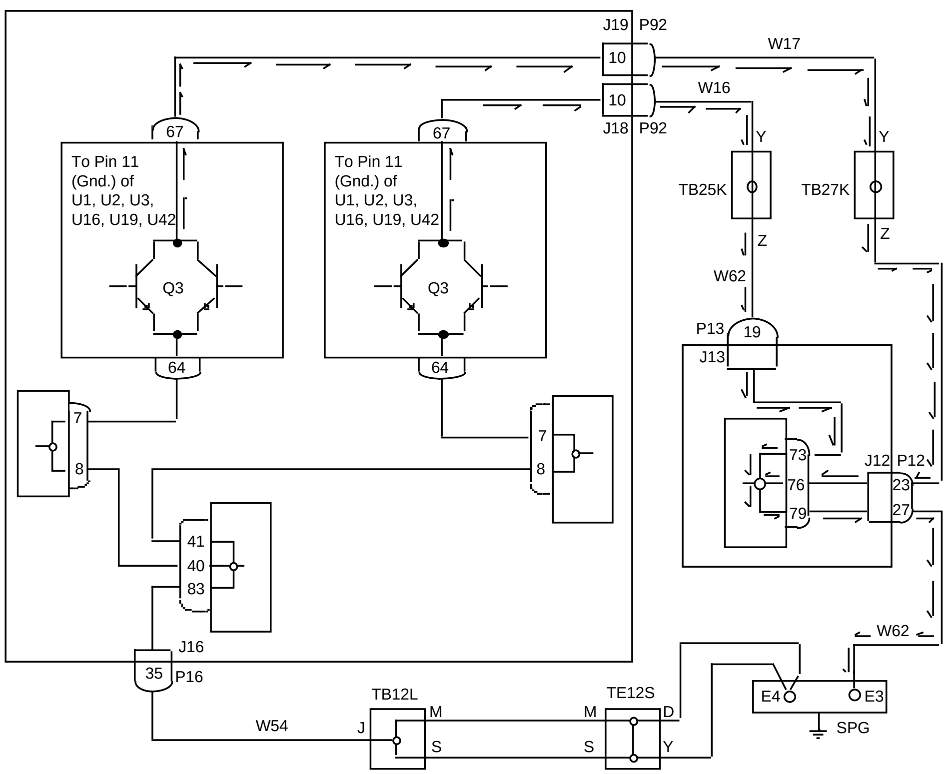
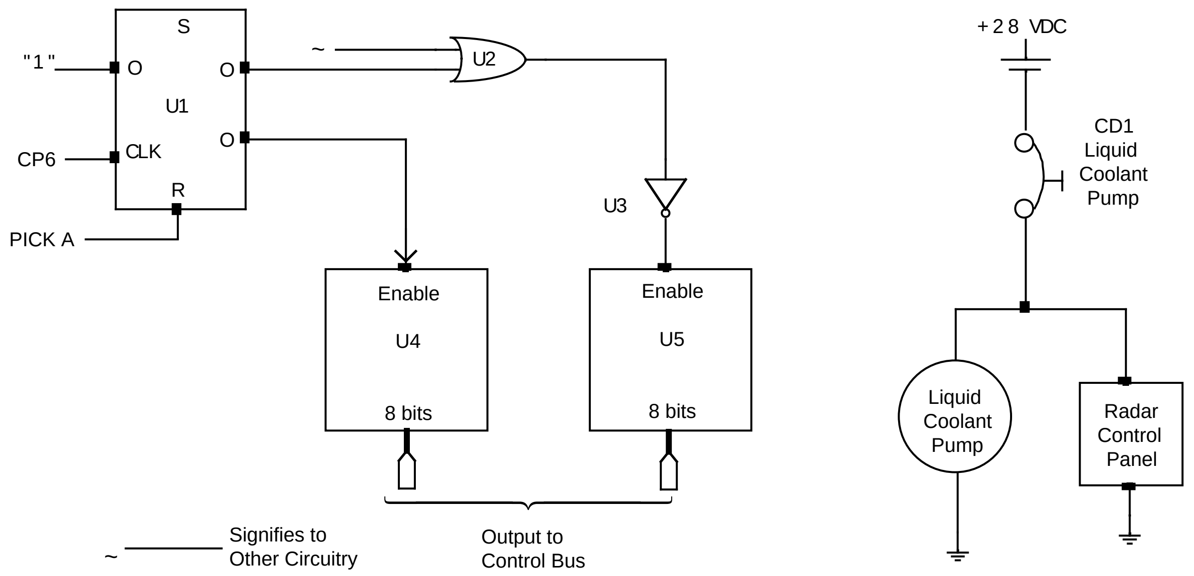


FIGURE 7.10-3: REDUNDANT CIRCUIT SWITCHED GROUND

Sneak Timing: A sneak timing condition is one which causes functions to be inhibited or to occur at an unexpected or undesired time. The example in Figure 7.10-4a illustrates a sneak that occurred in the digital control circuitry of a mine. The enable logic for U4 and U5 allows them, briefly, to be enabled simultaneously. Being CMOS devices in a "wired or" configuration, this allows a potential power-to-ground short through the two devices, damaging or destroying them during operation.

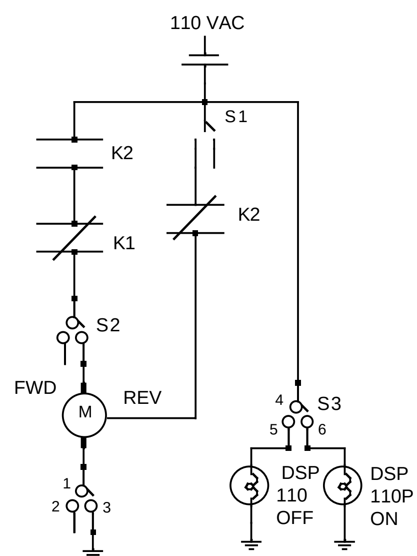
Sneak Indication: An indication which causes ambiguous or incorrect displays. Figure 7.10-4c illustrates a sneak indication which occurred in a sonar power supply system. The MOP (Motor Operated Potentiometer) OFF and ON indicators do not, in fact, monitor the status of the MOP motor. Switch S3 could be in the position shown, providing an MOP ON indication even through switches S1 or S2 or relay contacts K1 or K2 could be open, inhibiting the motor.

Sneak Label: A label on a switch or control device which would cause incorrect actions to be taken by operators. The example in Figure 7.10-4b taken from an aircraft radar system, involves a circuit breaker which provides power to two disparate systems, only one of which is reflected in its label. An operator attempting to remove power from the liquid coolant pump would inadvertently deactivate the entire radar.



(a) SNEAK TIMING

(b) SNEAK LABEL



(c) SNEAK INDICATOR

FIGURE 7.10-4: EXAMPLES OF CATEGORIES OF SNEAK CIRCUITS

7.10.5 SCA Methodology

7.10.5.1 Network Tree Production

The first major consideration that must be satisfied to identify sneak circuit conditions is to ensure that the data being used for the analysis represent the actual "as built" circuitry of the system. Functional, integrated, and system level schematics do not always represent the actual constructed hardware. Detailed manufacturing and installation schematics must be used, because these drawings specify exactly what was built, contingent on quality control checks, tests, and inspection. However, manufacturing and installation schematics rarely show complete circuits. The schematics are laid out to facilitate hookup by technicians without regard to circuit or segment function. As a result, analysis from detail schematics is extremely difficult. So many details and unapparent continuities exist in these drawings that an analyst becomes entangled and lost in the maze. Yet, these schematics are the data that must be used if analytical results are to be based on true electrical continuity. The first task of the sneak analyst is, therefore, to convert this detailed, accurate information into a form usable for analytical work. The magnitude of data manipulation required for this conversion necessitates the use of computer automation in most cases.

Automation has been used in sneak circuit analysis since 1970 as the basic method for tree production from manufacturing detail data. Computer programs have been developed to allow encoding of simple continuities in discrete "from-to" segments extracted from detail schematics and wire lists. The encoding can be accomplished without knowledge of circuit function. The computer connects associated points into paths and collects the paths into node sets. The node sets represent interconnected nodes that make up each circuit. Plotter output of node sets and other reports are generated by the computer to enable the analyst to easily sketch accurate topological trees. The computer reports also provide complete indexing of every component and data point to its associated tree. This feature is especially useful in cross indexing functionally related or interdependent trees, in incorporating changes, and in troubleshooting during operational support.

7.10.5.2 Topological Pattern Identification

Once the network trees have been produced, the next task of the analyst is to identify the basic topological patterns that appear in each tree. Five basic patterns exist for hardware SCA: (1) single line (no-node) topograph, (2) ground dome, (3) power dome, (4) combination dome, and (5) "H" pattern. These patterns are illustrated in Figure 7.10-5. One of these patterns or several in combination will characterize the circuitry shown in any given network tree. Although, at first glance, a given circuit may appear more complex than these basic patterns, closer inspection reveals that the circuit is actually composed of these basic patterns in combination. In examining each node in the network tree, the sneak circuit analyst must identify the topographical pattern or patterns incorporating the node and apply the basic clues that have been found to typify sneak circuits involving that particular pattern.

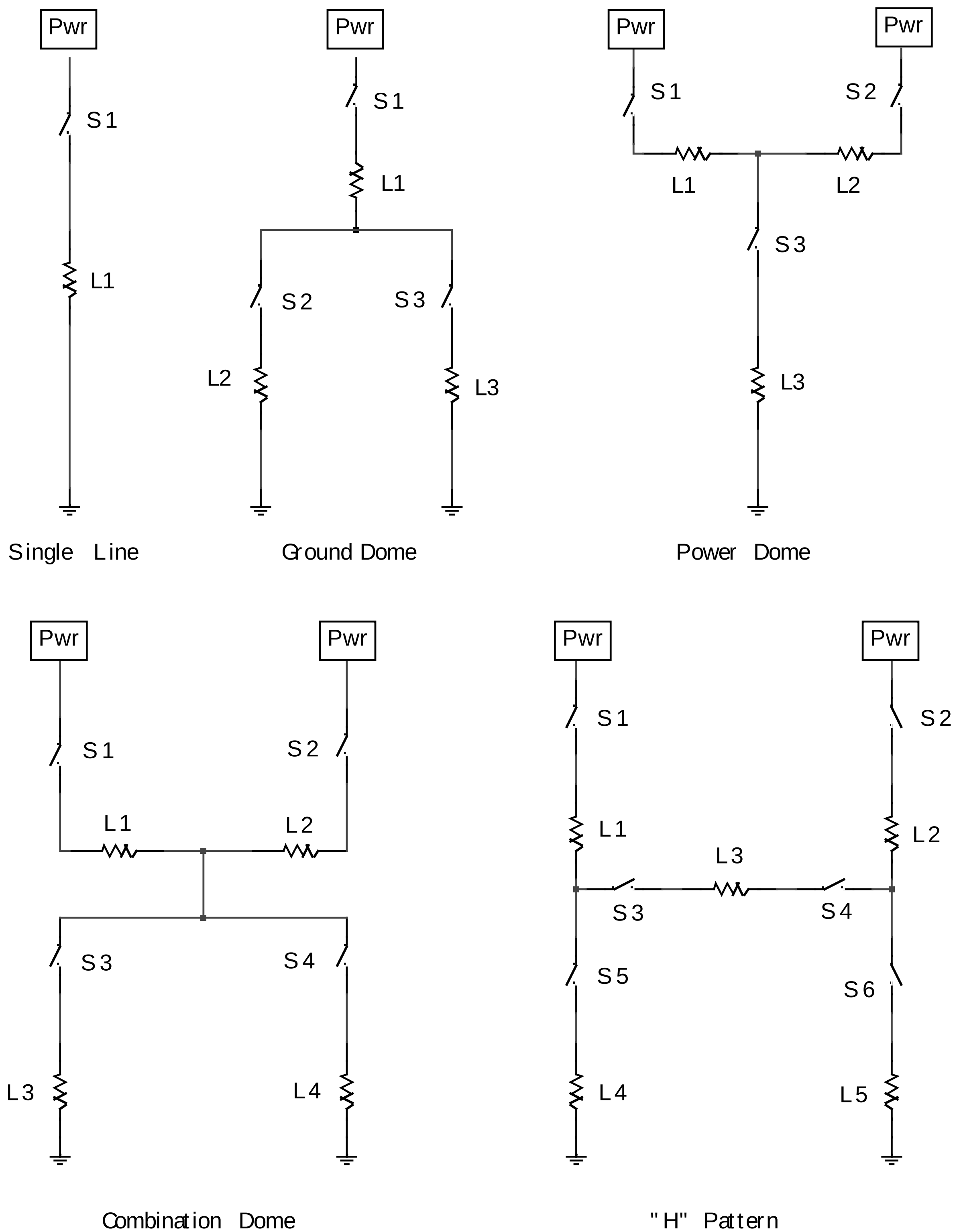


FIGURE 7.10-5: BASIC TOPOGRAPHS

7.10.5.3 Clue Application

Associated with each pattern is a list of clues to help the analyst identify sneak circuit conditions. These lists were first generated during the original study of historical sneak circuits. The lists were updated and revised during the first several years of applied sneak circuit analysis. Now, the list of clues provides a guide to all possible design flaws that can occur in a circuit containing one or more of the five basic topological configurations, subject to the addition of new clues associated with new technological developments. The lists consist of a series of questions that the analyst must answer about the circuit to ensure that it is sneak free.

As an example, the single line topograph (Figure 7.10-5) would have clues such as:

(a) Is switch S open when load L is desired?

(b) Is switch S closed when load L is not desired?

Obviously, sneak circuits are rarely encountered in this topograph because of its simplicity. Of course, this is an elementary example and is given primarily as the default case which covers circuitry not included by the other topographs.

With each successive topograph, the clue list becomes longer and more complicated. The clue list for the "H" pattern includes over 100 clues. This pattern, because of its complexity, is associated with more sneak circuits than any of the previous patterns. Almost half of the critical sneak circuits identified to date can be attributed to the "H" patterns. Such a design configuration should be avoided whenever possible. The possibility of current reversal through the "H" crossbar is the most commonly used clue associated with "H" pattern sneak circuits.

7.10.6 Software Sneak Analysis

In 1975, a feasibility study was performed resulting in the development of a formal technique, involving the use of mathematical graph theory, electrical sneak theory, and computerized search algorithms, to identify sneaks in software programs. A software sneak is defined as a logic control path which causes an unwanted operation to occur or which bypasses a desired operation, without regard to failures of the hardware system to respond as programmed.

The feasibility study concluded that:

(1) Software Sneak Analysis is a viable means of identifying certain classes of software problems.

(2) Software Sneak Analysis works equally well on different software languages.

(3) Software Sneak Analysis does not require execution of the software to detect problems.

The Software Sneak Analysis technique has evolved along lines very similar to hardware Sneak Circuit Analysis. Topological network trees are used with electrical symbology representing the software commands to allow easy cross analysis between hardware and software trees and to allow the use of a single standardized analysis procedure.

Since topological pattern recognition is the keystone of both Sneak Circuit Analysis and Software Sneak Analysis, the overall methodologies are quite similar. The software package to be analyzed must be encoded, processed, and reduced to a standardized topographical format, the basic topological patterns identified and the appropriate problem clues applied to each pattern. For software, it has been found that six basic patterns exist: the Single Line, the Return Dome, the Iteration/Loop Circuit, the Parallel Line, the Entry Dome, and the Trap Circuit, as shown in Figure 7.10-6.

Although at first glance, a given software tree may appear to be more complex than these basic patterns, closer inspection will reveal that the code is actually composed of these basic structures in combination. As each node in the tree is examined, the analyst must identify which pattern or patterns include that node. The analyst then applies the basic clues that have been found to typify the sneaks involved with that particular structure. These clues are in the form of questions that the analyst must answer about the use and interrelationships of the instructions that are elements of the structure. These questions are designed to aid in the identification of the sneak conditions in the instruction set which could produce undesired program outputs.

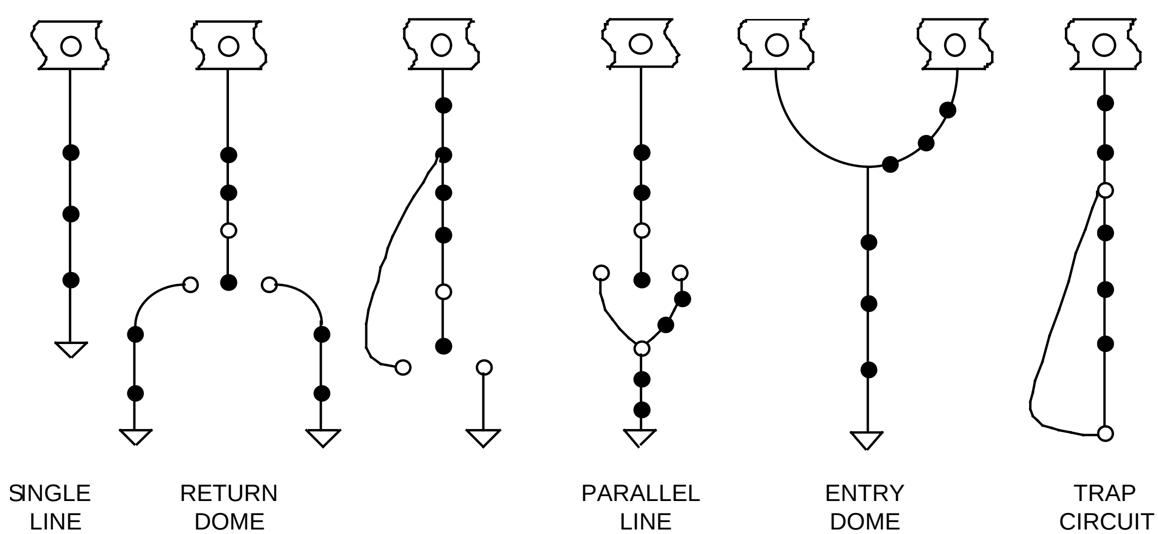


FIGURE 7.10-6: SOFTWARE TOPOGRAPHS

Software sneaks are classified into four basic types:

(1) Sneak Output: The occurrence of an undesired output.

(2) Sneak Inhibit: The undesired inhibition of an output.

(3) Sneak Timing: The occurrence of an undesired output by virtue of its timing or mismatched input timing

(4) Sneak Message: The program message does not adequately reflect the condition.

Figure 7.10-7 illustrates a software sneak which occurred in the operating software of a military aircraft. Figure 7.10-7a illustrates the design intent of the section of software with the sneak. When the actual code was produced, however, the two tests were inadvertently interchanged. The network tree of the actual software code (see Figure 7.10-7b) makes the sneak readily apparent. This historical problem was uncovered only during the software system integrated testing when it was found that the instructions represented by LOAD 1 could never be executed.

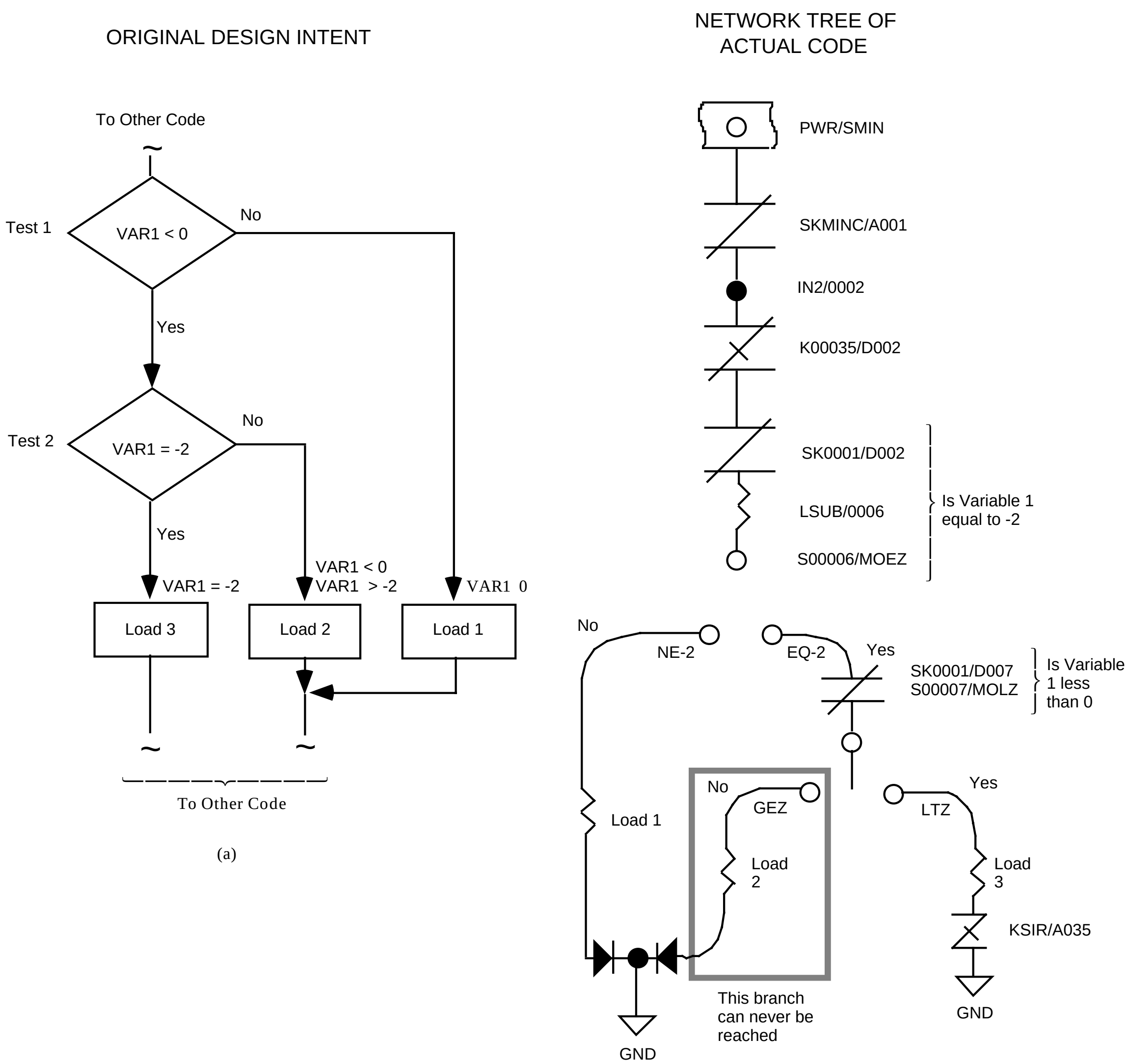


FIGURE 7.10-7: SOFTWARE SNEAK EXAMPLE

7.10.7 Integration of Hardware/Software Analysis

After a sneak circuit analysis and a software sneak analysis have been performed on a system, the interactions of the hardware and software can readily be determined. For this purpose, the analyst has diagramatic representations of these two elements of the system in a single standardized format. The effect of a control operation that is initiated by some hardware element can be traced through the hardware trees until it impacts the system software. The logic flow can then be traced through the software trees to determine its ultimate impact on the system. Similarly, the logic sequence of a software initiated action can be followed through the software and electrical network trees until its eventual total system impact can be assessed.

The joint analysis of a system's software and hardware circuitry previously described is simply called Sneak Analysis. Sneak Analysis helps provide visibility of the interactions of a system's hardware and software and hence will help reduce the difficulties involved in the proper integration of two such diverse, complex system designs. As hardware and software systems increase in complexity, the use of interface bridging analysis tools, such as Sneak Analysis, becomes imperative to help ensure the safety of the total system.

7.10.8 Summary

SCA is different from other analyses commonly performed in a reliability program in a number of important ways. SCA generally concentrates on the interconnections, interrelationships, and interactions of system components rather than on the components themselves. SCA concentrates more on what might go wrong in a system rather than on verifying that it works right under some set of test conditions. The SCA technique is based on a comparison with other systems which have "gone wrong", not because of part failures, but because of design oversight or because a human operator made a mistake. The consequence of this subtly different perspective may be very important, because it tends to concentrate on and find problems which may be hidden from the perspectives of other analytical techniques.

For example FMEA/FMECA differs from SCA in that it predicts and quantifies the response of a system to failures of individual parts or subsystems. An FMECA is an analysis of all expected failure modes and their effect on system performance. FMECA results are often used in maintainability predictions, in the preparation of maintenance dependency charts, and to establish sparing requirements. SCA, on the other hand, considers possible human error in providing system inputs while FMECA does not. In this regard the two types of analysis tend to complement one another.

Fault Tree Analysis is a deductive method in which a catastrophic, hazardous end result is postulated and the possible events, faults, and occurrences which might lead to that end event are determined. Thus, FTA overlaps SCA in purpose because the FTA is concerned with all possible faults, including component failures as well as operator errors.

Concerning the availability of SCA computer programs, the original SCA computer programs developed under government contract with (NASA), Johnson Spacecraft Center, Houston, Texas, on the Apollo program are available to all industry and government agencies. They can be purchased from Computer Software Management and Information Center (COSMIC), University of Georgia, 112 Barrow Hall, Athens, Georgia 30602. These programs may not be current. However, several companies have purchased these programs and updated them. The improved programs and the accompanying analysis techniques are considered proprietary by most companies.

References [86] - [93] provide more details on SCA.

7.11 Design Reviews

7.11.1 Introduction and General Information

The purpose of a design review is to ensure that each design has been studied to identify possible problems, to improve the item where necessary, and to provide assurance that the most satisfactory design has been selected to meet the specified requirements. Design reviews are critical audits of all pertinent aspects of the design and are conducted at critical milestones in an acquisition program. They are essential to reliability engineering.

The formal review (depicted in Figure 7.11-1) of equipment design concepts and design documentation for both hardware and software is an essential activity in any development program. Standard procedures should be established to conduct a review of all drawings, specifications, and other design information by a supplier's technical groups such as engineering, reliability engineering, and manufacturing engineering. (Ideally, representatives of these and other key groups would comprise one or more integrated product development teams (IPDTs)). This review should be accomplished prior to the release of design information for manufacturing operations. Such a review is an integral part of the design-checking reviews. Responsible members of each reviewing department meet to consider all design documents, resolve any problem areas uncovered, and signify their acceptance of the design documentation by approving the documents for their departments.

Reliability engineering, ideally as part of an IPDT, should conduct an intensive review of the system during initial design. A design review, from a reliability perspective, includes the following major tasks:

(1) Analysis of environment and specifications

(2) Formal design review of engineering information

(3) Reliability participation in all checking reviews

Prior to the formal review, the requirements defined in applicable specifications are reviewed. The expected environmental extremes of the system are studied to determine suspected detrimental effects on equipment performance. Checklists, based on these studies, are prepared to ensure that the objectives of formal design reviews are fulfilled.

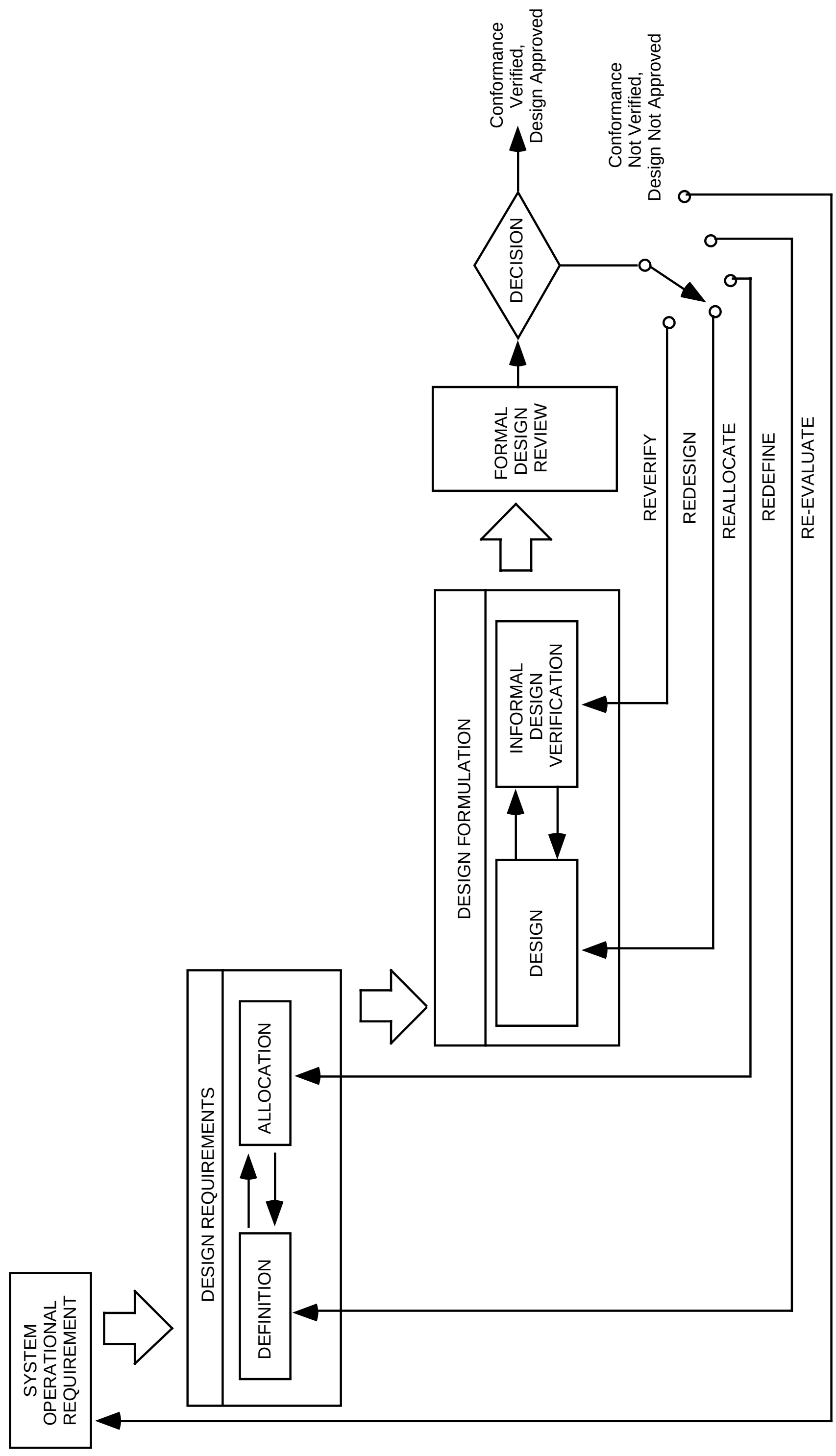


FIGURE 7.11-1: DESIGN REVIEW AS A CHECK VALVE IN THE SYSTEM ENGINEERING CYCLE

The formal design review, which is instituted prior to the release of drawings, is intended to do the following in regard to reliability:

(1) Detect any conditions that could degrade equipment reliability

(2) Provide assurance of equipment conformance to applicable specifications

(3) Ensure the use of preferred or standard parts as far as practical

(4) Ensure the use of preferred circuitry as far as possible

(5) Evaluate the electrical, mechanical, and thermal aspects of the design

(6) Review stress analysis to ensure adequate part derating

(7) Ensure accessibility of all parts that are subject to adjustment

(8) Ensure interchangeability of similar subsystems, circuits, modules, and subassemblies

(9) Ensure that adequate attention is given to all human factors aspects of the design

(10) Ensure that the quality control effort will be effective

Reviews should be made at appropriate stages of the design process. It may be necessary to conduct specific reviews to evaluate achievement of the reliability requirements on a timely basis. The reviews should include, to the extent applicable but not necessarily limited to: current reliability estimates and achievements for each mode of operation, as derived from reliability analyses or test(s); potential design or production (derived from reliability analyses) problem areas, and control measures necessary to preserve the inherent reliability; failure mode(s) and effect(s) and criticality analyses; corrective action on reliability critical items; effects of engineering decisions, changes and tradeoffs upon reliability achievements, potential and growth, within the functional model framework; status of supplier and vendor reliability programs; and status of previously-approved design review actions. The results of reliability reviews should be documented.

In order to satisfy the objectives of the design review, the review team must have sufficient breadth to handle aspects of the items under review, such as performance, reliability, etc., and the interfaces and interactions with adjacent items. The ultimate objective of the team is to arrive at a balanced and reliable design.

7.11.2 Informal Reliability Design Review

The informal reliability design review is conducted for the purpose of evaluating and guiding specified reliability characteristics and maintenance features "in process." That is, it is conducted while the design is in the evolutionary or formative stage and still amenable to major conceptual and configuration changes. Reviews are conducted on an unscheduled, "as required," informal basis. They are usually conducted at the request of the designer or the systems engineer to verify conformance throughout the team effort, to allocate requirements and design constraints, to verify the solution of problems identified in earlier design iterations, or to provide the basis for selection of design alternatives.

Even though the verification review is an informal working session, usually involving only a few selected reviewers, results of each review should be documented. The review may result in one of five alternatives being selected for further design iteration. These alternatives are:

(1) **Reverify Design Adequacy** to provide additional analytical or empirical proof of design adequacy to facilitate design review approval decision with more confidence than current data will substantiate

(2) **Redesign** to correct design discrepancies and marginal characteristics disclosed by the review

(3) **Reallocate Design Requirements** to rectify allocation errors identified in the review, or reallocate subsystem requirements on the basis of updated estimates of design feasibility or changes in relative criticality disclosed during the review

(4) **Redefine Design Requirements** to restudy previous requirements analyses and tradeoff studies, and redefine or refine baseline design and configuration requirements more nearly consistent with state-of-art and program constraints revealed during the design review.

(5) **Re-evaluate System Operational Requirements** to provide the basis for choosing one of two alternatives: (a) redefine system operational requirements consistent with current design state-of-art and program constraints; or (b) redefine program constraints, such as delivery schedule and funds, to rectify earlier estimating errors.

The recommended design review team membership, and functions of each member, are briefly summarized in Table 7.11-1. For these informal design reviews, customer participation is usually optional. The IPDT is the current and preferred approach to forming the design team.

7.11.3 Formal Design Reviews

Formal design reviews, when the customer is the government, are usually the subject of contractual agreement between the government and the supplier. Table 7.11-1 shows the recommended review team composition. Regardless of who the customer is, formal reviews normally include the following:

Preliminary Design Review (PDR): The PDR is conducted prior to the detail design process to evaluate the progress and technical adequacy of the selected design approach, determine its compatibility with the performance requirements of the specification; and establish the existence and the physical and functional interfaces between the item and other items of equipment or facilities. The basic design reliability tasks shown in Figure 7.11-3 should be accomplished for the PDR.

Eight suggested basic steps pertinent to the PDR are shown in Figure 7.11-2.

Critical Design Review: The CDR is conducted when detail design is essentially complete and fabrication drawings are ready for release. It is conducted to determine that the detail design satisfies the design requirements established in the specification, and establish the exact interface relationships between the item and other items of equipment and facilities.

Preproduction Reliability Design Review (PRDR): The PRDR is a formal technical review conducted to determine if the achieved reliability of a weapon system at a particular point in time is acceptable to justify commencement of production. For DoD acquisitions, details for the PRDR are usually provided in the individual Service documents or instructions, e.g., NAVAIR INST. 13070.5.

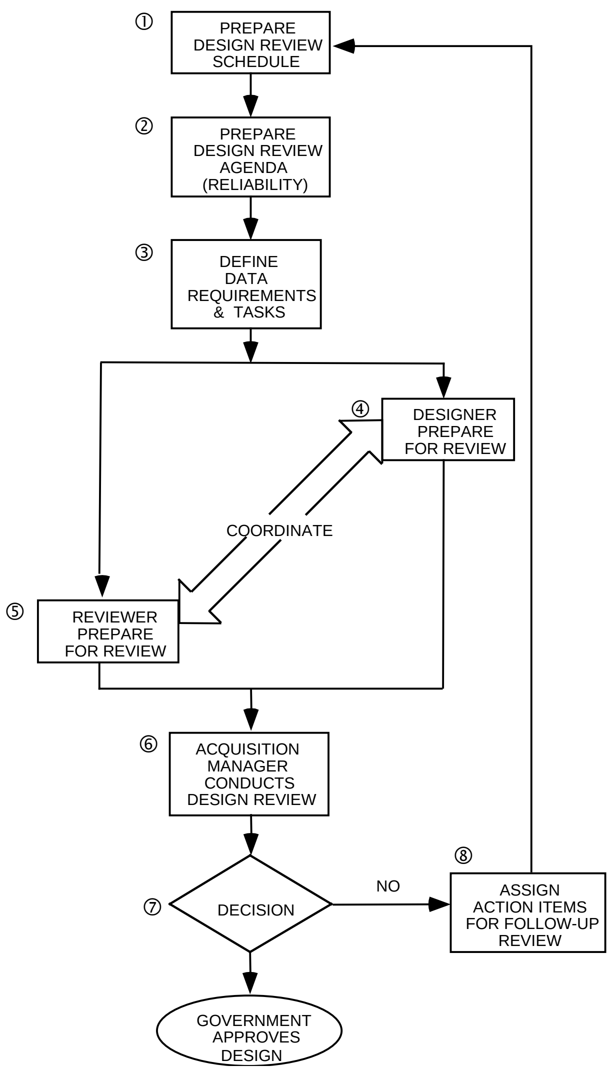
TABLE 7.11-1: DESIGN REVIEW GROUP, RESPONSIBILITIES AND MEMBERSHIP SCHEDULE

Group Member	Responsibilities
Chairman	Calls, conducts meetings of group, and issues interim and final reports
Design Engineer(s) (of product)	Prepares and presents design and substantiates decisions with data from tests or calculations
*Reliability Manager or Engineer	Evaluates design for optimum reliability, consistent with goals
Quality Control Manager or Engineer	Ensures that the functions of inspection, control, and test can be efficiently carried out
Manufacturing Engineer	Ensures that the design is producible at minimum cost and schedule
Field Engineer	Ensures that installation, maintenance, and operator considerations were included in the design
Procurement Representative	Assures that acceptable parts and materials are available to meet cost and delivery schedules
Materials Engineer	Ensures that materials selected will perform as required
Tooling Engineer	Evaluates design in terms of the tooling costs required to satisfy tolerance and functional requirements
Packaging and Shipping Engineer	Assures that the product is capable of being handled without damage, etc.
Design Engineers (not associated with unit under review)	Constructively review adequacy of design to meet all requirements of customer
Customer Representative (optional)	Generally voices opinion to acceptability of design and may request further investigation on specific items

*May have other titles within some companies. Other specialties, such as maintainability, human factors, and value engineering are also represented.

The PRDR is conducted just prior to production (and, for DoD programs, after completion of initial operational test and evaluation) to ensure the adequacy of the design from a reliability standpoint. The level of achieved reliability and adequacy of design will be evaluated primarily on initial technical and operational testing, e.g., test results, failure reports, failure analyses reports, reports of corrective action, and other documents which could be used as necessary for back-up or to provide a test history.

Suggested steps for a CDR are shown in Figure 7.11-4. The basic design reliability tasks shown in Figure 7.11-5 should be accomplished for the CDR.



① Design review requirements are established. Supplier prepares preliminary design review schedule compatible with master program milestone schedule.

② Supplier prepares and distributes agenda in advance of each review, defining: purpose and scope of review; specific items to be reviewed; date, time, and place for the review.

③ Supplier defines applicable data, to assist designer and reviewer in preparation for the design review.

④ Designer prepares for review, to include answers to questions accompanying the agenda, description of design concept, analysis, test results, problems, requirements, etc.

⑤ Review committee (or IPDT) members prepare for review; formulate questions and suggestions.

⑥⑦⑧ Review committee members approve or disapprove preliminary design.

FIGURE 7.11-2: BASIC STEPS IN THE PRELIMINARY DESIGN REVIEW (PDR) CYCLE

1. Identify the quantitative reliability requirements and compare preliminary predictions with specified requirements.

2. Review failure rate sources, derating policies, and prediction methods.

3. Identify planned actions when predictions are less than specified requirements.

4. Identify and review parts or items which have a critical life or require special consideration, and general plan for handling.

5. Identify applications of redundant elements. Evaluate the basis for their use and provisions for redundancy with switching.

6. Review critical signal paths to determine that a fail-safe/fail-soft design has been provided.

7. Review margins of safety between functional requirements and design provisions for elements, such as: power supplies, transmitter modules, motors, and hydraulic pumps. Similarly, review structural elements, i.e., antenna pedestals, dishes, and radomes to determine that adequate margins of safety are provided between operational stresses and design strengths.

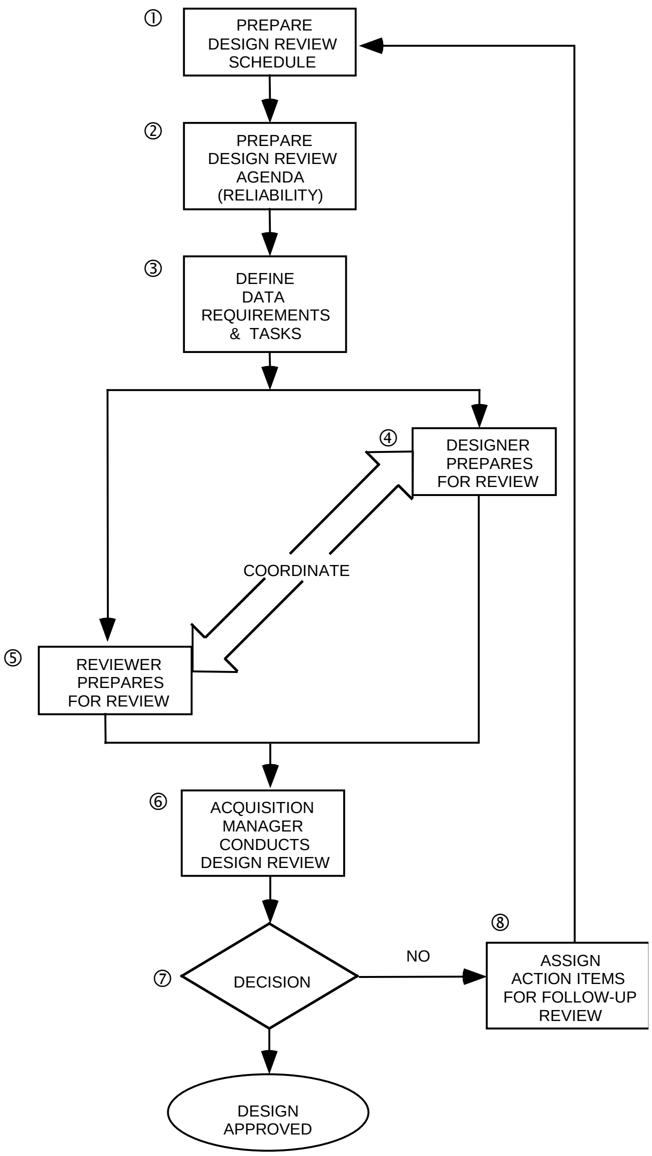
8. Review Reliability Design Guidelines to ensure that design reliability concepts shall be available and used by equipment designers. Reliability Design Guidelines should include, part application guidelines (electrical derating, thermal derating, part parameter tolerances), part selection order of preference, prohibited parts/materials, reliability allocations/predictions, and management procedures to ensure compliance with the guidelines.

9. Review preliminary reliability demonstration plan: failure counting ground rules, accept-reject criteria, number of test articles, test location and environment, planned starting date, and test duration.

10. Review elements of reliability program plan to determine that each task has been initiated toward achieving specified requirements.

11. Review vendor reliability controls.

FIGURE 7.11-3: DESIGN RELIABILITY TASKS FOR THE PDR



①② Design review requirements are established. Supplier prepares and distributes agenda in advance of Critical Design Review (CDR) defining: purpose and scope of review; specific items to be reviewed; date, time and place for the review.

③ Supplier defines applicable data, to assist designer and reviewer in preparation for the design review.

④ Designer prepares for pre-critical design review, to include answers to questions accompanying the agenda, description of design concept, analyses, test results, problems, requirements, etc.

⑤ Review committee (or IPDT) members prepare for review; formulate questions and suggestions.

⑥ Acquisition Manager conducts the critical design review meeting.

⑦ Decisions made either to approve the design or to withhold approval pending correction of deficiencies.

⑧ Action items for correction of deficiencies assigned and schedule for follow-up review established.

FIGURE 7.11-4: BASIC STEPS IN THE CDR CYCLE

1. Review the most recent predictions or assessments of quantitative reliability and compare against specified requirements. Substantiate predictions by review of parts application stress data and substantiate assessments by reviewing any test data.

2. Review application of parts or items with minimum life, or those which require special consideration to insure their affect on system performance is minimized.

3. Review completed Reliability Design Review Checklist to insure principles have been satisfactorily reflected in the design.

4. Review applications of redundant elements to establish that expectations have materialized since the PDR.

5. Review detailed reliability demonstration plan for compatibility with specified test requirements. Review the number of test articles, schedules, location, test conditions, and personnel involved to insure a mutual understanding of the plan and to provide overall planning information to activities concerned.

FIGURE 7.11-5: DESIGN RELIABILITY TASKS FOR THE CRITICAL DESIGN REVIEW (CDR)

7.11.4 Design Review Checklists

A design review checklist delineates specific areas to be considered for the item under review. In order to ensure that every consideration has been appropriately taken into account, a checklist for design should be prepared. Figure 7.11-6 is a typical list of areas to be considered in various stages of a design review (not to be considered all inclusive). Table 7.11-2 is a typical example of a Reliability Actions Checklist.

Technical checklists can be in question format to ensure that critical factors will not be overlooked. Figure 7.11-7 illustrates typical questions which could be asked at various stages of the design review.

1. System concept/alternative approaches
2. System performance and stability
3. Design documentation
4. Design changes
5. Tradeoff studies
6. Materials and Processes
7. Construction, Fabrication, Maintenance and Service
8. Analyses (Failure Mode, Effects and Criticality, Tolerance, etc.
9. Equipment compatibility
10. Environmental effects
11. Test data
12. Reliability allocation/prediction/assessment
13. Redundancy
14. Cost and procurement considerations
15. Life and controls
16. Interchangeability, spares and repair parts
17. Weight
18. Supplier design
19. Safety
20. Critical functions

FIGURE 7.11-6: TYPICAL AREAS TO BE COVERED IN A DESIGN REVIEW

TABLE 7.11-2: RELIABILITY ACTIONS CHECKLIST

DESIGN TITLE			NUMBER		Notes
			Responsibility		&
No.	Item	Completed	Design	Reliability	Comments
1.	System Constraints		D	X	
	a. Success Criteria		D	X	
	b. Environmental Stresses		D	X	
	c. Compatibility Factors		D	X	
	d. User Skill Levels		D	X	
2.	Feasibility Study		D	X	
3.	Reliability Apportionment			R	
4.	Preliminary Reliability Review		D	R	
5.	Trade-Off Studies		D	X	
6.	Functional Schematics		D	X	
7.	Block Diagram		D	X	
8.	Cause and Effect Analysis		D	X	
9.	Worst Case Analysis		D	X	
10.	Subsystem and Equipment Reliability Prediction				
	a. Part Failure Rate Method		D	X	
	b. Safety Margin Method		D	X	
	c. Drift Rate and Tolerance Method		D	X	
11.	Intermediate Design Review		D	R	
12.	Time/ Cycle Recording Requirements		D	X	
13.	Failure Reporting Requirements		D	X	
14.	Serialization Requirements		D	X	
15.	Procurement Specification Review			R	
16.	Vendor Proposal Review			R	

CODE
D - Prime Action by Designer - check off, sign and date as completed.
R - Prime Action by Reliability Engineer - check off, sign and date as completed.
X - Check by Reliability Engineer - initial and date.

TABLE 7.11-2: RELIABILITY ACTIONS CHECKLIST (CONT'D)

DESIGN TITLE			NUMBER		Notes
			Responsibility		&
No.	Item	Completed	Design	Reliability	Comments
17.	Source Selection Review			R	
18.	Parts Selection and Application Review		D	X	
19.	Reliability Signoff - Top Assy. & Inst. Dwgs.			R	
20.	Vendor Design Review			R	
21.	Critical Design Review		D	R	
22.	Process Controls		D	X	
23.	Manufacturing Procedure Controls			X	
24.	Qualification Test Review		D	X	
25.	Acceptance Test Review		D	X	
26.	Integration Test Review		D	X	
27.	Reliability Demonstration Test Review		D	X	
28.	System Test:				
	a. Test Requirements Review		D	X	
	b. Test Plans Review		D	X	
	c. Reliability Tests			R	
29.	Reliability Summary Sheet			R	

CODE
D - Prime Action by Designer - check off, sign and date as completed.
R - Prime Action by Reliability Engineer - check off, sign and date as completed.
X - Check by Reliability Engineer - initial and date.

1. Is the design simple?
2. Does the design reflect an integrated system approach rather than an accumulation of parts?
3. Is the item compatible with the system in which it is used?
4. Are there adequate indicators to verify critical functions?
5. Has reliability of spares and repair parts been considered?
6. Are reliability requirements established for critical items? For each part?
7. Are there specific reliability design criteria for each item?
8. Have reliability tests been established?
9. Are appropriate parts being used properly?
10. Are unreliable parts identified?
11. Has the failure rate for each part or part class been established?
12. Have parts been selected to meet reliability requirements?
13. Has shelf life been determined?
14. Have limited-life parts been identified, and inspection, and replacement requirements specified?
15. Have critical parts which required special procurement, testing, and handling been identified?
16. Have stress analyses been accomplished?
17. Have derating factors been used in the application of parts?
18. Have safety factors and safety margin been used in the application of parts?
19. Are circuit safety margins ample?
20. Have standard and proven circuits been used?
21. Has the need for the selection of parts (matching) been eliminated?
22. Have circuit studies been made considering variability and degradation of electrical parameters of parts?
23. Is the reliability or MTBF of the item based on actual application of the parts?
    a. Comparison made with reliability goal?
    b. Provision for necessary design adjustments?

FIGURE 7.11-7: TYPICAL QUESTIONS CHECKLIST FOR THE DESIGN REVIEW

24. Are the best available methods for reducing the adverse effects of operational environments on critical parts being utilized?

25. Has provision been made for forecasting failures, including the use of marginal testing?

26. Is there a process for identifying improvements to eliminate design inadequacies observed in tests?

27. Have normal modes of failure and the magnitude of each mode for each item or critical part been identified?

28. Have the following effects been considered?
    a. External effects on the next higher level in which the item is located.
    b. Internal effects on the item.
    c. Common effects, or direct effect of one item on another item, because of mechanical or electro-mechanical or electro-mechanical linkage.

30. Has redundancy been provided where needed to meet specified reliability?

31. Have failure mode and effects analyses been adequately conducted for the design?

32. Have the risks associated with critical item failures been identified? Accepted? Has design action been taken?

33. Does the design account for early failure, useful life and wear-out?

FIGURE 7.11-7: TYPICAL QUESTIONS CHECKLIST FOR THE DESIGN REVIEW

7.12 Design for Testability

Testability, an important subset of maintainability, is a product design characteristic reflecting the ability to determine the status (operable, inoperable or degraded) of an item, and to isolate faults within the item in a timely and efficient manner. Therefore, a great deal of attention must be paid to ensuring that all designs incorporate features that allow testing to occur without a great deal of effort. The design must be such that testing is efficient in terms of detecting and isolating only failed items, with no removal of good items. The removal of good items continues to be a problem in many industries, with obvious impacts on troubleshooting times and repair and logistics costs.

Design guides and analysis tools must be used rigorously to ensure a testable design. Not doing so leads to greater costs in the development of manufacturing and field tests, as well as in the development of test equipment. Trade-offs must be made up front on the use of built-in-test (BIT) versus other means of fault detection and isolation. Further, the expected percentage of faults that can be detected and isolated to a specified or desired level of ambiguity must be determined - it is an important input to the logistics analysis process. The consequences of poor testability are higher manufacturing costs, higher support costs, and lower customer satisfaction.

7.12.1 Definition of Testability and Related Terms

Testability is a discipline that has many unique terms that may be unfamiliar to some readers. Many of these terms are defined in Section 3. Some additional terms are defined here to help in understanding the material that follows. Additional terms and definitions related to testability are provided (References [94] and [95]).

- Controllability: A design attribute that defines or describes the degree of test control that can be realized at internal nodes of interest.
- General Purpose Test Equipment (GPTE): Test equipment used to measure a range of parameters common to two or more systems of basically different design.
- Observability: A design attribute that describes the extent to which signals of interest can be observed.
- On-line Test: Testing of a UUT in its normal operating environment.
- Off-line Test: Testing of a UUT removed from its normal operating environment.
- Troubleshooting: A procedure for locating and diagnosing malfunctions or breakdowns in equipment using systematic checking or analysis.

7.12.2 Distinction between Testability and Diagnostics

Whereas testability is related to the physical design characteristics of a product, diagnostics are related to the means by which faults are detected and isolated. This includes the actual on-line and off-line tests themselves, as well as the means (BIT, BIT Equipment, GPTE, External Test Equipment, etc.) by which tests are performed. Achieving good diagnostics involves determining the diagnostic capability required in a product. A diagnostic capability can be defined as all capabilities associated with detecting, isolating, and reporting faults, including testing, technical information, personnel, and training. In comparing testability with diagnostics, we see that testability is an inherent design characteristic, while diagnostics involves factors other than those associated with the design itself. Attention paid to both in all design phases will impact not only the cost of producing a product, but certainly the cost and time associated with troubleshooting failures of the product once it has been fielded.

7.12.3 Designing for Testability

Although a subset of maintainability, testability has become recognized as a separate design discipline in its own right. Because of the impact of poor testability on production and maintenance costs, it will continue to be treated as a distinct discipline, at least in the foreseeable future. Therefore, it is important to develop a testability program plan as an integral part of the systems engineering process, and to elevate testability to the same level of importance accorded to other product assurance disciplines. Plans must be established that define the need to analyze a design to assure it contains characteristics that allow efficient and effective fault detection and isolation.

Ensuring that a product is testable requires adherence to some basic testability design principles. A brief description of the most common testability design principles follows.

- Physical and functional partitioning - The ease or difficulty of fault isolation depends to a large extent upon the size and complexity of the units that are replaceable. Partitioning the design such that components are grouped by function (i.e., each function is implemented on a single replaceable unit), or by technology (e.g., analog, digital) whenever possible will enhance the ability to isolate failures.

- Electrical partitioning - Whenever possible, a block of circuitry being tested should be isolated from circuitry not being tested via blocking gates, tristate devices, relays, etc.

- Initialization - The design should allow an item to be initialized to a known state so it will respond in a consistent manner for multiple testing of a given failure.

- Controllability - The design should allow external control of internal component operation for the purpose of fault detection and isolation. Special attention should be given to independent control of clock signals, the ability to control and break up feedback loops, and tri-stating components for isolation.

- Observability - Sufficient access to test points, data paths and internal circuitry should be provided to allow the test system (machine or human) to gather sufficient signature data for fault detection and isolation.

- Test System Compatibility - Each item to be tested should be designed to be electrically and mechanically compatible with selected or available test equipment to eliminate or reduce the need for a large number of interface device (ID) designs.

In addition to the preceding principles, checklists of testability design practices have been developed that are specific to technologies, such as analog, digital, mechanical, and so forth. See 7.12.6.1.2 for one such checklist.

Determining the amount of testability necessary in a design will be driven by the requirements for fault *detection* and fault *isolation*. Fault detection requirements are typically stated as the percentage of faults that can be detected, using defined means (BIT, semi-automatic/automatic test, etc.), out of all possible faults. For instance, a system may have a requirement of 95% fault detection, indicating that 95% of all possible failures are to be detectable by the diagnostic capability of the system. Fault isolation requirements are typically stated as the percentage of time fault isolation is possible to a specified number of components. As an example, a system may have a requirement of 90% isolation to a single replaceable unit (RU), 95% isolation to an ambiguity group of 2 or fewer RUs and 100% isolation to an ambiguity group of 3 or fewer RUs.

Mathematically, fault detection and isolation are defined in the following equations for the fraction of faults detectable (FFD) and the fraction of faults isolatable (FFI).

$$FFD = FD/FA$$

where:

FA = total number of actual faults occurring over time

FD = no. of actual failures correctly identified using defined means

Equation 1 is used to calculate predicted fault resolution. To use the equation, data are required that correlate each detected failure with the signature, or "error syndrome", that each failure produces during testing. The data are most conveniently ordered by signature and by failed module within each signature. The signature, then, is the observed test response when a particular failure occurs. This information typically is generated from an FMEA, or in the case of electronics design, especially digital, from a fault simulation program. The collection of test responses, or failure signatures, represents a fault dictionary. In many instances, several failures will produce the same observed (usually at the system output(s)) signature, creating ambiguity. The fault resolution predicted by equation 1 measures the amount of ambiguity that exists, for a given level of test capability. As noted, for each signature, a list of suspect modules is created, providing the input data needed to apply the following equation:

$$FFI_L = \left(\frac{100}{\lambda_d}\right) \sum_{i=1}^{N} X_i \sum_{j=1}^{M_i} \lambda_{ij}$$

where:

X_i = 1 if $M_i \leq L$; 0 otherwise
N = number of unique test responses
L = number of modules isolated to (i.e., ambiguity group size)
i = signature index
M_i = number of modules listed in signature i
j = module index within signature
λ_{ij} = failure rate for jth module for failures having signature i

λ_d = overall failure rate of detected failures = $\sum_{i=1}^{N} \sum_{j=1}^{M_i} \lambda_{ij}$

Additional quantitative measures of testability may include fault isolation time, which is derived from the Mean Time To Repair (MTTR).

Mean Fault isolation time = Mean [repair time - (operation time + disassembly time + interchange time + reassembly time + alignment time + verification time)]

Note that the first two measures are interrelated in that before you can isolate a fault, you must first detect it. Therefore, a testability analysis program is designed to analyze the effectiveness of the *detection* scheme, and then to analyze the effectiveness of the *isolation* scheme. For complex designs, the analysis of testability often requires the use of testability design and analysis tools that provide information on fault detection and isolation, for a given diagnostic approach, or diagnostic capability.

False alarms (in which a failure is “detected” even though none occurred) is a problem related to both testability and a system's diagnostic design. Manifesting themselves in varying degrees in avionics and other types of equipment, false alarms are a drain on maintenance resources and reduce a system’s mission readiness. The two most commonly reported symptoms of false alarms are CND and RTOK.

False alarms occur for many reasons, including external environmental factors (temperature, humidity, shock, etc.), design of diagnostics, equipment degradation due to age, design tolerance factors, maintenance-induced factors (e.g., connectors, wire handling, etc.), or combinations of these factors. External environmental factors may cause failures of avionics or other equipment that do not occur under ambient conditions and are believed to be a leading cause of false alarms. When the environmental condition is removed, the “failure” cannot be found. One solution to the problem is to use a stress measurement device to record the environmental stresses before, during, and after a system anomaly. Subsequent diagnosis can use this data to determine what occurred and whether any action (maintenance, modifications, etc.) is needed.

The Time Stress Measurement Device (TSMD) is a stress measurement device that has been developed over the past few years by the Air Force. TSMDs measure and record selected environmental parameters and fault signatures and record a time stamp, for use in subsequent failure correlation analysis. TSMD has been adapted to record an image of all of the environmental data prior to, during, and after a system anomaly. These recorded events can be used to identify environmental stress-related conditions that may be causing intermittent or hard failures. The TSMD data aids in reducing RTOK, and CND conditions by correlating the event with the conditions that existed when the anomaly was detected.

Several different models of TSMDs have been developed by different manufacturers. They feature both 8 bit (Ref. [96]) and 32 bit (Ref. [97]) internal microprocessors and RS-232 and RS-485 interfaces. Typically they are powered by 5 volts DC drawn from the host system and dissipate 1 watt or less. They may be powered by an external battery for operation under power-off conditions, e.g., shipping or storage, or when host system power is impractical or too costly to provide.

Many commercial stress measurement devices are also in use or under study. A RAC publication (Ref. [98]) provides a compendium of such commercially available devices, including their sensing and storing capabilities. This publication is part of an on-going market survey aimed at identifying sources of stand-alone environmental stress data collection systems.

7.12.4 Developing a Diagnostic Capability

Defining and developing a product's diagnostic capability depends on factors such as:

- The product's performance and usage requirements
- Maintenance support requirements (e.g., levels of maintenance)
- Technology available to: improve diagnostics in terms of test effectiveness; reduce the need for test equipment, test manuals, personnel, training, and skill levels; and reduce cost
- The amount of testability designed into the product
- Previously known diagnostic problems on similar systems

Each of these factors will play a role in determining the approach to detecting and isolating faults. A typical approach to diagnostics includes the use of BIT. BIT is an integral capability of the mission equipment which provides an on-board, automated test capability. This capability consists of software or hardware (or both) components that detect, diagnose, or isolate product (system) failures. The fault detection and isolation capability is used for periodic or continuous monitoring of a system's operational health, and for observation and diagnosis as a prelude to maintenance action. BIT reduces the need for maintenance manpower and External Test Equipment. Other approaches may consider the use of automatic or semi-automatic test equipment, manual testing using benchtop test equipment, or visual inspection procedures. In all cases, tradeoffs are required among system performance, cost, and test effectiveness.

It must be remembered that the effectiveness of the diagnostic capability, and the cost of development, is greatly influenced by how well testability has been designed into the system. Should there be a lack of test points available to external test equipment, for example, then the ability to isolate failures to smaller ambiguity group sizes may be adversely affected. The result is higher costs to locate the failure to a single replaceable item. The cost of test development may also increase. BIT design should be supported by the results of a failure modes and effects analysis (FMEA). An FMEA should be used to define those failures that are critical to system performance, and to identify when the effects of a failure can be detected using BIT. Without such information, BIT tests will be developed based only on the test engineer's knowledge of how the system works, and not on whether a test needs to be developed for a particular fault.

7.12.5 Designing BIT

Planning for BIT at all levels within the system design is becoming more important for several reasons. First, surface mount devices (SMDs) are increasingly being used in circuit cards. The use of SMDs, and devices with higher packaging density (including double-sided boards), decreases the accessibility required for guided-probe testing, while increasing the risks of such testing. Incorporating BIT in such designs therefore becomes critical to effective diagnostics. Second, many component vendors of integrated circuits (ICs), such as Application Specific ICs (ASICs) are incorporating some form of BIT into their designs. Higher-level designs (i.e., board, module, etc.) that use such devices must take advantage of this fact by planning to integrate lower-level BIT capabilities with higher-level BIT designs. Doing this will increase the vertical testability of an entire system, wherein factory-level test programs can be used in field operations as well as the factory. Further, tests performed using BIT at higher levels of support (e.g., depot or intermediate) can also be used at lower levels (i.e., intermediate and organizational). This characteristic of the diagnostic system will help to maintain consistency across maintenance levels and may reduce the high incidences of false alarms. False alarms are often reflected by such measures as Retests OK (RTOK) or Can Not Duplicates (CNDs). (Note that not all the military services either use these terms or define them the same way).

The most important factor in BIT design is early planning. Without planning for BIT early in the life cycle, it will be harder to maximize any advantages offered by the use of BIT while minimizing any negative impacts such as increased design cost, higher hardware overhead, and increased failure rate. In “Chip-To-System Testability” (Interim Report submitted to Rome Laboratory under Contract No. F30602-94-C0053, 1996, Research Triangle Institute and Self-Test Services), five axioms are given that will allow designers to capitalize on the use of BIT. These axioms are:

- Plan for BIT starting at the earliest stage (e.g., proposal stage) of the program
- Design BIT in conjunction with the functional design, not as an afterthought
- Use the same high degree of engineering cleverness and rigor for BIT that is used for the functional design
- Take advantage of computer aided design (CAD) tools for the BIT design process whenever possible
- Incorporate the subject of BIT into peer, design and program reviews

BIT must be a part of the product’s design to avoid the risks and consequences shown in Table 7.12-1.

TABLE 7.12-1: RISKS AND CONSEQUENCES OF NOT MAKING BIT PART OF PRODUCT DESIGN

Risks	**Consequences**
BIT is designed independently of the product	BIT fails to support operational and maintenance needs
BIT is designed after the fact	BIT's MTBF is less than that of the product
Production personnel are not consulted on BIT	BIT is not effective in the factory

7.12.6 Testability Analysis

Testability analysis is important at all levels of design and can be accomplished in a variety of ways. For instance, when designing complex integrated circuits (ICs), such as Application Specific ICs, or ASICs, it is important to develop test vectors that will detect a high percentage of 'stuck at' faults (i.e., signal stuck at logic '1' or '0'). This is almost always determined via logic simulation wherein a model of the design is developed in an appropriate fault simulation language. Once the model is compiled and ready to be simulated, a set of test vectors are applied to the model. The fault simulation program then produces a list of faults detected by the test vectors, as well as reporting the percentage (or fraction) of faults detected. Many such programs also identify specific signals that were not detected such that adjustments can be made either in the design or in the test vectors themselves in order to increase the fault detection percentage.

For non-digital electronics, fault detection efficiency is typically determined with the aid of an FMEA. The FMEA will identify those faults that result in an observable failure and can therefore be detected. The test engineer then must develop a test that will verify operation and detect any malfunctions identified in the FMEA. Fault detection percentages are determined by summing the number of faults identified in the FMEA that are detected versus the total number identified as being detectable. This process can occur at all levels of design. The fault grading methods described in the preceding paragraph are primarily applied at the IC and printed circuit card levels.

In addition to determining fault detection percentage, a testability analysis should be performed to determine the fault isolation effectiveness of designed tests. For digital electronics, many of the tools used to grade test vectors also provide statistics on fault isolation percentages. This is typically provided by creating a fault dictionary. During fault simulation, the response of the circuit is determined in the presence of faults. These responses collectively form the fault dictionary. Isolation is then performed by matching the actual response obtained from the circuit or test item with one of the previously computed responses stored in the fault dictionary. Fault simulation tools can determine from the fault dictionary the percentage of faults that are uniquely isolatable to an ambiguity group of size n (n = 1, 2, 3, ...). These tools can be used to verify fault isolation goals or requirements via analysis, prior to actual testing. For non-digital circuits, hybrid circuits or even digital systems above the printed circuit card level, analysis of fault isolation capability can be performed with the aid of a diagnostic model and a software tool that analyzes that model. Examples are dependency modeling tools such as the Weapon System

Testability Analyzer (WSTA), System Testability Analysis Tool (STAT) or the System Testability and Maintenance Program (STAMP)[7]. These tools, and others like them, can be used to determine the fault isolation capability of a design based on the design topology, order of test performance, and other factors such as device reliability. Statistics such as percentage of faults isolatable to an ambiguity of group size n are provided, as is the identification of which components or modules are in an ambiguity group for a given set of tests. Test effectiveness and model accuracy are the responsibility of the test designer, however.

7.12.6.1 Dependency Analysis

Assessing testability via dependency analysis has gained in popularity recently, and it is therefore prudent to provide some additional information on this technique. Dependency analysis starts with the creation of a dependency model of the item to be analyzed. The model is designed to capture the relationship between tests or test sites within a system, and those components and failure modes of components that can affect the test. As an example, consider the simple functional block diagram shown in Figure 7.12-1.

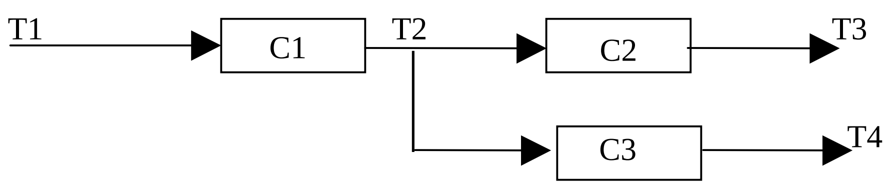


FIGURE 7.12-1: SIMPLE SYSTEM SHOWING TEST DEPENDENCIES

The dependency model for the system, in the form of a tabular list of tests and their dependencies is provided in Table 7.12-2.

TABLE 7.12-2: FIRST ORDER DEPENDENCY MODEL FOR SIMPLE SYSTEM

Test	First-Order Dependencies
T1	None
T2	C1, T1
T3	C2, T2
T4	C3, T2

Figure 7.12-1 has been labeled to identify each potential test site within the system, where in this example, exactly one test is being considered at each node. The dependency model shown in Table 7.12-2 is a list of "first-order dependencies" of each test. For example, the first order dependency of test T3 is C2 and T2. This would indicate that T3 *depends* upon the health of component C2 and any inputs to C2, which is T2 in this case. For this simple system, it is also

[7] STAT is a registered trademark of DETEX Systems, Inc. and STAMP is a registered trademark of the ARINC Research Corporation. WSTA is a tool developed by the US Navy and available to most US Government contractors and US Government employees.

obvious that T3 will also depend on C1 and T1, but these are considered higher-order dependencies. Each of the tools mentioned previously (i.e., STAT, STAMP and WSTA), determine all higher order dependencies based on a first order dependency input model.

Dependency modeling is attractive due to its applicability to any kind or level of system. Note in the example that neither the nature nor level of the system is required to process the model. Consequently, this methodology is applicable to most any type of system technology and any level (i.e., component to system).

Based on the input model, the analysis tools can determine the percentage of time isolation to an ambiguity group of n or fewer components will occur. In addition, each of the tools discussed will also identify which components or failures will be in the same ambiguity group with other components or failures. Furthermore, any test feedback loops that exist, including those components contained within the feedback loop, will also be identified. Note that the ambiguity group sizes and statistics are based on a binary test outcome (i.e., test is either good or bad), and in most cases the tools assume that the test is 100% effective. This means that if the model indicates that a particular test depends on a specified set of components, the tools assume that should the test pass, all components within the dependency set are good. Conversely, a failed test makes all of the components within the dependency set suspect. Therefore, the accuracy of the model, in terms of what components and component failure modes are actually covered by a particular test are the responsibility of the model developer. The coverage is very much dependent upon test design and knowledge of the system's functional behavior.

Even before intimate knowledge of what tests are to be performed is known, such as in the early stages of system development, a model can be created that assumes a test at every node, for instance. The system design can be evaluated as to where feedback loops reside, which components are likely to be in ambiguity, and where more visibility, in terms of additional test points, need to be added to improve the overall testability of the design. Once the design is more developed, and knowledge of each test becomes available, the dependency model can then be refined. Given that the analyst is satisfied with the model results, each of the tools discussed can be used to develop optimal test strategies based on system topology and one or more weighting factors such as test cost, test time, component failure rates, time to remove an enclosure to access a test point, etc.

One of the drawbacks in the past to dependency modeling has been the time it takes to create a model. However, translation tools exist and are continuously being developed that can translate a design captured in a CAD format, such as the Electronic Data Interchange Format (EDIF), into a dependency model compatible with the specific dependency analysis tool being used. The analyst is still responsible for verifying the accuracy of the model, however, as in some cases, not all dependencies will be 100% correctly translated. Despite this fact, the amount of time that can be saved in translation out weighs any additional time it may take to verify the model.

7.12.6.1.1 Dependency Analysis Tools

The three tools mentioned, STAT, STAMP and WSTA, provide the same basic kinds of outputs as just discussed. Each tool has other features that may be attractive depending on the system being analyzed, CAD tools being used in the design process, etc. Therefore, more information should be gathered on these and other similar tools prior to making a final decision as to which one to acquire.

The key points to remember about any of these tools is that model accuracy is most important. Therefore, it is important to understand how the system behaves in the presence of a failure, and which tests can be developed to detect such behavior. Thus, to gain the most benefit from the model development process, experts in design and test should be involved.

7.12.6.2 Other Types of Testability Analyses

Other types of analyses that do not require the use of a software tool are ad hoc procedures, such as reviewing a design against a known set of testability design practices. Grumman, and later Raytheon, developed such a procedure for the US Air Force Rome Laboratory that rates a design based on the presence or absence of design features that increase or decrease ease of test. The result is a score that is subjectively evaluated as indicating the design is anywhere between untestable without redesign to very testable. Used in conjunction with a design guide, also developed as part of the process by the mentioned companies, this method can be very effective in making the test engineer's job easier and less costly. The report, RL-TR-92-12 (Ref. [99]), VOLUMES I & II - Testability Design Rating System: Testability Handbook (VOL. I) & Analytical Procedure (VOL. II), include testability design.

In addition to specific diagnostics testability and diagnostics guidelines, RL-TR-92-12 provides the following general guidance regarding testability.

Redundancy - Built-in-Test (BIT) can be implemented by repeating the functional circuitry (the redundancy) to be tested by BIT. The same functional signal(s) is input into the redundant element and Circuit Under Test (CUT). Therefore, the circuitry of the CUT exists twice in the design and the outputs can be compared. If the output values are different and their difference exceeds a limit (analog circuits), then a fault exists. Due to the expense of this technique, redundant BIT design is usually implemented only in critical functions

An example of a BIT design using redundancy is shown in Figure 7.12-2. In this example, an analog circuit is repeated and the difference between the output levels is compared. If the difference exceeds a predefined threshold, then a fault signal is generated and latched.

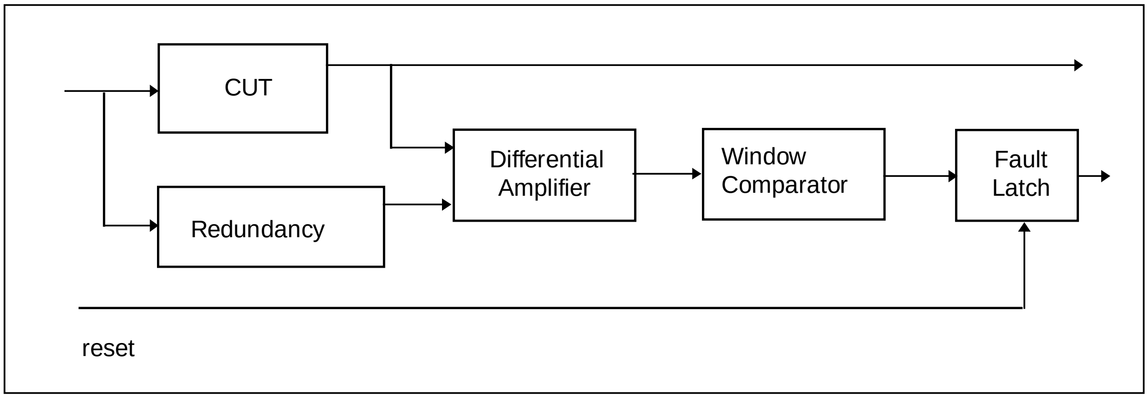


FIGURE 7.12-2: REDUNDANCY BIT (SOURCE: RADC-TR-89-209, VOL. II)

Wrap-around BIT - Wrap-around BIT requires and tests microprocessors and their input and output devices. During test, data leaving output devices is routed to input devices of the module. The BIT routine is stored in on-board read-only memory (ROM). Wrap-around can be done by directing output signals from the processor back to the input signals and verifying the input signal values. Wrap-around BIT can be applied to both digital and analog signals concurrently. An example of wrap-around BIT testing both analog and digital devices is shown in Figure 7.12-3. In this example, during normal operation processor outputs are converted from digital to analog outputs and analog inputs are converted to digital input signals. When the BIT is initiated, the analog outputs are connected to the analog inputs and the signals are verified by the processor.

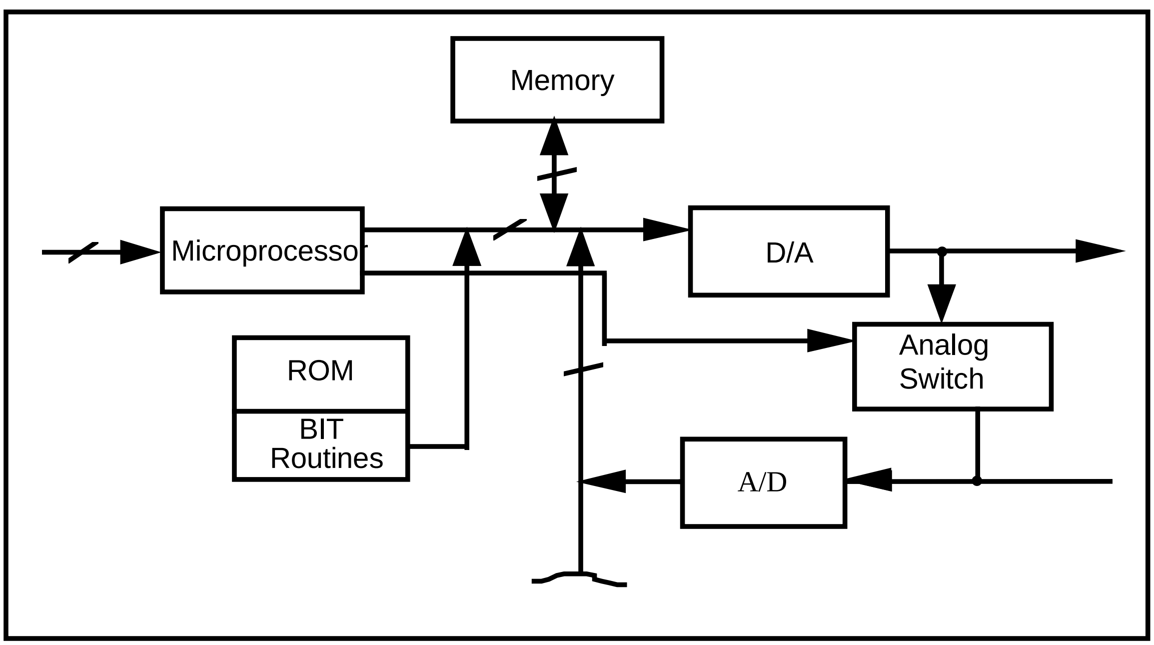


FIGURE 7.12-3: WRAP-AROUND BIT (SOURCE: RADC-TR-89-209, VOL II)

The remainder of RL-TR-92-12, VOL I, provides detailed guidance on testability design techniques and structured test techniques for various categories of part types and technologies.

In addition to the practical design guide information found in RL-TR-92-12, VOL I, Reference [100], provides an inherent testability checklist. It is reprinted here, in a slightly different format, as Table 7.12-3. Refer to Reference [100] for further guidance on testability program planning.

7.13 System Safety Program

7.13.1 Introduction

Reliability and safety are closely related subjects. Many of the analyses are complementary. For these reasons, a discussion of a system safety program is included here.

The principal objective of a system safety program is to ensure that safety, consistent with mission requirements, is designed into systems, subsystems, equipment and facilities, and their interfaces.

Within the DoD, MIL-STD-882, "System Safety Program Requirements," provides uniform guidelines for developing and implementing a system safety program of sufficient comprehensiveness to identify the hazards of a system and to impose design requirements and management controls to prevent mishaps by eliminating hazards or reducing the associated risk to a level acceptable to the managing activity.

Four different types of program elements are addressed: (a) Program Management and Control Elements, (b) Design and Integration Elements, (c) Design Evaluation Elements and (d) Compliance and Verification Elements.

(a) Program Management and Control Elements are those relating primarily to management responsibilities dealing with the safety of the program and less to the technical details involved.

(b) Design and Integration Elements focus on the identification, evaluation, prevention, detection, and correction or reduction of the associated risk of safety hazards by the use of specific technical procedures.

(c) Design Evaluation Elements focus on risk assessment and the safety aspects of tests and evaluations of the system and the possible introduction of new safety hazards resulting from changes.

(d) Compliance and Verification Elements are those directly related to the actual verification or demonstration that all legal and contractual safety requirements have been compiled with.

TABLE 7.12-3: INHERENT TESTABILITY CHECKLIST

Mechanical Design Checklist (for electronic designs)	
• Is a standard grid layout used on boards to facilitate identification of components? • Are the number of I/O pins in an edge connector or cable connector compatible with the I/O capabilities of the selected test equipment? • Are connector pins arranged such that the shorting of physically adjacent pins will cause minimum damage? • Is the design free of special set-up requirements (special cooling) which would slow testing? • Does the item warm up in a reasonable amount of time? • Has provision been made to incorporate a test-header connector into the design to enhance ATE testing of surface-mounted devices?	• Is defeatable keying used on each board so as to reduce the number of unique interface adapters required? • Is each hardware component clearly labeled? • Are all components oriented in the same direction (pin 1 always in same position)? • Does the board layout support guided-probe testing techniques? • When possible, are power and ground included in the I/O connector or test connector? • Have test and repair requirements impacted decisions on conformal coating? • Is enough spacing provided between components to allow for clips and test probes?
Partitioning Checklist (for electronic functions)	
• Is each function to be tested placed wholly upon one board? • Within a function, is the size of each block of circuitry to be tested small enough for economical fault detection and isolation? • Is the number of power supplies required compatible with the test equipment? • If more than one function is placed on a board, can each be tested independently?	• If required, are pull up resistors located on the same board as the driving component? • Is the number and type of stimuli required compatible with the test equipment? • Within a function, can complex digital and analog circuitry be tested independently? • Are analog circuits partitioned by frequency to ease tester compatibility? • Are elements which are included in an ambiguity group placed in the same package?
Test Control Checklist	
• Are connector pins not needed for operation used to provide test stimulus and control from the tester to internal nodes? • Is it possible to disable on-board oscillators and drive all logic using a tester clock? • Is circuitry provided to by-pass any (unavoidable) one-shot circuitry? • In microprocessor-based systems, does the tester have access to the data bus, address bus and important control lines? • Are active components, such as demultiplexers and shift registers, used to allow the tester to control necessary internal nodes using available input pins? • Can circuitry be quickly and easily driven to a known initial state? (master clear, less than N clocks for initialization sequence)?	• Can long counter chains be broken into smaller segments in test mode with each segment under tester control? • Can feedback loops be broken under control of the tester? • Are test control points included at those nodes which have high fan-in (test bottlenecks)? • Are redundant elements in design capable of being independently tested? • Can the tester electrically partition the item into smaller independent, easy-to-test segments? (placing tri-state element in a high impedance state). • Have provisions been made to test the system bus as a stand-alone entity? • Are input buffers provided for those control point signals with high drive capability requirements?

TABLE 7.12-3 INHERENT TESTABILITY CHECKLIST (CONT'D)

Parts Selection Checklist	
• Is the number of different part types the minimum possible? • Is a single logic family being used? If not, is a common signal level used for interconnections?	• Have parts been selected which are well characterized in terms of failure modes? • Are the parts independent of refresh requirements? If not, are dynamic devices supported by sufficient clocking during testing?
Test Access	
• Are unused connector pins used to provide additional internal node data to the tester? • Are test access points placed at those nodes which have high fan-out? • Are active components, such as multiplexers and shift registers, used to make necessary internal node test data available to the tester over available output pins? • Are signal lines and test points designed to drive the capacitive loading represented by the test equipment? • Are buffers employed when the test point is a latch and susceptible to reflections?	• Are all high voltages scaled down within the item prior to providing test point access so as to be consistent with tester capabilities? • Are test points provided such that the tester can monitor and synchronize to onboard clock circuits? • Are buffers or divider circuits employed to protect those test points which may be damaged by an inadvertent short circuit? • Is the measurement accuracy of the test equipment adequate compared to the tolerance requirement of the item being tested?
Analog Design Checklist	
• Is one test point per discrete active stage brought out to the connector? • Are circuits functionally complete without bias networks or loads on some other UUT? • Is a minimum number of complex modulation or unique timing patterns required? • Are response rise time or pulse width measurements compatible with test capabilities? • Does the design avoid or compensate for temperature sensitive components? • Is each test point adequately buffered or isolated from the main signal path? • Is a minimum number of multiple phase-related or timing-related stimuli required?	• Are stimulus frequencies compatible with tester capabilities? • Are stimulus amplitude requirements within the capability of the test equipment? • Does the design allow testing without heat sinks? • Are multiple, interactive adjustments prohibited for production items? • Is a minimum number of phase or timing measurements required? • Do response measurements involve frequencies compatible with tester capabilities? • Does the design avoid external feedback loops? • Are standard types of connectors used?
Performance Monitoring Checklist	
• Have critical functions been identified (by FMECA) which require monitoring for the system operation and users? • Have interface standards been established that ensure the electronic transmission of data from monitored systems is compatible with centralized monitors?	• Has the displayed output of the monitoring system received a human engineering analysis to ensure that the user is supplied with the required information in the best useable form?

TABLE 7.12-3: INHERENT TESTABILITY CHECKLIST (CONT'D)

RF Design Checklist	
• Do transmitter outputs have directional couplers or similar signal sensing/attenuation techniques employed for BIT or off-line test monitoring purposes, or both? • Has provision been made in the off-line ATE to provide switching of all RF stimulus and response signals required to test the subject RF UUT? • Are the RF test input/output access ports of the UUT mechanically compatible with the off-line ATE I/O ports? • Have adequate testability (controllability/ observability) provisions for calibrating the UUT been provided? • If an RF transmitter is to be tested utilizing off-line ATE, has suitable test fixturing (anechoic chamber) been designed to safely test the subject item over its specified performance range of frequency and power? • Have all RF testing parameters and quantitative requirements for these parameters been explicitly stated at the RF UUT interface for each RF stimulus/ response signal to be tested? • Has the UUT/ATE RF interface been designed so that the system operator can quickly and easily connect and disconnect the UUT without special tooling?	• Have RF compensation procedures and data bases been established to provide calibration of all stimulus signals to be applied and all response signals to be measured by BIT or off-line ATE to the RF UUT interface? • Have suitable termination devices been employed in the off-line ATE or BIT circuitry to accurately emulate the loading requirements for all RF signals to be tested? • Does the RF UUT employ signal frequencies or power levels in excess of the core ATE stimulus/ measurement capability? If so, are signal converters employed within the ATE to render the ATE/UUT compatible? • Has the RF UUT been designed so that repair or replacement of any assembly or subassembly can be accomplished without major disassembly of the unit? • Does the off-line ATE or BIT diagnostic software provide for compensation of UUT output power and adjustment of input power, so that RF switching and cable errors are compensated for in the measurement data?
Electro-optical (EO) Design Checklist	
• Have optical splitters/couplers been incorporated to provide signal accessibility without major disassembly? • Has temperature stability been incorporated into fixture/UUT design to assure consistent performance over a normal range of operating environments? • Have optical systems been functionally allocated so that they and associated drive electronics can be independently tested? • Are the ATE system, light sources, and monitoring systems of sufficient wave-length to allow operation over a wide range of UUTs? • Does the test fixturing intended for the off-line test present the required mechanical stability? • Is there sufficient mechanical stability and controllability to obtain accurate optical registration? • Can requirements for boresighting be automated or eliminated?	• Do monitors possess sufficient sensitivity to accommodate a wide range of intensities? • Can optical elements be accessed without major disassembly or realignment? • Do they possess sufficient range of motion to meet a variety of test applications? • Has adequate filtering been incorporated to provide required light attenuation? • Can all modulation models be simulated, stimulated, and monitored? • Can targets be automatically controlled for focus and aperture presentation? • Do light sources provide enough dynamics over the operating range? • Do test routines and internal memories test pixels for shades of gray? • Are optical collimators adjustable over their range of motion via automation?

TABLE 7.12-3 INHERENT TESTABILITY CHECKLIST (CONT'D)

Digital Design Checklist	
• Does the design contain only synchronous logic? • Does the design avoid resistance capacitance one-shots and dependence upon logic delays to generate timing pulses? • Is the design free of WIRED-ORs? • Will the selection of an unused address result in a well defined error state? • Are all clocks of differing phases and frequencies derived from a single master clock? • Is the number of fan-outs for each board output limited to a predetermined value? Are latches provided at the inputs to a board in those cases where tester input skew could be a problem? • For multilayer boards, is the layout of each major bus such that current probes or other techniques may be used for fault isolation beyond the node?	• If the design incorporates a structured testability design technique (scan path, signature analysis), are all the design rules satisfied? • Is the number of fan-outs for each internal circuit limited to a predetermined value? • Are all memory elements clocked by a derivative of the master clock? (Avoid elements clocked by data from other elements.) • Does the design include data wrap-around circuitry at major interfaces? • Is a known output defined for every word in a read only memory? • Are sockets provided for microprocessors and other complex components? • Does the design support testing of "bit slices"? • Do all buses have a default value when unselected?
Diagnostic Capability Integration	
• Have vertical testability concepts been established, employed, and documented? • Has the diagnostic strategy (dependency charts, logic diagrams) been documented?	• Has a means been established to ensure compatibility of testing resources with other diagnostic resources at each level of maintenance (technical information, personnel, and training)?
Mechanical Systems Condition Monitoring (MSCM) Checklist	
• Have MSCM and battle damage monitoring functions been integrated with other performance monitoring functions?	• Are preventive maintenance monitoring functions (oil analysis, gear box cracks) in place? • Have scheduled maintenance procedures been established?
Sensors Checklist	
• Are pressure sensors placed very close to pressure sensing points to obtain wideband dynamic data? • Has the selection of sensors taken into account the environmental conditions under which they will operate?	• Have procedures for calibration of sensing devices been established? • Has the thermal lag between the test media and sensing elements been considered?
Test Requirements Checklist	
• Has a "level of repair analysis" been accomplished? • For each maintenance level, has a decision been made for each item on how BIT, ATE, and General Purpose Electronic Test Equipment (GPETE), will support fault detection and isolation?	• For each item, does the planned degree of testability design support the level of repair, test mix, and degree of automation decisions? • Is the planned degree of test automation consistent with the capabilities of the maintenance technician?

TABLE 7.12-3 INHERENT TESTABILITY CHECKLIST (CONT'D)

Built-in-Test (BIT) Checklist	
• Can BIT in each item be exercised under control of the test equipment? • Does the BIT use a building-block approach (all inputs to a function are verified before that function is tested)? • Does on-board ROM contain self-test routines? • Does BIT include a method of saving on-line test data for the analysis of intermittent failures and operational failures which are non-repeatable in the maintenance environment? • Is the additional volume due to BIT within stated constraints? • Does the allocation of BIT capability to each item reflect the relative failure rate of the items and the criticality of the items' functions? • Are the data provided by BIT tailored to the differing needs of the system operator and the system maintainer? • Is sufficient memory allocated for confidence tests and diagnostic software? • Are BIT threshold limits for each parameter determined as a result of considering each parameter's distribution statistics, the BIT measurement error and the optimum fault detection/false alarm characteristics? • Is BIT optimally allocated in hardware, software, and firmware? • Have means been established to identify whether hardware or software has caused a failure indication?	• Is the failure latency associated with a particular implementation of BIT consistent with the criticality of the function being monitored? • Is the test program set designed to take advantage of BIT capabilities? • Does building-block BIT make maximum use of mission circuitry? • Is the self-test circuitry designed to be testable? • Is the predicted failure rate contribution of the BIT circuitry within stated constraints? • Is the additional power consumption due to BIT within stated constraints? • Are BIT threshold values, which may require changing as a result of operational experience, incorporated in software or easily-modified firmware? • Are on-board BIT indicators used for important functions? Are BIT indicators designed such that a BIT failure will give a "fail" indication? • Is the additional weight due to BIT within stated constraints? • Is the additional part count due to BIT within stated constraints? • Is processing or filtering of BIT sensor data performed to minimize BIT false alarms? • Does mission software include sufficient hardware error detection capability?
Test Data Checklist	
• Do state diagrams for sequential circuits identify invalid sequences and indeterminate outputs? • For computer-assisted test generation, is the available software sufficient in terms of program capacity, fault modeling, component libraries, and post-processing of test response data? • If a computer-aided design system is used for design, does the CAD data base effectively support the test generation process and test evaluation process? • Is the tolerance band known for each signal on the item?	• Are testability features included by the system designer documented in the Test Requirement Document (TRD) in terms of purpose and rationale for the benefit of the test designer? • For large scale ICs used in the design, are data available to accurately model the circuits and generate high-confidence tests? • Are test diagrams included for each major test? Is the diagram limited to a small number of sheets? Are inter-sheet connections clearly marked?

7.13.2 Definition of Safety Terms and Acronyms

The meanings of some terms and acronyms are unique to this section and are therefore included here to aid the reader.

Fail Safe: A design feature that either ensures that the system remains safe, or, in the event of a failure, forces the system to revert to a state which will not cause a mishap.

Hazard: A condition that is prerequisite to a mishap.

Hazard Probability: The aggregate probability of occurrence of the individual events that create a specific hazard.

Hazardous Material: Anything that due to its chemical, physical, or biological nature causes safety, public health, or environmental concerns that result in an elevated level of effort to manage.

Mishap: An unplanned event or series of events that result in death, injury, occupational illness, or damage to or loss of equipment or property or damage to the environment. An accident.

Risk: An expression of the possibility of a mishap in terms of hazard severity and hazard probability.

Risk Assessment: A comprehensive evaluation of the risk and its associated impact.

Safety: Freedom from those conditions that can cause death, injury, occupational illness, or damage to or loss of equipment or property or damage to the environment.

Safety Critical: A term applied to a condition, event, operation, process or item of whose proper recognition, control, performance or tolerance is essential to safe operation or use; e.g., safety critical function, safety critical path or safety critical component.

Safety-Critical Computer Software Components: Those computer software components and units whose errors can result in a potential hazard, or loss of predictability or control of a system.

System Safety: The application of engineering and management principles, criteria, and techniques to optimize safety within the constraints of operational effectiveness, time, and cost throughout all phases of the system life cycle.

7.13.3 Program Management and Control Elements

7.13.3.1 System Safety Program

A basic system safety program consists of the following safety-related elements.

7.13.3.2 System Safety Program Plan

This plan describes in detail those elements and activities of safety system management and system safety engineering required to identify, evaluate, and eliminate hazards, or reduce the associated risk to a level acceptable to the managing activity throughout the system life cycle. It normally includes a description of the planned methods to be used to implement a system safety program plan, including organizational responsibilities, resources, methods of accomplishment, milestones, depth of effort, and integration with other program engineering and management activities and related systems.

7.13.3.3 Integration/Management of Associate Contractors, Subcontractors, and Architect and Engineering Firms

This element consists of appropriate management surveillance procedures to ensure uniform system safety requirements are developed.

7.13.3.4 System Safety Program Reviews/Audits

This element is a forum for reviewing the system safety program, to periodically report the status of the system safety program, and, when needed, to support special requirements, such as certifications and first flight readiness reviews.

7.13.3.5 System Safety Group/System Safety Working Group Support

This element is a forum for suppliers and vendors to support system safety groups (SSGs) and system safety working groups (SSWGs) established in accordance with government regulations or as otherwise defined by the integrating supplier.

7.13.3.6 Hazard Tracking and Risk Resolution

This element is a single closed-loop hazard tracking system to document and track hazards from identification until the hazard is eliminated or the associated risk is reduced to an acceptable level.

7.13.3.7 System Safety Progress Summary

This element consists of periodic progress reports summarizing the pertinent system safety management and engineering activity that occurred during the reporting period.

7.13.4 Design and Integration Elements

7.13.4.1 Preliminary Hazard List

This element is a preliminary hazard list (PHL) identifying any especially hazardous areas for added management emphasis. The PHL should be developed very early in the development phase of an item.

7.13.4.2 Preliminary Hazard Analysis

The purpose of the Preliminary Hazard Analysis (PHA) is to identify safety critical areas, evaluate hazards, and identify the safety design criteria to be used.

7.13.4.3 Safety Requirements/Criteria Analysis

The Safety Requirements/Criteria Analysis (SRCA) relates the hazards identified to the system design and identifies or develops design requirements to eliminate or reduce the risk of the hazards to an acceptable level. The SRCA is based on the PHL or PHA, if available. The SRCA is also used to incorporate design requirements that are safety related but not tied to a specific hazard.

7.13.4.4 Subsystem Hazard Analysis

The Subsystem Hazard Analysis (SSHA) identifies hazards associated with design of subsystems including component failure modes, critical human error inputs, and hazards resulting from functional relationships between components and equipments comprising each subsystem.

7.13.4.5 System Hazard Analysis

The System Hazard Analysis (SHA) documents the primary safety problem areas of the total system design including potential safety critical human errors.

7.13.4.6 Operating and Support Hazard Analysis

The Operating and Support Hazard Analysis (O&SHA) identifies associated hazards and recommends alternatives that may be used during all phases of intended system use.

7.13.4.7 Occupational Health Hazard Assessment

The Occupational Health Hazard Assessment (OHHA) identifies human health hazards and proposes protective measures to reduce the associated risks to levels acceptable to the managing activity.

7.13.5 Design Evaluation Elements

7.13.5.1 Safety Assessment

This element is a comprehensive evaluation of the mishap risk that is being assumed prior to the test or operation of a system or at the contract completion.

7.13.5.2 Test and Evaluation Safety

The purpose of this element is to ensure that safety is considered (and safety responsibility assigned) in test and evaluation, to provide existing analysis reports and other safety data, and to respond to all safety requirements necessary for testing in-house, at other supplier facilities, and at Government ranges, centers, or laboratories.

7.13.5.3 Safety Review of Engineering Change Proposals and Requests for Deviation/Waiver

This element consists of performing and documenting the analyses of engineering change proposals (ECPs) and requests for deviation/waiver to determine the safety impact, if any, upon the system.

7.13.6 Compliance and Verification

7.13.6.1 Safety Verification

Safety Verification is conducted to verify compliance with safety requirements by defining and performing tests and demonstrations or other verification methods on safety critical hardware, software, and procedures.

7.13.6.2 Safety Compliance Assessment

The element consists of performing and documenting a safety compliance assessment to verify compliance with all military, federal, national, and industry codes imposed contractually or by law. This element is intended to ensure the safe design of a system, and to comprehensively evaluate the safety risk that is being assumed prior to any test or operation of a system or at the completion of the contract.

7.13.6.3 Explosive Hazard Classification and Characteristics Data

The purpose of this element is to ensure the availability of tests and procedures need to assign an Explosive Hazard Classification (EHC) to new or modified ammunition, explosives (including solid propellants), and devices containing explosives, and to develop hazard characteristics data for these items.

7.13.6.4 Explosive Ordinance Disposal Source Data

The purpose of this element is to ensure that the following resources are available as needed: source data, explosive ordnance disposal procedures, recommended “render safe” procedures, and test items for new or modified weapons systems, explosive ordnance items, and aircraft systems.

7.13.7 Tailoring Guidelines

A system safety program needs to be matched to the scope and complexity of the development program, i.e., tailored to the program requirements. The requirements of MIL-STD-882 are tailored primarily by the selection of the applicable elements. Tables 7.13-1, and 7.13-2 taken from MIL-STD-882, Appendix A, are element application matrices used to indicate the applicable elements for development programs, and for facilities acquisition programs.

7.14 Finite Element Analysis

7.14.1 Introduction and General Information

Finite element analysis (FEA) is an automated technique for determining the effects of mechanical loads and thermal stress on a structure or device. It is a computer simulation that can predict the material response or behavior of a model of that device or structure represented as a network of simple elements.

FEA is a powerful method for identifying areas of stress concentration that are susceptible to mechanical failure. A device is modeled by decomposing it into a collection of simple shapes, such as plate elements or three dimensional brick elements. The elements are connected together at node points. The analysis can provide material temperatures and stresses at each node point by simulating thermal or dynamic loading situations.

FEA can be used to assess the potential for thermal and mechanical failures before manufacture and testing. It may be used to analyze mechanical systems ranging in size from a portion of a microcircuit chip to a large space antenna. For this reason, FEA is an important numerical analysis technique.

7.14.2 Finite Element Analysis Application

FEA is most appropriately applied to structures too large to test economically, to irregular shaped objects or those composed of many different materials, which do not lend themselves to direct analysis, and to microelectronic devices that may exist only as electronic design representations. In each case, it will reveal areas at risk from mechanical or thermal stress.

A realistic test of a tower, large antenna, etc., cannot be done without going through the expense of constructing the structure. In most cases, this is much too costly, yet it is too risky to commit the design for a large structure to production without assurance of its reliability. FEA can provide the necessary assurance at a relatively insignificant expense. It can also be used when tests are impossible, such as when the structure is intended for use in outer space.

TABLE 7.13-1: APPLICATION MATRIX FOR SYSTEM PROGRAM DEVELOPMENT

				PROGRAM PHASE		
TASK	TITLE	TASK TYPE	CONCEPT	VALID	FSED	PROD
100	System Safety Program	MGT	G	G	G	G
101	System Safety Program Plan	MGT	G	G	G	G
102	Integration/Management of Associate Contractors, Subcontractors, and AE Firms	MGT	S	S	S	S
103	System Safety Program Reviews	MGT	S	S	S	S
104	SSG/SSWG Support	MGT	G	G	G	G
105	Hazard Tracking and Risk Resolution	MGT	S	G	G	G
106	Test and Evaluation Safety	MGT	G	G	G	G
107	System Safety Progress Summary	MGT	G	G	G	G
108	Qualifications of Key System Safety Personnel	MGT	S	S	S	S
201	Preliminary Hazard List	ENG	G	S	S	N/A
202	Preliminary Hazard Analysis	ENG	G	G	G	GC
203	Subsystem Hazard Analysis	ENG	N/A	G	G	GC
204	System Hazard Analysis	ENG	N/A	G	G	GC
205	Operating and Support Hazard Analysis	ENG	S	G	G	GC
206	Occupational Health Hazard Assessment	ENG	G	G	G	GC
207	Safety Verification	ENG	S	G	G	S
208	Training	MGT	N/A	S	S	S
209	Safety Assessment	MGT	S	S	S	S
210	Safety Compliance Assessment	MGT	S	S	S	S
211	Safety Review of ECPs and Waivers	MGT	N/A	G	G	G
212	- RESERVED -	—	—	—	—	—
213	GFE/GFP System Safety Analysis	ENG	S	G	G	G
301	Software Req. Hazard Analysis	ENG	S	G	G	GC
302	Top-Level Design Hazard Analysis	ENG	S	G	G	GC
303	Detailed Design Hazard Analysis	ENG	S	G	G	GC
304	Code-Level Software Hazard Analysis	ENG	S	G	G	GC
305	Software Safety Testing	ENG	S	G	G	GC
306	Software/User Interface Analysis	ENG	S	G	G	GC
307	Software Change Hazard Analysis	ENG	S	G	G	GC

Notes: TASK TYPE
ENG - System Safety Engineering
MGT - Management

APPLICABILITY CODES
S - Selectively Applicable
G - Generally Applicable
N/A Not Applicable

PROGRAM PHASE
CONCEPT - Conceptual
VALID - Validation
FSED - Full Scale Engineering Development
PROD - Production

TABLE 7.13-2: APPLICATION MATRIX FOR FACILITIES ACQUISITION

TASK	TITLE	TASK TYPE	PROGRAM PHASE: CONCEPT	VALID	FSED	PROD
100	System Safety Program	MGT	G	G	G	G
101	System Safety Program Plan	MGT	S	G	G	S
102	Integration/Management of Associate Contractors, Subcontractors, and AE Firms	MGT	S	S	S	S
103	System Safety Program Reviews	MGT	G	G	G	G
104	SSG/SSWG Support	MGT	G	G	G	G
105	Hazard Tracking and Risk Resolution	MGT	G	G	G	G
106	Test and Evaluation Safety	MGT	G	G	G	G
107	System Safety Progress Summary	MGT	S	S	S	S
108	Qualifications of Key System Safety Personnel	MGT	S	S	S	S
201	Preliminary Hazard List	ENG	G	N/A	N/A	N/A
202	Preliminary Hazard Analysis	ENG	G	S	N/A	N/A
203	Subsystem Hazard Analysis	ENG	N/A	S	G	GC
204	System Hazard Analysis	ENG	N/A	G	G	GC
205	Operating and Support Hazard Analysis	ENG	S	G	G	GC
206	Occupational Health Hazard Assessment	ENG	G	S	N/A	N/A
207	Safety Verification	ENG	N/A	S	S	S
208	Training	MGT	S	S	S	S
209	Safety Assessment	MGT	N/A	S	G	S
210	Safety Compliance Assessment	MGT	N/A	S	S	S
211	Safety Review of ECPs and Waivers	MGT	S	S	S	S
212	- RESERVED -	—	—	—	—	—
213	GFE/GFP System Safety Analysis	ENG	S	S	S	S
301	Software Req. Hazard Analysis	ENG	S	S	S	GC
302	Top-Level Design Hazard Analysis	ENG	S	S	S	GC
303	Detailed Design Hazard Analysis	ENG	S	S	S	GC
304	Code-Level Software Hazard Analysis	ENG	S	S	S	GC
305	Software Safety Testing	ENG	S	S	S	GC
306	Software/User Interface Analysis	ENG	S	S	S	GC
307	Software Change Hazard Analysis	ENG	S	S	S	GC

Notes: TASK TYPE
ENG - System Safety Engineering
MGT - Management

APPLICABILITY CODES
S - Selectively Applicable
G - Generally Applicable
N/A Not Applicable

Conventional mathematical analysis of structures become intractable when they are complex or composed of many different materials. These same factors confound the estimation of temperatures within the structure. Judiciously applied, FEA can reduce the risks of using less conservative structural designs. Even smaller designs can benefit from using FEA simulated structures to reduce the need for prototypes and expensive tests.

Mechanical systems have historically been designed with large safety margins. However, many applications preclude this. Airborne structures, for example, must be lightweight, severely limiting the selection and the amount of material available. To accommodate such constraints without courting disaster requires a comprehensive stress analysis. Even when large safety factors are possible, the knowledge provided by FEA permits sound design to be achieved with the minimum amount of materials, thus generating significant cost savings.

The optimum time to detect a structural design flaw is before any construction begins. Changing a design while it is still only a file in a computer is almost trivial. The cost of fixing design errors after prototypes or production models are produced can be significant. The most costly fixes are those required after the system is operational, and the need for these is often revealed by some disaster. FEA provides the means for the early detection of problems in proposed structures, and hence, economical corrective action.

FEA, however, can be time consuming and analysis candidates must be carefully selected. Candidates for FEA include devices, components, or design concepts that: (a) are unproven and for which little or no prior experience or test information is available; (b) use advanced or unique packaging or design concepts; (c) will encounter severe environmental loads; or (d) have critical thermal or mechanical performance and behavior constraints. The most difficult and time consuming portion of an FEA is creating the model. This aspect of FEA is being addressed by the development of intelligent modeling software and automated mesh generators.

FEA can take many different forms, some specific types of FEA include:

(1) Linear Static Analysis - Responses of a linear system to statically applied loads

(2) Linear and Modal Dynamic Analyses - Responses to time-dependent loads

(3) Heat Transfer Analysis - Analyses the flow or transfer of heat within a system

(4) FEAP - Analyzes mechanical stress effects on electronic equipment, printed circuit boards (PCB), avionic equipment, etc.

Many commercial general purpose and special purpose software products for FEA are available.

7.14.3 Finite Element Analysis Procedure

The following is a brief outline of a typical Finite Element Analysis - that of a hypothetical microcircuit/printed circuit board interface application.

First, the entire device (or a symmetrical part of the entire device) is modeled with a coarse mesh of relatively large sized elements such as 3-dimensional brick elements. The loading, material property, heat sink temperature, and structural support data are entered into the data file in the proper format and sequence as required by the FEA solver. The deflections and material stresses for all node point locations, see Figure 7.14-1, on the model are the desired output from the FEA.

Step 1: Perform FEA

(1) Establish FEA mesh

(2) Apply loading and boundary conditions

(3) Perform simulation

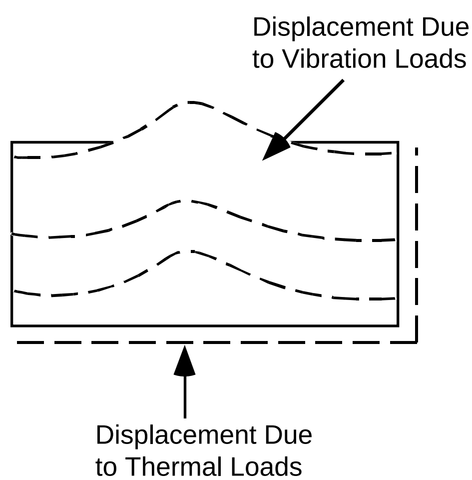
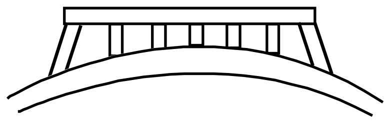


FIGURE 7.14-1: NODAL ANALYSIS

Step 2: Interpretation of Local Displacements/Stresses

For microelectronic devices, second or third follow-on models of refined regions of interest may be required because of the geometrically small feature sizes involved. The boundary nodes for the follow-on model are given initial temperatures and displacements that were acquired from the circuit board model. Figure 7.14-2 shows a refined region containing a single chip carrier and its leads. The more refined models provide accurate temperature, deflection, and stress information for reliability analyses. For example, the results of Step 2 could be a maximum stress value in a corner lead of a chip carrier caused by temperature or vibration cycling.



Vibration and Thermal Displacements
Component Relative to Board

FIGURE 7.14-2: DISPLACEMENT/STRESS INTERPRETATION

Step 3: Perform Life Analysis

A deterministic life analysis is then made by locating the stress value, S_1, on a graph of stress versus cycles-to-failure for the appropriate material, reading cycles to failure, N_1, on the abscissa as shown in Figure 7.14-3. Cycles to failure and time to failure are related by the temperature cycling rate or the natural frequency for thermal or dynamic environments, respectively.

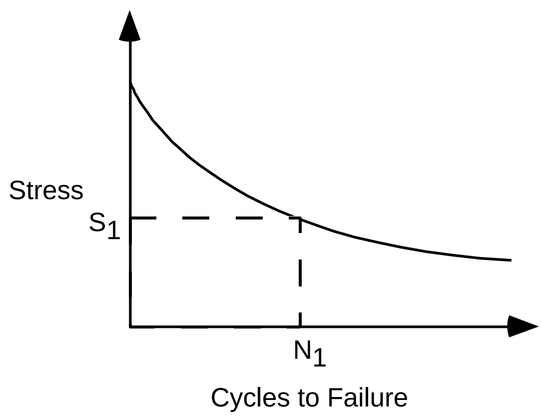


FIGURE 7.14-3: DETERMINISTIC ANALYSIS

Step 4: Estimate Circuit Board Lifetime

A distribution of stress coupled with a distribution of strength (i.e. scatter in fatigue data) will result in a probability distribution function and an estimate of the circuit board lifetime as shown in Figure 7.14-4.

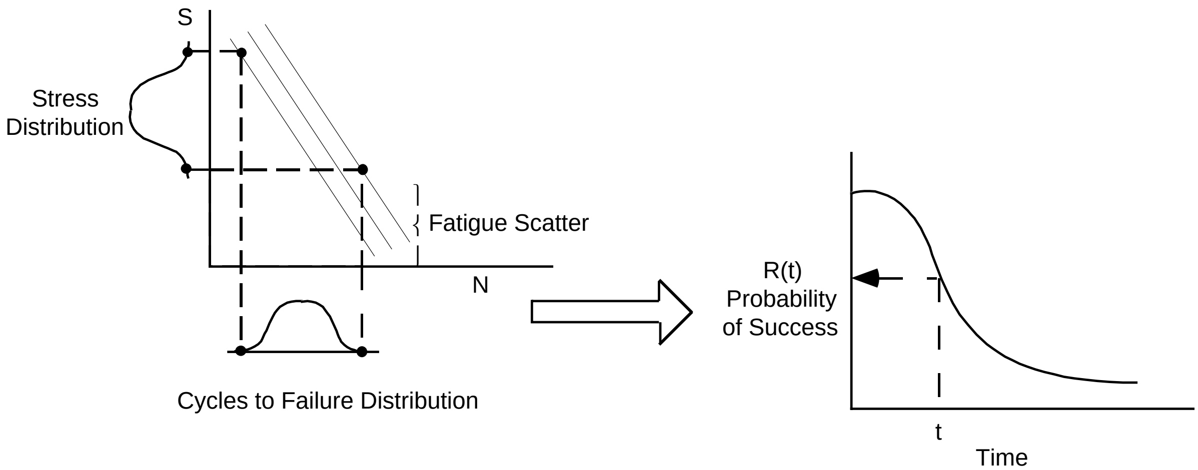


FIGURE 7.14-4: LIFETIME ESTIMATE

7.14.4 Applications

Two examples of how an FEA might be applied are:

a. assess the number of thermal or vibration cycles to failure of an electronic device

b. determine the probability of a fatigue failure at a critical region or location within a device after a given number of operating hours

7.14.5 Limitations

The adequacy of FEA is determined, or limited, by the following factors:

a. Numerical accuracy

b. Model accuracy

c. Material properties

7.15 References for Section 7

1. Parts Selection, Application, and Control (PASC), Reliability Analysis Center, Rome, NY, 1993.

2. Reliable Application of Plastic Encapsulated Microcircuits, (PEM2), Reliability Analysis Center, Rome, NY, 1995.

3. Analog Testing Handbook, (ATH), Reliability Analysis Center, Rome, NY, 1993.

4. 767 AWACS Strategies for Reducing Diminishing Manufacturing Sources (DMS) Risk, DMSMS (Diminishing Manufacturing Sources and Material Shortages) 96 Conference, Montgomery, Texas, 7-9 May 96.

5. Best Practices - How to Avoid Surprises in the World's Most Complicated Technical Process, NAVSO P6071.

6. Precondition of Plastic Surface Mount Devices Prior to Reliability Testing, JESD 22-A113.

7. General Standard for Statistical Process Control, JEDEC Publication 19.

8. JEDEC Registered and Standard Outlines for Semiconductor Devices, JEDEC Publication 95.

9. Impact of Moisture on Plastic IC Package Cracking IPC-SM-786.

10. Test Method, Surface Mount Component Cracking, IPC-TM-650.

11. Buying Commercial and Nondevelopmental Items: A Handbook, SD-2, Office of the Under Secretary of Defense for Acquisition and Technology, April 1996.

12. Lipson, C., et al., Reliability Prediction -- Mechanical Stress/Strength Interference Models, RADC-TR-68-403, March 1967.

13. Lipson, C., et al., Reliability Prediction--Mechanical Stress/Strength Interference (nonferrous), RADC-TR-68-403, December 1968.

14. Nonelectronic Reliability Notebook, RADC-TR-85-194, October 1985.

15. Electronic Engineers' Handbook. Fink, D.G. and D. Christiansen, ed., New York, NY: McGraw Hill Book Co., 1982.

16. Engineering Design Handbook: Reliable Military Electronics. AMCP 706-124, Headquarters U.S. Army Materiel Command, 5001 Eisenhower Ave, Alexandria, VA 22333, AD#A025665.

17. IEEE Recommended Practice on Surge Voltages in Low Voltage AC Power Circuits , IEEE Standard C62.41-1991.

18. Engineering Design Handbook: Design for Reliability, AMCP 706-196, AD#A027230, January 1976.

19. Lewis, E.E., Introduction to Reliability Engineering, John Wiley & Sons, Inc., New York, 1996.

20. Practical Reliability, Vol. 1 - Parameter Variations Analysis, NASA CR-1126, Research Triangle Institute, Research Triangle Park, NC, July 1968.

21. Ross, P., Taguchi Techniques for Quality Engineering, McGraw-Hill, New York, 1988.

22. Klion, J., A Redundancy Notebook, RADC-TR-77-287, December 1977, AD#A050837.

23. Shooman, M., Probabilistic Reliability: An Engineering Approach, New York, NY, McGraw-Hill Book Co., 1968.

24. Barrett, L.S., Reliability Design and Application Considerations for Classical and Current Redundancy Schemes, Lockheed Missiles and Space Co., Inc., Sunnyvale, CA, September 30,1973.

25. Application of Markov Techniques, IEC 1165, 1995.

26. Engineering Design Handbook: Environmental Series, Part One: Basic Environmental Concepts, AMCP 706-115, AD#784999.

27. Engineering Design Handbook: Environmental Series, Part Two: natural Environmental Factors, AMCP 706-116, AD#012648.

28. Engineering Design Handbook: Environmental Series, Part Three: Induced Environmental Factors, AMCP 706-117, AD#023512.

29. Engineering Design Handbook: Environmental Series, Part Four: Life Cycle Environments, AMCP 706-118, AD#0151799.

30. Engineering Design Handbook: Environmental Series, Part Five: Glossary of Environmental Terms, AMCP 706-119.

31. Engineering Design Handbook: Design for Reliability. AMCP 706-196, AD#A027370, January 1976.

32. Arsenault, J.E. and J.A. Roberts, "Reliability and Maintainability of Electronic Systems," Computer Science Press, 9125 Fall River lane, Potomac, MD 20854, 1980.

33. Pavia, R.V., An Investigation into Engine Wear Caused by Dirt, Aeronautical Research Committee Report, ACA-50, July 1950.

34. Engineering Design Handbook: Design Engineer's Nuclear Effects Manual, Vol. I, Munitions and Weapon Systems (U), AMCP 706-335 (SRD).

35. Engineering Design Handbook: Design Engineer's Nuclear Effects Manual, Vol. II, Electronic Systems and Logistical Systems (U), AMCP 706-336 (SRD).

36. Engineering Design Handbook: Design Engineer's Nuclear Effects Manual, Vol. III, Nuclear Environment (U), AMCP 706-337 (SRD).

37. Engineering Design Handbook: Design Engineer's Nuclear Effects Manual, Vol. IV, Nuclear Effects (U), AMCP 706-338 (SRD).

38. Dougherty, E.M. and J.R. Fragola, Human Reliability Analysis, Wiley, 1988.

39. Lee, K.W., F.A. Tillman, and J.J. Higging, "A Literature search of the Human Reliability Component in a Man-Machine System," *IEEE Transactions on Reliability*, Vol. 37, No. 1, 1988 Apr, pp. 24-34.

40. Meister, D., "A Comparative Analysis of Human Reliability Models, Final Report, Contract N00024-71-C-1257," Naval Sea Systems Command, 1971 Nov.

41. Apostolakis, G.E., G. Mancini, R.W. van Otterloo, and F.R. Farmer, eds., "Special Issue on Human Reliability Analysis," *Reliability Engineering & System Safety*, Vol. 29, No. 3, ISBN:0951-8320, 1990.

42. LaSala, K.P., "Survey of Industry Human Performance Reliability Practices," *IEEE Reliability Society Newsletter*, Vol. 36, No. 2, 1990 Apr, pp. 7-8.

43. D.D. Woods, L.J. Johannesen, Richard I Cook, N.B. Sarter, Behind Human Error: Cognitive Systems, Computers, and Hindsight, Crew Systems Ergonomics Information Analysis Center, 1994.

44. Watson, P.A. and W. Hebenstreit, "Manpower, Personnel, and Training Workshop Group Report (IDA Record Document D-35)," Institute of Defense Analysis, 1983.

45. Reason, J., Human Error, Cambridge University Press, 1992.

46. Siegel, A.I., K.P. LaSala, and C. Sontz, Human Reliability Prediction System User's Manual, Naval Sea Systems Command, 1977 Dec.

47. Blanchard, B.S., System Engineering Management, Wiley, 1991.

48. Advisory Group on Reliability of Electronic Equipment (AGREE), "Reliability of Military Electronic Equipment, "Office of the Assistant Secretary of Defense, 1957 4 Jun, pp. 52-57.

49. Blanchard, B.S. and W.J. Fabricky, System Engineering Analysis, Prentice-Hall, 1990.

50. Taha, H., Operations Research: An Introduction, Macmillan, 1971.

51. Bazovski, I., Sr., "Weapon System Operational Readiness," *Proceedings 1975 R&M Symposium*, IEEE, 1975, pp. 174-178.

52. Van Cott, H.P. and R.G. Kinkade, Human Engineering Guide to Equipment Design (2nd Edition), Joint Army-Navy-Air Force Steering Committee, US Government Printing Office, 1972.

53. Boff, K.R., L. Kaufman, and J. Thomas, Handbook of Perception and Human Performance (Vols. 1 and 2), Wiley, 1986.

54. Boff, K. R. and Lincoln, J. E., Engineering Data Compendium: Perception and Performance (Vols. 1-3), Wright-Patterson AFB, OH: Armstrong Aerospace Medical Research Laboratory, 1988.

55. Booher, H. R., Ed, MANPRINT: An Approach to Systems Integration, Van Nostrand Reinhold, 1990.

56. Munger, S.J., R.W. Smith, and D. Paynes, "An Index of Electronic Equipment Operability: Data Store," (Air-C-43-1/62-RP[l]) (DTIC No. AD 607161), Pittsburgh PA, American Institute for Research, 1962 Jan.

57. Topmiller, D.A., J.S. Eckel, and E.J. Kozinsky, "Human Reliability Data Bank for Nuclear Power Plant Operations, Volume l: A Review of Existing Human Error Reliability Data Banks" (NUREG/CR-2744/1 of 2 and SAND82-70571/1 of 2, AN, RX), Dayton OH, General Physics Corp., 1982 Dec .

58. Haney, L.N., H.S. Blackman, B.J. Bell, S.E. Rose, D.J. Hesse, L.A. Minton, and J.P. Jenkins, "Comparison and Application of Quantitative Human Reliability Analysis

Methods for the Risk Methods Integration and Evaluation Program (RMIEP)," NUREG/CR-4835, Idaho National Engineering Laboratory, Idaho Falls, ID, Jan. 1989.

59. Lydell, B.O.Y., "Human Reliability Methodology. A Discussion of the State of the Art," *Reliability Engineering and System Safety*, 36 (1992), pp. 15-21.

60. Dougherty, E.M., "Human Reliability Analysis; Need, Status, Trends, and Limitations," *Reliability Engineering and System Safety*, 29(1990), pp. 283-289.

61. Swain, A.D., "Human Reliability Analysis: Need, Status, Trends and Limitations," *Reliability Engineering and Systems Safety*, 29(1990), pp. 301-313.

62. Swain, A.D., "THERP", SC-R-64-1338, Sandia National Laboratories, Albuquerque, NM, 1964 Aug.

63. Swain, A. D., and Guttman, H.E., Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications (NUREG/CR-1278, SAND800 200, RX, AN), Sandia National Laboratories, Albuquerque, NM, 1983 Aug.

64. Poucet, A., "The European Benchmark Exercise on Human Reliability Analysis," Proceedings of the International Topical Meeting on Probability, Reliability, and Safety Assessment, PSA '89, American Nuclear Society, Inc., La Grange, IL, 1989, pp. 103-110.

65. Guassardo, G., "Comparison of the Results Obtained from the Application of Three Operator Action Models," Proceedings of the International Topical Meeting on Probability, Reliability, and Safety Assessment, PSA '89, American Nuclear Society, Inc., La Grange , IL, 1989, pp. 111-119.

66. Krois., P.A., P.M. Haas, J.J. Manning, and R. Bovell, "Human Factors Review for Severe Accident Sequence Analysis" NUREG/CR-3887, Oak Ridge National Laboratory, Oak Ridge TN, 1985 Nov., p. 26.

67. Hannaman, G.W., A.J. Spurgin, and Y.D. Lukic, "Human Cognitive Reliability for PRA Analysis," NUS-4531, NUS Corp., 1984, Dec.

68. Joksimovich, V., A.J. Spurgin, D.D. Orvis, and P. Moieni, "EPRI Operator Reliability Experiments Program: Model Development/Testing," *Proceedings of the International Topical Meeting on Probability, Reliability, and Safety Assessment, PSA '89*, American Nuclear Society, Inc., La Grange Park IL, 1989 Apr., pp. 120-127.

69. Dhillon, B.S., Human Reliability With Human Factors, Pergamon, 1988.

70. Dhillon, B.S., "Modeling Human Errors in Repairable Systems", Proceedings of the 1989 Reliability and Maintainability Symposium, IEEE, New York, NY, pp. 418-423.

71. Siegel, A.I., and J.J. Wolf, Man-Machine Simulation Models, Wiley, 1969.

72. Siegel, A.I., W.D. Barter, J.J. Wolf, and H.E. Knee, "Maintenance Personnel Performance Simulation (MAPPS) Model: Description of Model Content, Structure, and Sensitivity Testing," NUREG/CR-3626, Oak Ridge National Laboratory, Oak Ridge, TN, 1984 Dec.

73. Woods, D.D., E.M. Roth, and H. Pople, Jr, "Cognitive Environment Simulation: An Artificial Intelligence System for Human Performance Assessment," NUREG/CR-4862, Westinghouse Research and Development Center, Pittsburgh, PA, 1987 Nov.

74. Embrey, D.E., "The Use of Performance Shaping Factors and Quantified Expert Judgement in the Evaluation of Human Reliability: An Initial Appraisal," NUREG/CR-2986, Brookhaven National Laboratory, 1983.

75. Embrey, D.E., P. Humphreys, E.A.Rosa, B. Kirwan, K. Rea, "SLIM-MAUD, An Approach to Assessing Human Error Probabilities Using Structured Expert Judgement," NUREG/CR-3518, U.S. Nuclear Regulatory Commission, 1984.

76. Rosa, E.A., P.C. Humphreys, C.M. Spettell, and D.E. Embrey, "Application of SLIM-MAUD: A Test of an Interactive Computer-based Method for Organizing Expert Assessment of Human Performance and Reliability," NUREG/CR-4016, Brookhaven National Laboratory, Upton, NY, 1985 Sep.

77. Ireson, W.G., and C.F. Coombs, Handbook of Reliability Engineering and Management, New York, McGraw-Hill, 1988, Ch 12.

78. Procedures for Performing a Failure Mode, Effects, and Criticality Analysis, MIL-STD-1629, 28 Nov 1984.

79. Analysis Techniques for System Reliability - Procedure for Failure Mode and Effects Analysis (FMEA) , IEC 812, 1985.

80. Surface Vehicle Recommended Practice: Potential Failure Mode and Effects Analysis in Design (Design FMEA) and Potential Failure Mode and Effects in Manufacturing (Process FMEA) Reference Manual , SAE J-1739, July 1994.

81. Michels, J.M., "Computer Evaluation of the Safety Fault Tree Model," Proceedings System Safety Symposium, 1965, available from University of Washington Library, Seattle, WA.

82. Henley, E.J., and J.W. Lynn (Eds), Generic Techniques in System Reliability Assessment, Nordhoff, Leyden, Holland, 1976.

83. Vesely, W.E., "A Time-Dependent Methodology for Fault Tree Evaluation," Nuclear Engineering and Design, 13, 2 August 1970.

84. Fault Tree Handbook. NUREG-0492, available from Division of Technical Information and Document Control, U.S. Nuclear Regulatory Commission, Washington, DC 20555.

85. Fault Tree Analysis , IEC 1025, 1990.

86. Sneak Circuit Analysis for the Common Man, RADC-TR-89-223.

87. Integration of Sneak Circuit Analysis with Design, RADC-TR-90-109.

88. Automated Sneak Circuit Analysis Technique, Rome Air Development Center, Griffiss Air Force Base, N.Y., 13441, 1990.

89. SCAT: Sneak Circuit Analysis Tool, Version 3.0, RL-TR-95-232.

90. Clardy, R.C., "Sneak Circuit Analysis Development and Application," 1976 Region V, IEEE Conference Digest, 1976, pp. 112-116.

91. Hill, E.J., and L.J. Bose, "Sneak Circuit Analysis of Military Systems," Proceedings of the 2nd International System Safety Conference, July 1975, pp. 351-372.

92. Buratti, D.L., and Goday, S.G., Sneak Analysis Application Guidelines. RADC-TR-82-179, Rome Air Development Center, Griffiss Air Force Base, N.Y., 13441, June 1982.

93. Godoy, S.G., and G.J. Engels, "Sneak Circuit and Software Sneak Analysis," Journal of Aircraft, Vol. 15, August 1978, pp. 509-513.

94. Definition of Terms for Testing Measurement and Diagnostics, MIL-STD-1309, February 1992.

95. Testability Program for Electronic Systems and Equipments, MIL-HDBK-2165, July 1995.

96. Skeberdis, P.W., E.G. White, Fault Logging Using a Micro Time Stress Measurement Device, RL-TR-95-289, Westinghouse Electronics Systems, January 1996.

97. Havey, G., S. Louis, S. Buska, Micro-Time Stress Measurement Device Development, RL-TR-94-196, Honeywell, Inc., November 1994.

98. Environmental Characterization Device Sourcebook (ECDS), Reliability Analysis Center, PO Box 4700, Rome, NY 13342-4700, September 1995.

SECTION 7: RELIABILITY ENGINEERING DESIGN GUIDELINES

99. Testability Design Rating System: Testability Handbook and Analytical Procedure, (2 Vols.), RL-TR-92-12.

100. Testability Program for Systems and Equipment, MIL-HDBK-2165, January 1985.

101. Electromagnetic Properties and Effects of Advanced Composite Materials: Measurement and Modeling, RADC-TR-78-156.

102. Electromagnetic Shielding Effectiveness for Isotropic and Anisotropic Materials, RADC-TR-81-162.

8.0 RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

8.1 Introduction

Successful or satisfactory operation - the goal of all design efforts - yields little information on which to base improvements. Failures, on the other hand, contribute a wealth of data on "what to improve" or "what to design against" in subsequent efforts. The feedback of information obtained from the analysis of failures is one of the principal stepping stones of progress.

The prediction or assessment of reliability is actually an evaluation of unreliability, the rate at which failures occur. The nature and underlying cause of failures must be identified and corrected to improve reliability. Reliability data consist of reports of failures and reports of duration of successful operation of the monitored equipment/system.

Reliability data is used for three main purposes:

(1) To verify that the equipment is meeting its reliability requirements

(2) To discover deficiencies in the equipment to provide the basis for corrective action

(3) To establish failure histories for comparison and for use in prediction

Reliability data can also be useful in providing information about logistics, maintenance, and operations. The data can provide a good estimate of the degradation and wearout characteristics of parts and components and how spare parts requirements are affected.

From this information, not only can effective preventive maintenance routines to control frequent trouble areas be developed, but also an estimate can be obtained of the number of maintenance manhours required to assure a desired level of reliability.

It is important that the data be factual so that a high degree of credence may be placed in the conclusions derived from it. Incomplete and inaccurate reporting will inevitably lead to either complete loss of confidence in the data or to incorrect conclusions and, hence, incorrect decisions and actions based on the conclusions.

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

Reliability/failure data can be obtained from a number of sources.

(1) An in-house failure analysis and corrective action system (FRACAS)

(2) Reliability test data

(3) Subcontractor or vendor data

(4) Field data

(5) Reliability data banks

The most useful of the above sources are (1) and (2), and possibly (5). The other sources are not as reliable since they are, in most cases, incomplete. For example, the military maintenance collection systems for collecting field data (e.g., the Army's TAMMS, the Navy's 3M, and the Air Force's REMIS and other maintenance data collection systems) are primarily maintenance oriented (see Section 11). Thus, field reliability cannot be assessed by using data from these systems alone. All of the factors influencing the data need to be clearly understood. These factors include the ground rules for collecting the data, assumptions made during analysis, and so forth. Clearly understanding these factors assures that the data will be properly interpreted and that conclusions will be credible.

The following section provides more details on a FRACAS system. The sections on Reliability Testing and Growth discuss the collection and analysis of reliability test data.

8.2 Failure Reporting, Analysis, and Corrective Action System (FRACAS) and Failure Review Board (FRB)

8.2.1 Failure Reporting, Analysis and Corrective Action System (FRACAS)

The purpose of FRACAS is to collect failure data, provide procedures to determine failure cause, and document corrective action taken. It requires the contractor to have a system that collects, analyzes and records failures that occur for specified levels of assembly prior to acceptance of the hardware by the procuring activity.

Failure reporting and analysis is necessary to ensure that a product's reliability and maintainability will be achieved and sustained. The FRACAS program is a key element in "failure recurrence" control for newly developed and production equipment. A FRACAS program must include provisions to ensure that failures are accurately reported and thoroughly analyzed and that corrective actions are taken on a timely basis to reduce or prevent recurrence.

An in-plant FRACAS determines the basic causes of failures associated with design or manufacturing, and provides a closed-loop method of implementing corrective action. The system should emphasize the investigation and analysis of all failures, regardless of their apparent frequency or impact, and classification of failures according to categories of design/part procurement, manufacture, or assembly and inspection. It is well known that the most economical repair of a failure occurs at the component part level. A conventional rule of thumb is that a repair action at the subassembly level costs an order of magnitude more than at the part level, and a repair at the product level costs an order of magnitude more than a repair at the subassembly level.

Data on electronic equipment malfunctions can be obtained from any or all of the following types of data sources:

(1) Design verification tests

(2) Pre-production tests

(3) Production tests

(4) Subcontractor tests

(5) Field data

The FRACAS system must provide essential information on:

(1) What failed

(2) How it failed

(3) Why it failed

(4) How future failures can be eliminated

8.2.1.1 Closed Loop Failure Reporting/Corrective Actions System

Figure 8.2-1 indicates the main steps in a closed-loop FRACAS. As shown in Figure 8.2-1, a typical FRACAS consists of fourteen steps.

(1) A failure is observed during some operation or test.

(2) The observed failure is fully documented, including, as a minimum
  (a) Location of failure
  (b) Date and time of failure

(c) Part number of the failed system/equipment
(d) Serial number of the failed system/equipment
(e) Model number of the failed system/equipment
(f) Observed failure symptoms
(g) Name of the individual who observed the failure
(h) All significant conditions which existed at the time of the observed failure

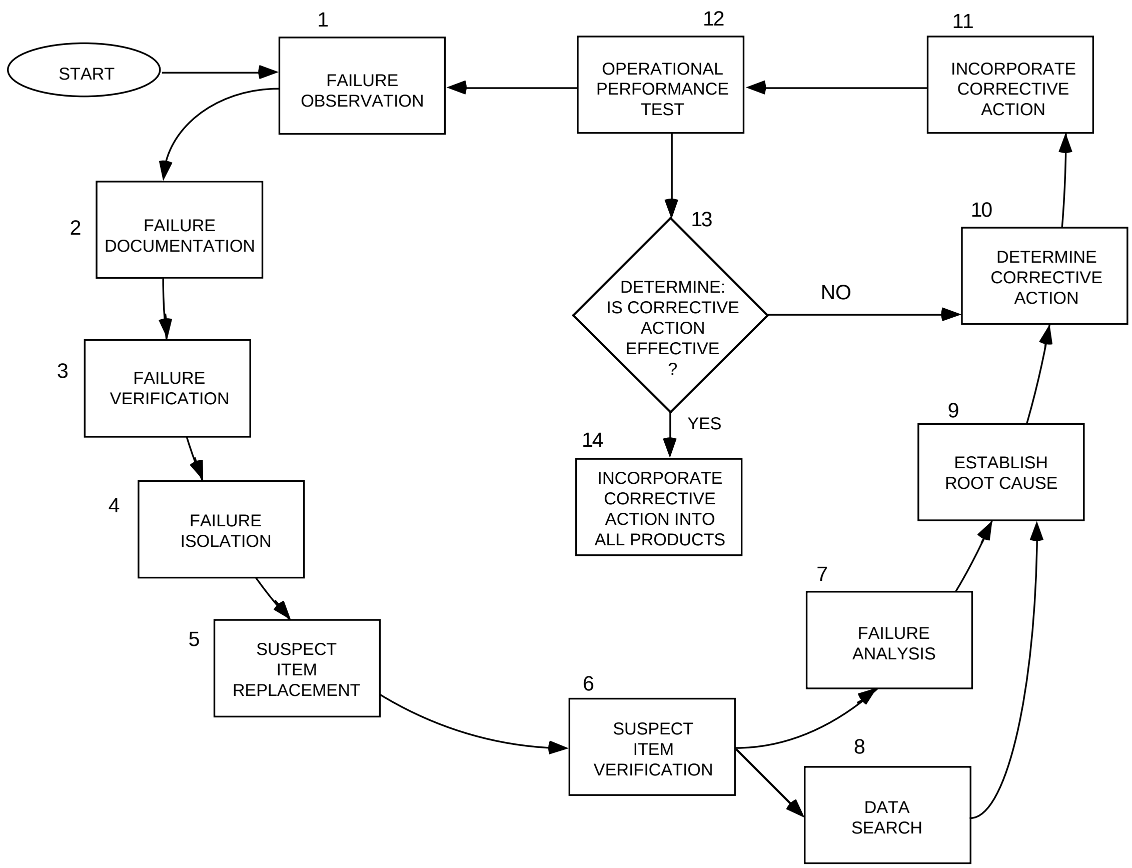


FIGURE 8.2-1: CLOSED LOOP FAILURE REPORTING AND CORRECTIVE ACTION SYSTEM

(3) Failure verification (i.e., reconfirmation of the validity of the initial failure observation).

(4) Failure isolation (i.e., localization of the failure to the lowest replaceable defective item within the system/equipment).

(5) Replacement of the suspected defective item with a known good item and retest of the system/equipment to provide assurance that the replacement item does in fact correct the originally reported failure.

(6) Retest of the suspect item at the system/equipment level or at a lower level to verify that the suspect item is defective.

(7) Failure analysis of the defective item to establish the internal failure mechanism responsible for the observed failure or failure mode.

(8) A search of existing data to uncover similar failure occurrences in this or related items (i.e., establishing the historical perspective of the observed failure mode/failure mechanism).

(9) Utilizing the data derived from Steps 7 and 8, determine the antecedent or root cause of the observed failure.

(10) Determine the necessary corrective action, design change, process change, procedure change, etc. to prevent future failure recurrence. The decision regarding the appropriate corrective action should be made by an interdisciplinary design team.

(11) Incorporation of the recommended corrective action into the original test system/equipment.

(12) Retest of the system/equipment with the proposed corrective action modification incorporated.

(13) After suitable retest and review of all applicable data, determine if proposed corrective action is effective.

(14) After the effectiveness of the proposed corrective action has been proven, the corrective action is then incorporated into the deliverable systems/equipment.

There are several “keys” that make the failure reporting and corrective action cycle effective. These are:

(1) The discipline of the report writing itself must be maintained so that an accurate description of failure occurrence and proper identification of the failed items are ensured.

(2) The proper assignment of priority and the decision for failure analysis must be made with the aid of cognizant design engineers and systems engineers.

(3) The status of all failure analyses must be known. It is of prime importance that failure analyses be expedited as priority demands and that corrective action be implemented as soon as possible.

(4) The root cause of every failure must be understood. Without this understanding, no logically derived corrective actions can follow.

(5) There must be a means of tabulating failure information for determining failure trends and the mean times between failures of system elements. There should also be a means for management visibility into the status of failure report dispositions and corrective actions.

(6) The system must provide for high level technical management concurrence in the results of failure analysis, the soundness of corrective action, and the completion of formal actions in the correction and recurrence prevention loop.

(7) An extremely valuable assurance mechanism is to have active Government involvement in surveillance of the adequacy of the failure reporting, analysis, and corrective action effort.

The contractor's program plan should clearly describe his proposed FRACAS. Furthermore it should identify those provisions incorporated therein to ensure that effective corrective actions are taken on a timely basis. The applicable statement of work (SOW) should identify the extent to which the contractor's FRACAS must be compatible with the procuring agency's data system. It should also identify the levels of assembly and test to be addressed by the FRACAS, give definitions for each of the failure cause categories, identify the applicable logistics support requirements and identify the data items required for delivery.

8.2.1.2 Failure Reporting Systems

Normally a manufacturer's reliability engineering organization is responsible for instituting and managing FRACAS. They establish policy, provide direction, and monitor the status of FRACAS investigations. The cognizant inspection and testing organizations, including reliability and quality engineering, are responsible for initiating failure reports promptly as they are observed. The project management office generally reviews recommendations, coordinates analyses and test activities with the government, authorizes the implementation of acceptable fixes or corrective measures and provides direction relative to continuation of tests. Often, it is the quality assurance organization that transmits reports to the government and coordinates implementation of corrective actions.

8.2.1.3 Failure Reporting Forms

It is imperative that failure reporting and resultant corrective actions be documented. Therefore, failure reporting and corrective actions forms must be designed to meet the needs of the individual system development and production program as well as the organizational responsibilities, requirements, and constraints of the manufacturer. Figure 8.2-2 is an example of a typical failure report form used in a FRACAS system.

8.2.1.4 Data Collection and Retention

Maintaining accurate and up-to-date records through the implementation of the data reporting, analysis and corrective action system described in the preceding subsections provides a dynamic, expanding experience base. This experience base, consisting of test failures and corrective actions, is not only useful in tracking current programs but can also be applied to the development of subsequent hardware development programs. Furthermore, the experience data can be used to:

(1) Assess and track reliability

(2) Perform comparative analysis and assessments

(3) Determine the effectiveness of quality and reliability activities

(4) Identify critical components and problem areas

(5) Compute historical part failure rates for new design reliability prediction (in lieu of using generic failure rates found in MIL-HDBK-217, for example)

TRAC FAILURE AND DEFECT REPORT

DOC. NO. ☐ SECONDARY DOCUMENT ☐ CONTINUED DOCUMENT	TRAC FAILURE AND DEFECT REPORT				DATE
	MFG. JOB ITEM	PT. NO. (FAILED ITEM)	NOMENCLATURE	SER .NO. OR DATE CODE	CONTINUED TO DOC. NO.

IF EITHER, ENTER PRIMARY DOC. NO. HERE	SHOP ORDER	SUBMISSION NO.	TEST NO.	TEST COMPLETED	FATP	LOT SIZE	SAMPLE SIZE	# ACCEPTED	# REJECTED	TEST UNIT	TESTER	COGNIZANT FOREMAN

DEF. #	CHAR. #	DEFECT CODE	QTY.	DESCRIPTION OF DEFECT	WHEN MORE THAN ONE SN OR DATE CODE, LIST EACH HERE	DISP. #	AUTHORIZATION

TROUBLESHOOTING/REPAIR — PART VERIFICATION

DEF. #	REPAIR CODE	DESCRIPTION	FAILURE CAUSE	CIRCUIT DESIGNATOR	PART NAME	REPLACED ITEM PART NO.	SUB-ASSY S/N OR PT. DATE CODE	CONNECTED BY	INSP.	VERIFIED GOOD	VERIFICATION FINDINGS	SECONDARY DOC #

QUALITY & COMPONENT REVIEW

DEF #	DISP. #	COMMENT	CONNECTED BY	INSP.	CLEARED BY
Q.E.	DATE	P&E/ET/PROGRN.	ENGR		DATE

FAILURE ANALYSIS REPORT

DEF #	MANUFACTURER	ANALYST	DATE	FAILURE ANALYSIS REPORT NO.
				FAILURE CAUSE CODE

THIS DOCUMENT STAYS WITH ORIGINAL FAILED ITEM UNTIL CLEARED. FOREMAN SHOULD MAKE COPIES FOR EACH SEPARATED COMPONENT PART. FOREMAN SHOULD ENTER APPROPRIATE DEFECT NUMBER ON COPIES IN QUALITY AND COMPONENT REVIEW SECTION. AT COMPLETION OF INITIAL DISPOSITION FOR SEPARATED COMPONENT PARTS A SECONDARY DOCUMENT SHOULD BE FILLED OUT FOR EACH SEPARATE ASSEMBLY.

FIGURE 8.2-2: EXAMPLE OF FAILURE REPORT FORM

8.2.2 Failure Review Board

For the acquisition of certain critical (extremely expensive and complex) systems and equipments, a separate Failure Review Board (FRB) may sometimes be established specifically to oversee the effective functioning of the FRACAS. The Failure Review Board activity is reliability management. A closed loop FRACAS with an FRB is illustrated in Figure 8.2-3.

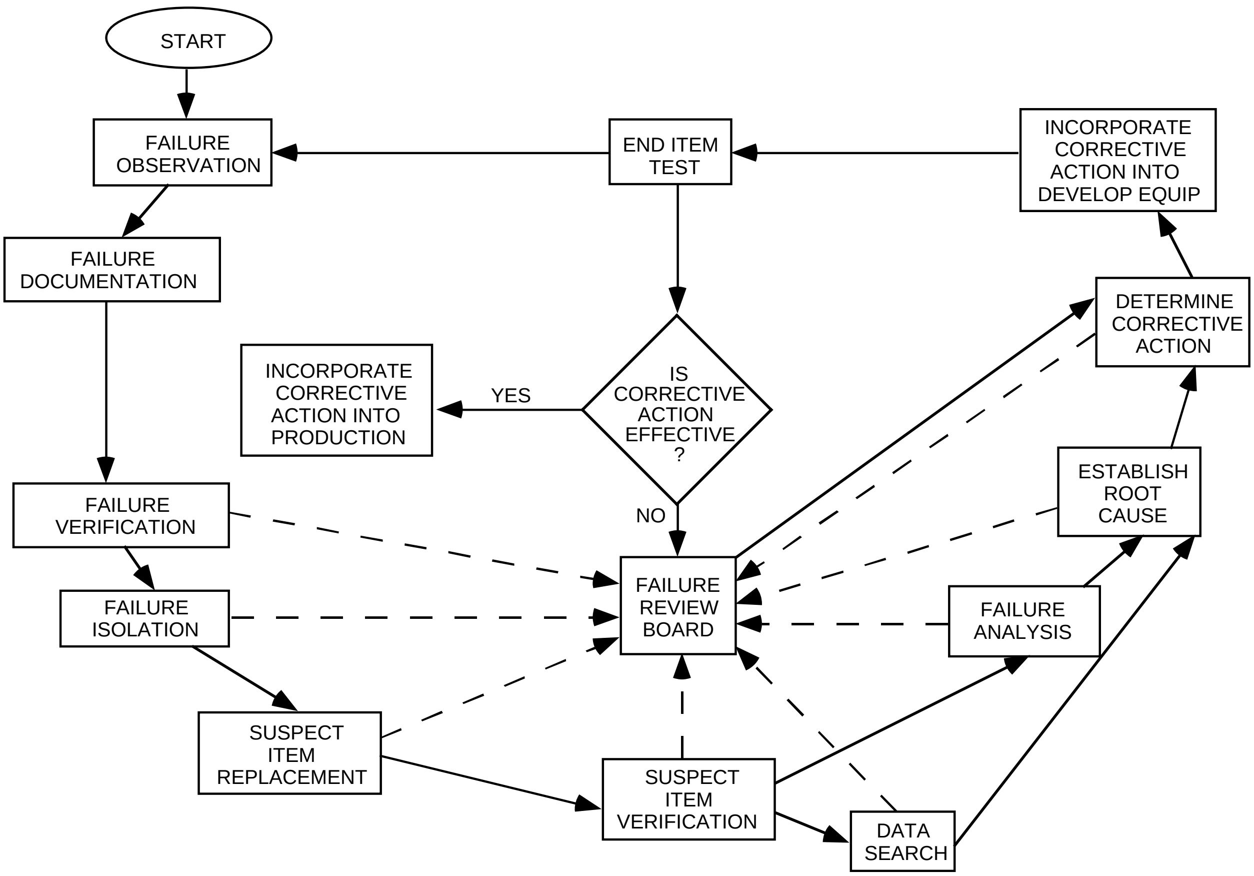


FIGURE 8.2-3: CLOSED LOOP FAILURE REPORTING AND CORRECTIVE ACTION SYSTEM WITH FAILURE REVIEW BOARD

The purpose of the Failure Review Board is to provide increased management visibility and control of the FRACAS. Its intent is to improve reliability and maintainability of hardware and associated software by the timely and disciplined utilization of failure and maintenance data. The FRB consists of a group of representatives from appropriate organizations with sufficient level of responsibility to ensure that failure causes are identified with enough detail to generate and implement effective corrective actions which are intended to prevent failure recurrence and to simplify or reduce the maintenance tasks.

The FRB usually consists of higher level management personnel who possess the authority to set priorities, establish schedules, assign specific responsibility and authorize adequate funding to insure the implementation of any necessary changes when dealing with complex and difficult problems. The acquiring activity usually reserves the right to appoint a representative to the FRB as an observer.

8.3 Reliability Data Analysis

From a reliability assessment viewpoint, failure data are used to:

(1) Determine the underlying probability distribution of time to failure and estimate its parameters (if not already known)

(2) Determine a point estimate of a specific reliability parameter, e.g., MTBF

(3) Determine a confidence interval that is believed to contain the true value of the parameter

Two methods are used to analyze failure data:

(1) Graphical methods

(2) Statistical analysis

In many practical cases, graphical methods are simple to apply and produce adequate results for estimating the underlying distribution. They are virtually always a useful preliminary to more detailed statistical analysis. The two methods will be discussed in more detail in the following subsections.

8.3.1 Graphical Methods

The basic idea of graphical methods is the use of special probability plotting papers in which the cumulative distribution function (cdf) or the cumulative hazard function can be plotted as a straight line for the particular distribution being studied. Since a straight line has two parameters (slope and intercept), two parameters of the distribution can be determined. Thus, reliability data can be evaluated quickly, without a detailed knowledge of the statistical mathematics being necessary. This facilitates analysis and presentation of data.

Graphical curve-fitting techniques and special probability-plotting papers have been developed for all of the distributions commonly associated with reliability analysis (Refs. [4], [5]).

Ranking of Data

Probability graph papers are based upon plots of the variable of interest against the cumulative percentage probability. The data, therefore, need to be ordered, and the cumulative probability calculated. For reliability work, the data are ordered from the smallest to largest; this is referred to as order statistics. For example, consider the data on times-to-failure of 20 items (Table 8.3-1). For the first failure, the cumulative percent failed is 1/20 or 5%. For the second, the cumulative percent failed is 2/20 or 10%, and so on to 20/20 or 100% for the 20th failure. However, for probability plotting, it is better to make an adjustment to allow for the fact that each failure represents a point on a distribution. Thus, considering that the whole population of 20 items represents a sample, the times by which 5, 10, ..., 100% will have failed in several samples of 20 will be randomly distributed. However, the data in Table 8.3.1-1 show a bias, in that the first failure is shown much further from the zero cumulative percentage point than is the last from 100% (in fact, it coincides). To overcome this, and thus to improve the accuracy of the estimation, mean or median ranking of cumulative percentages is used for probability plotting. Mean ranking is used for symmetrical distributions, e.g., normal; median ranking is used for skewed distributions, e.g., Weibull.

The usual method for mean ranking is to use (n + 1) in the denominator, instead of n, when calculating the cumulative percentage position. Thus in Table 8.3-1 the cumulative percentages (mean ranks) would be:

$$\frac{1}{20 + 1} = .048 \cong 5\%$$

$$\frac{2}{20 + 1} = .096 \cong 10\%$$

$$\vdots$$

$$\frac{20}{20 + 1} = .952 \cong 95\%$$

TABLE 8.3-1: DATA ON TIMES TO FAILURE OF 20 ITEMS

Order No.	Time to Failure (hours)	Cumulative % (Cdf)	Mean Rank (%) (Cdf)
1	175	5	5
2	695	10	10
3	872	15	14
4	1250	20	19
5	1291	25	24
6	1402	30	29
7	1404	35	33
8	1713	40	38
9	1741	45	43
10	1893	50	48
11	2025	55	52
12	2115	60	57
13	2172	65	62
14	2418	70	67
15	2583	75	71
16	2725	80	76
17	2844	85	81
18	2980	90	86
19	3268	95	90
20	3538	100	95

These data are shown plotted on normal probability paper in Figure 8.3-1 (circles). The plotted points show a reasonably close fit to the straight line drawn by eye.' Therefore, we can say that the data appear to fit the cumulative normal distribution represented by the line.

Median ranking, as was previously stated, is used for skewed distributions such as the Weibull because it provides a better correction. The most common approximation for median ranking (Ref. [4]) is given by:

$$\text{Median rank } (n,i) = r_i = \frac{i - 0.3}{n + 0.4}$$

where r_i is the i^{th} order value and n is the sample size. Median ranking is the method most used in probability plotting, particularly if the data are known not to be normally distributed. Also, to save calculations, tables of median ranks are available for use. These are included in Table 8.3-2 and will be used in the examples to be described later.

8.3.1.1 Examples of Graphical Methods

Reference [5] provides an excellent discussion of caveats that must be considered in graphical estimation. Now, let us turn to some examples.

Example 1: Normal Distribution

1. When to Use

This method estimates μ and σ, the mean and standard deviation when failure times are normally distributed. This method yields a less accurate estimate than statistical analysis but requires very minimal calculations.

2. Conditions for Use

   a. Failure times must be collected, but may be censored; censored data is discussed in the next section.

   b. Normal probability paper is required.

3. Method	Example
a. On normal probability paper plot the i^{th} failure time in a sample of n ordered failure times on the lower axis vs. $\frac{i}{n+1}$ on the right hand axis.	a. The sample data used in Table 8.3-1 are repeated here, with the necessary plotting positions (mean ranks).

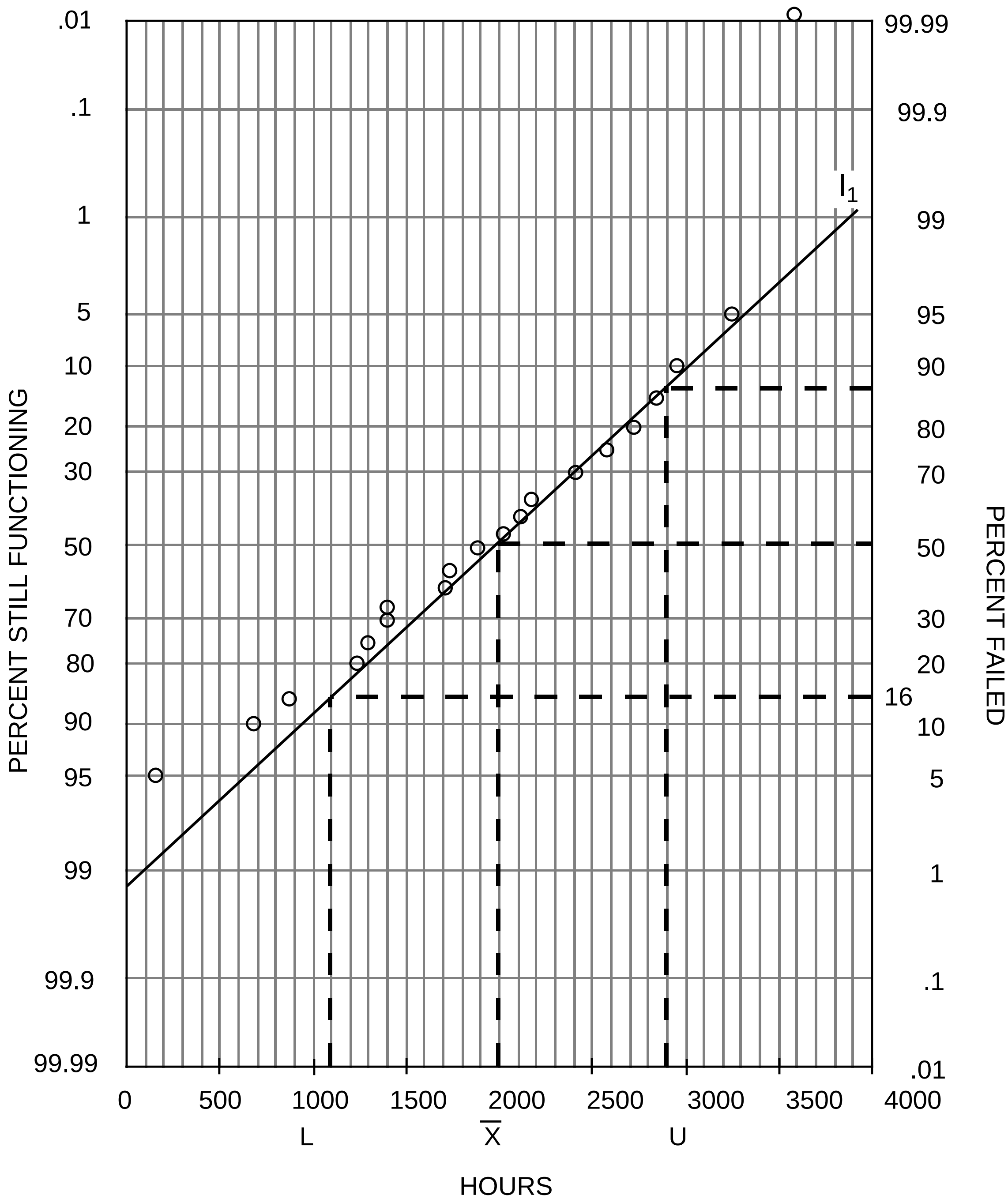


FIGURE 8.3-1: GRAPHICAL POINT ESTIMATION FOR
THE NORMAL DISTRIBUTION

TABLE 8.3-2: MEDIAN RANKS

sample size = n
failure rank = i

n

i	1	2	3	4	5	6	7	8	9	10
1	.5000	.2929	.2063	.1591	.1294	.1091	.0943	.0830	.0741	.0670
2		.7071	.5000	.3864	.3147	.2655	.2295	.2021	.1806	.1632
3			.7937	.6136	.5000	.4218	.3648	.3213	.2871	.2594
4				.8409	.6853	.5782	.5000	.4404	.3935	.3557
5					.8706	.7345	.6352	.5596	.5000	.4519
6						.8906	.7705	.6787	.6065	.5481
7							.9057	.7979	.7129	.6443
8								.9170	.8194	.7406
9									.9259	.8368
10										.9330

n

i	11	12	13	14	15	16	17	18	19	20
1	.0611	.0561	.0519	.0483	.0452	.0424	.0400	.0378	.0358	.0341
2	.1489	.1368	.1266	.1788	.1101	.1034	.0975	.0922	.0874	.0831
3	.2366	.2175	.2013	.1873	.1751	.1644	.1550	.1465	.1390	.1322
4	.3244	.2982	.2760	.2568	.2401	.2254	.2125	.20099	.1905	.1812
5	.4122	.3789	.3506	.3263	.3051	.2865	.2700	.2553	.2421	.2302
6	.5000	.4596	.4253	.3958	.3700	.3475	.3275	.3097	.2937	.2793
7	.5878	.5404	.5000	.4653	.4350	.4085	.3850	.3641	.3453	.3283
8	.6756	.6211	.5747	.5347	.5000	.4695	.4425	.4184	.3968	.3774
9	.7634	.7018	.6494	.6042	.5650	.5305	.5000	.4728	.4484	.4264
10	.8511	.7825	.7240	.6737	.6300	.5915	.5575	.5272	.5000	.4755
11	.8389	.8632	.7987	.7432	.6949	.6525	.6150	.5816	.5516	.5245
12		.9439	.8734	.8127	.7599	.7135	.6725	.6359	.6032	.5736
13			.9481	.8822	.8249	.7746	.7300	.6903	.6547	.6226
14				.9517	.8899	.8356	.7875	.7447	.7063	.6717
15					.9548	.8966	.8450	.7991	.7579	.7207
16						.9576	.9025	.8535	.8095	.7698
17							.9600	.9078	.8610	.8188
18								.9622	.9126	.8678
19									.9642	.9169
20										.9659

3. Method (continued) | Example

Failure Time (Hours)	Plotting Position $\frac{i}{n+1}$
175	.05
695	.10
872	.14
1250	.19
1291	.24
1402	.29
1404	.33
1713	.38
1741	43
1893	.48
2025	.52
2115	.57
2172	.62
2418	.67
2583	.71
2725	.76
2844	.81
2980	.86
3268	.90
3538	.95

b. Draw the line of best fit through the plotted points by using the last point plotted as a reference point for a straight edge and dividing the rest of the points into two equal groups above and below the line.

b. Figure 8.3-1 is the plot of this data on normal paper. The normal line has been labeled ℓ_1.

c. The mean, μ, is estimated by projecting the 50% probability of failure point on the right hand axis to the line and then projecting that intersection point down to the lower axis. The estimate of μ, $\bar{x}$, is read there.

c. The value of $\overline{X}$ is read as 2000 hours.

3. Method	Example
d. The estimate of σ is obtained by first projecting the intersection of the 84% probability of failure point on the right hand axis with the normal line to the lower axis. Call that point on the lower axis U.	d. U = 2900 hours
e. Repeat step d with the 16% point. Call the point L	e. L = 1100 hours
f. The estimate of σ is $s = \frac{U - L}{2}$	f. The sample standard deviation is $s = \frac{U - L}{2} = \frac{2900 - 1100}{2} = 900$ hours
g. The 95% confidence limits around the mean are given by $\bar{X} \pm t\, s/\sqrt{n}$ where t is shown below for various sample sizes, n.	g. $2000 \pm (2.09)\,(900)/\sqrt{20}$ 2000 ± 420 hours

n	t
5	2.57
10	2.23
20	2.09
30	2.04
50	2.00
∞	1.96

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

Example 2: Weibull Distribution

1. When to Use. The flexibility of the Weibull distribution makes it useful in describing the probability density function for a variety of cases. The Weibull cumulative distribution function is given by:

$$F(t) = 1 - e^{\left[-(t/\theta)^{\beta}\right]}, \quad 0 \le t \le \infty$$

The Weibull distribution is used to describe the distribution of times to failure and of strengths of brittle materials, and the weakest link phenomena. It is an appropriate failure law model whenever the system consists of a number of components, and failure is essentially due to the "most severe" fault among a large number of faults in the system. By making an appropriate choice of the shape parameter, β, either an increasing or a decreasing failure rate can be obtained. Estimates of the Weibull shape (β) and scale (θ) parameters may be obtained graphically using *ln-ln* (or a special Weibull probability) graph paper. Less accurate than statistical methods, this method can be done quickly and easily.

2. Steps in Using the Graphical Method

   a. Collect failure times for items under test, put in ascending order, and assign an order number to each. The failure times are the values to be plotted on the x-axis. Note that failure time may be in hours, cycles, or whatever measure of life is appropriate for the item in question.

   b. Assign the median rank for each order number. The median ranks are the values to be plotted on the y-axis. The median rank is one model used for the cumulative probability of failures, F(t). It is usable when the number of failures is greater than 20. The formula is:

$$\text{Median Rank } (n,i) = \frac{i - 0.3}{n + 0.4}$$

   where: n = number of failures
   i = order number

   c. Plot the pairings of median ranks and failure times on Weibull probability graph paper. Draw a straight line that best fits the data (i.e., roughly an equal number of data points will be on either side of the line).

d. The slope of the line is β. The slope is calculated using the following equation:

$$\beta = \frac{\Delta Y}{\Delta X} = \frac{ln\,ln\left(\frac{1}{1-F(t_2)}\right) - ln\,ln\left(\frac{1}{1-F(t_1)}\right)}{ln\,t_2 - ln\,t_1}$$

Note 1: This equation assumes that *ln* paper is used. Log-log paper can also be used with the following equation:

$$\beta = \frac{log\;ln\left(\frac{-1}{1-F(t_2)}\right) - \log ln\left(\frac{-1}{1-F(t_1)}\right)}{log\;t_2 - log\;t_1}$$

Note 2: Some special Weibull graph paper allows β to be read directly.

3. Example

The following failure data are collected from a test in which 20 items were tested to failure.

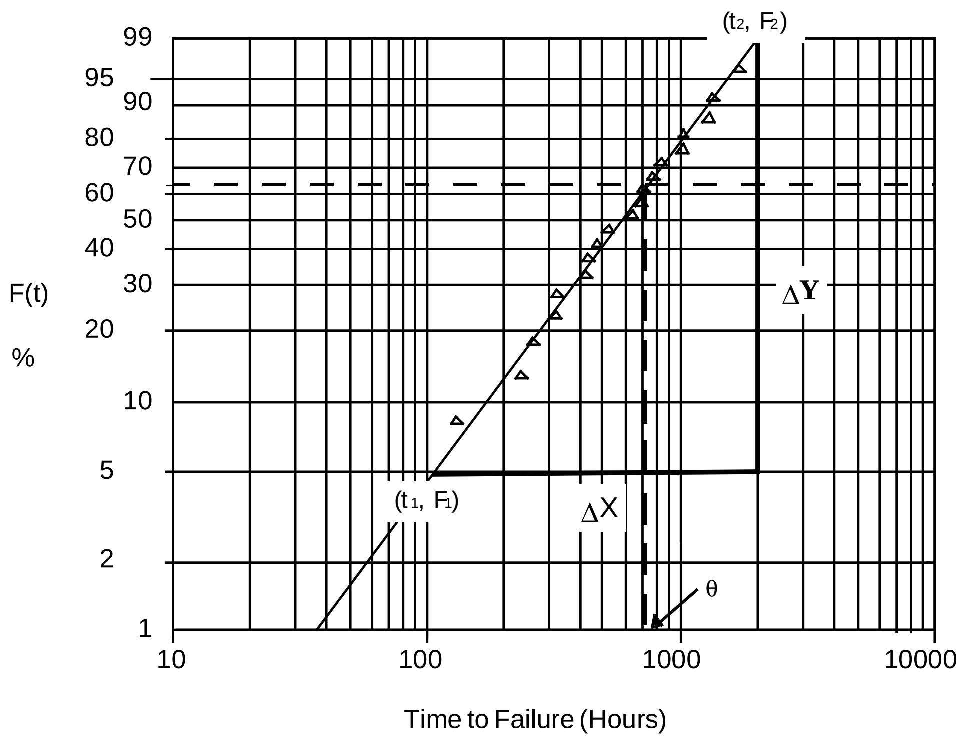
Order Number	Failure Time (in hours)	Median Rank (%)
1	92	3.41
2	130	8.31
3	233	13.22
4	260	18.12
5	320	23.02
6	325	27.93
7	420	32.83
8	430	37.74
9	465	42.64
10	518	47.55
11	640	52.45
12	700	57.36
13	710	62.26
14	770	67.17
15	830	72.07
16	1010	76.98
17	1020	81.88
18	1280	86.78
19	1330	91.69
20	1690	96.59

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

Figure 8.3-2 shows the data plotted on *ln-ln* graph paper. From the graph, θ is 739.41 hours. β is:

$$\beta = \frac{\Delta Y}{\Delta X} = \frac{ln ln\left(\frac{1}{1-.99}\right) - ln ln\left(\frac{1}{1-.05}\right)}{ln\,2000 - ln\,105} = 1.53$$

The reliability at t = 1000 hours is found by drawing a line up vertically from t=1000 on the abscissa to the line. Then, from that point a horizontal line is drawn to the ordinate. It intersects the ordinate at F(t) = 80%. The reliability is 1- F(t) = 20% (i.e., 20% percent probability of no failure).



θ = 739.41

$$\beta = \frac{\Delta Y}{\Delta X} = \frac{ln ln\left(\frac{1}{1-F(t_2)}\right) - ln ln\left(\frac{1}{1-F(t_1)}\right)}{ln\,t_2 - ln\,t_1} = 1.53$$

FIGURE 8.3-2: GRAPHICAL POINT ESTIMATION FOR THE WEIBULL DISTRIBUTION

Example 3: Exponential Distribution

A simple graphical procedure to test the validity of the exponential distribution is to plot the cumulative test or operating time against the cumulative number of failures as shown in Figure 8.3-3. If the plot is reasonably close to a straight line, then a constant failure rate is indicated. An exponential distribution of failures may be assumed.

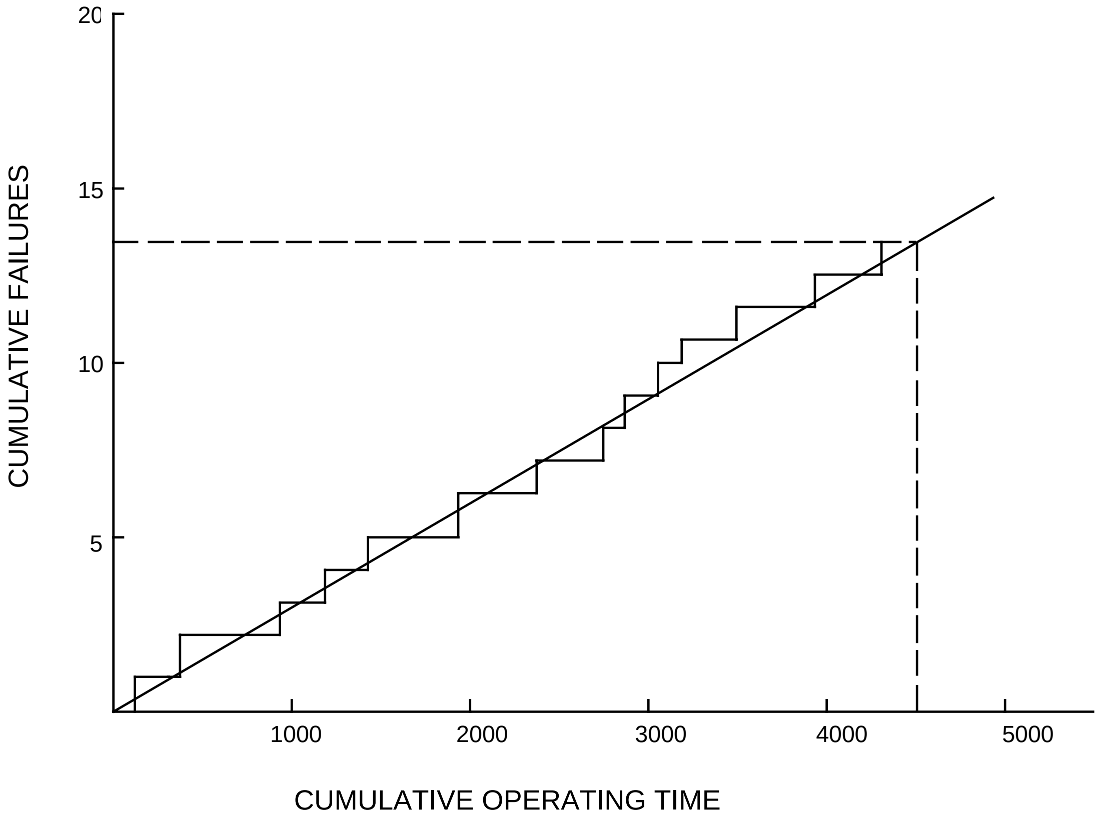


FIGURE 8.3-3: DISTRIBUTION GRAPHICAL EVALUATION

8.3.2 Statistical Analysis

8.3.2.1 Introduction

Since the available data usually only constitute a sample from the total population, statistical methods are used to estimate the reliability parameters of interest, e.g., MTBF, failure rate, probability of survival, etc.

The main advantage of statistics is that it can provide a measure of the uncertainty involved in a numerical analysis. The secondary advantage is that it does provide methods for estimating effects that might otherwise be lost in the random variations in the data.

It is important to keep in mind the fact that data constitute a sample from the total population, that random sampling peculiarities must be smoothed out, that population density parameters must be estimated, that the estimation errors must themselves be estimated, and - what is even more difficult - that the very nature of the population density must be estimated. To achieve these ends, it is necessary to learn as much as one can about the possible population density functions, and especially what kind of results we can expect when samples are drawn, the data are studied, and we attempt to go from data backward to the population itself. It is also important to know what types of population densities are produced from any given set of engineering conditions. This implies the necessity for developing probability models, or going from a set of assumed engineering characteristics to a population density.

It is customary, even necessary, in statistical analysis to develop, from physical engineering principles, the nature of the underlying distribution. The sample of data is then compared against the assumed distribution.

The usual parameter of interest in reliability is the distribution of times to failure, called the probability density function or failure density function. The failure density function may be discrete, that is, only certain (integer) values may occur, as in tests of an explosive squib. Success or failure will occur on any trial, time not being considered. Or it may be continuous, any value of time to failure being possible.

Typically histograms are plotted (e.g., time-to-failure plots) and statistical techniques used to first test the data to determine the applicable form of the probability distribution, and then identify and evaluate the relationship between the reliability parameter(s), such as failure rate, and the critical hardware characteristics/attributes which affect reliability (such as technology, complexity, application factors, etc.) as defined by the data.

8.3.2.2 Treatment of Failure Data

Failure data are usually obtained from a) test results or b) field failure reports. Experience has shown that a good way to present these data is to compute and plot either the failure density function, f(t), or the hazard rate, h(t), as a function of time.

Remember from Section 5 that f(t) is given by the ratio of the number of failures occurring in the time interval to the size of the original population, divided by the length of the time interval. The hazard rate, h(t), on the other hand, is given by the ratio of the number of failures occurring in the time interval to the number of survivors at the beginning of the time interval, divided by the length of the time interval.

Although f(t) and h(t) are defined as continuous functions, piecewise continuous functions of f(t) and h(t) are computed, graphed results are examined, and a continuous model is chosen which best fits the data.

Once having found f(t) from the data, F(t) (the cumulative distribution of time to failure) and R(t) = 1 - F(t), the reliability function or survival probability, can be readily determined from the relationships.

$$F(t) = \int_{-\infty}^{t} f(t)\,dt \tag{8.1}$$

$$R(t) = 1 - F(t) \tag{8.2}$$

Two examples follow.

Example 4:

TABLE 8.3-3: FAILURE DATA FOR TEN HYPOTHETICAL ELECTRONIC COMPONENTS

Failure Number	Operating Time, Hr.
1	8
2	20
3	34
4	46
5	63
6	86
7	111
8	141
9	186
10	266

From Table 8.3-4 and Eq. (8.1) and (8.2) one can calculate and plot F(t) and R(t). The data plots for the various function of interest are shown in Figure 8.3-4.

Note, from the dashed lines of Figure 8.3-4 (a) and (b), that the exponential distribution of time to failure represents a good approximation to the data.

TABLE 8.3-4: COMPUTATION OF DATA FAILURE DENSITY AND DATA HAZARD RATE

Time Interval, Hour t	Failure Density per Hour $f(t) \times 10^{-2}$	Hazard Rate per Hour $h(t) \times 10^{-2}$
0 - 8	$\frac{1}{10 \times 8} = 1.25$	$\frac{1}{10 \times 8} = 1.25$
8 - 20	$\frac{1}{10 \times 12} = 0.83$	$\frac{1}{9 \times 12} = 0.93$
20 - 34	$\frac{1}{10 \times 14} = 0.71$	$\frac{1}{8 \times 14} = 0.89$
34 - 46	$\frac{1}{10 \times 12} = 0.83$	$\frac{1}{7 \times 12} = 1.19$
46 - 63	$\frac{1}{10 \times 17} = 0.59$	$\frac{1}{6 \times 17} = 0.98$
63 - 86	$\frac{1}{10 \times 23} = 0.43$	$\frac{1}{5 \times 23} = 0.87$
86 - 111	$\frac{1}{10 \times 25} = 0.40$	$\frac{1}{4 \times 25} = 1.00$
111 - 141	$\frac{1}{10 \times 30} = 0.33$	$\frac{1}{3 \times 30} = 1.11$
141 - 186	$\frac{1}{10 \times 45} = 0.22$	$\frac{1}{2 \times 45} = 1.11$
186 - 266	$\frac{1}{10 \times 80} = 0.13$	$\frac{1}{1 \times 80} = 1.25$

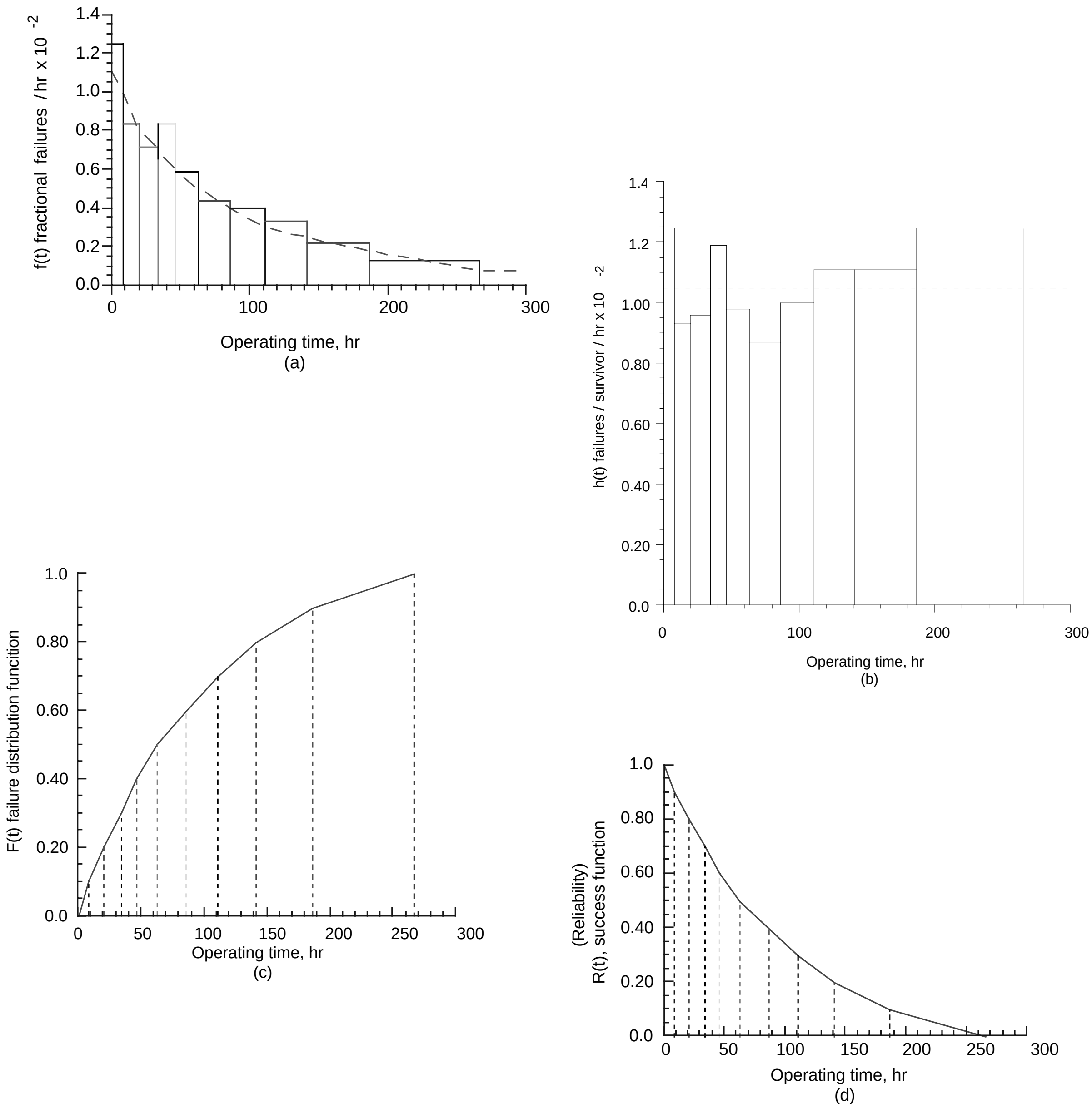


FIGURE 8.3-4: HAZARD AND DENSITY FUNCTIONS FOR TABLE 8.3-3

Example 5:

Data for a single B-52 performing 1000 missions of 2 to 24 hours, or the equivalent of 1000 B-52s performing a single mission of 2 to 24 hours, are shown in Tables 8.3-5 and 8.3-6 and Figure 8.3-5 (Ref. [6]). This data shows that the B-52 is most likely to fail in the first two hours of its mission. The exponential distribution of times to failure does not fit well to the data.

TABLE 8.3-5: FAILURE DATA FOR 1,000 B-52 AIRCRAFT

Time Until Failure, Hour	Number of Failures in Interval	f(t) Failure Density/Hr.	h(t) Hazard Rate/Hr.
0 - 2	222	$\frac{222}{1,000 \times 2} = 0.1110$	$\frac{222}{1,000 \times 2} = 0.1110$
2 - 4	45	$\frac{45}{1,000 \times 2} = 0.0225$	$\frac{45}{778 \times 2} = 0.0289$
4 - 6	32	$\frac{32}{1,000 \times 2} = 0.0160$	$\frac{32}{733 \times 2} = 0.0218$
6 - 8	27	$\frac{27}{1,000 \times 2} = 0.0135$	$\frac{27}{701 \times 2} = 0.0192$
8 - 10	21	$\frac{21}{1,000 \times 2} = 0.0105$	$\frac{21}{674 \times 2} = 0.0156$
10 - 12	15	$\frac{15}{1,000 \times 2} = 0.0075$	$\frac{15}{653 \times 2} = 0.0113$
12 - 14	17	$\frac{17}{1,000 \times 2} = 0.0085$	$\frac{17}{638 \times 2} = 0.0133$
14 - 16	7	$\frac{7}{1,000 \times 2} = 0.0035$	$\frac{7}{621 \times 2} = 0.0056$
16 - 18	14	$\frac{14}{1,000 \times 2} = 0.0070$	$\frac{14}{614 \times 2} = 0.0114$
18 - 20	9	$\frac{9}{1,000 \times 2} = 0.0045$	$\frac{9}{600 \times 2} = 0.0075$
20 - 22	8	$\frac{8}{1,000 \times 2} = 0.0040$	$\frac{8}{591 \times 2} = 0.0068$
22 - 24	3	$\frac{3}{1,000 \times 2} = 0.0015$	$\frac{3}{583 \times 2} = 0.0026$
TOTAL	420		

TABLE 8.3-6: TIME-TO-FAILURE DATA FOR S = 1000 MISSION HOURS

TIME-TO- FAILURE (HOURS)	CUMULATIVE FAILURES = F	$R = \frac{1000 - F}{1000}$
2	222	.778
4	267	.733
6	299	.701
8	326	.674
10	347	.653
12	362	.638
14	379	.621
16	386	.614
18	400	.600
20	409	.591
22	417	.583
24	420	.580

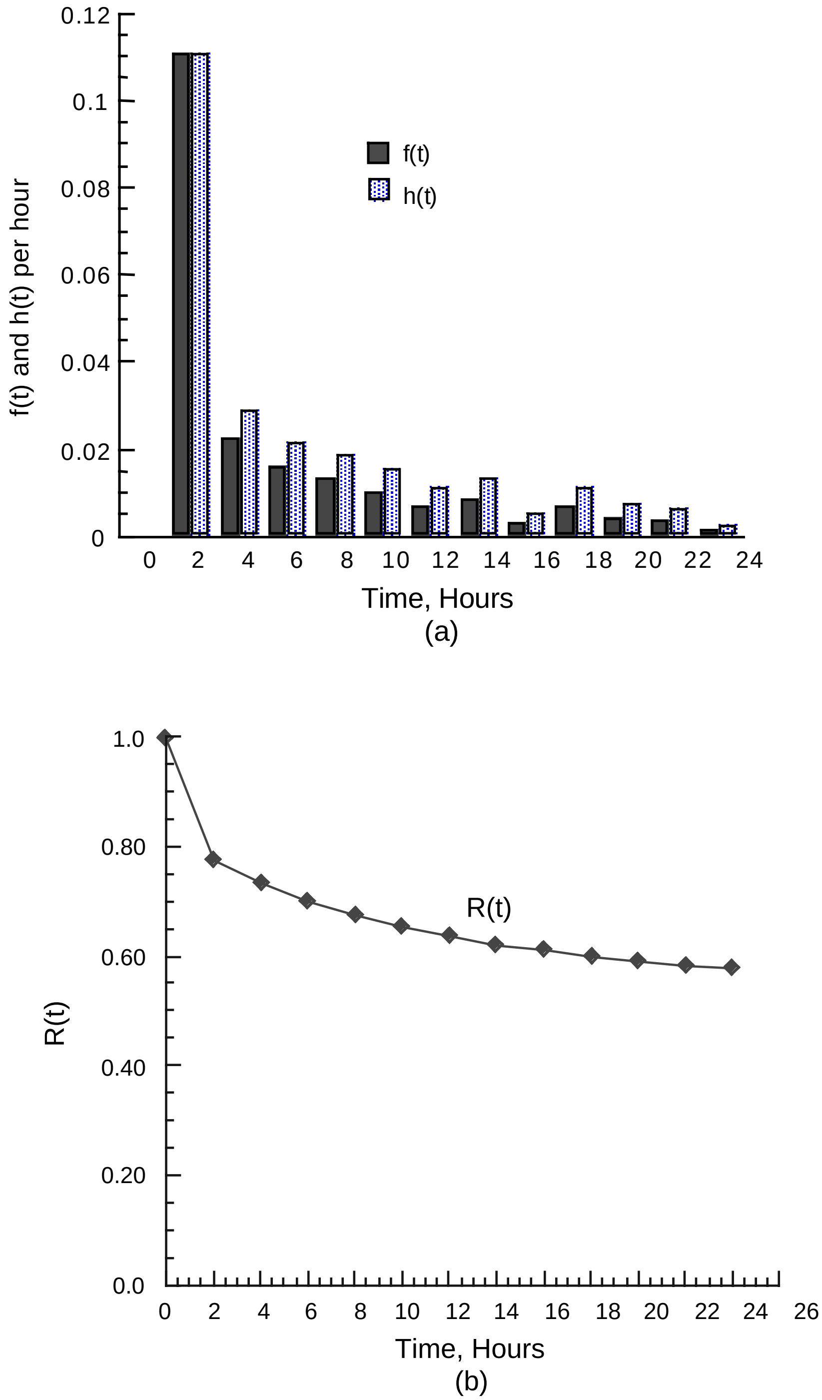


FIGURE 8.3-5: RELIABILITY FUNCTIONS FOR THE EXAMPLE GIVEN IN TABLE 8.3-4

8.3.2.3 Reliability Function (Survival Curves)

A survival curve or reliability function, R(t), is a graphic representation of the relationship between the probability of survival and time. Here, probability of survival is synonymous with probability of nonfailure or probability of satisfactory performance. Three types of survival curves are of primary interest. The first is a discrete or point-type curve derived from observed data by nonparametric or distribution-free methods. The second type is a continuous curve based on an assumption as to the form of the distribution (Gaussian, exponential, etc.) and on values of the distribution parameters estimated from the observed data. The third type of curve is the true reliability function of the population from which the sample observations were drawn. This last function can only be estimated (i.e., not determined precisely), although the limits within which it will fall a given percentage of the time can be defined.

Figure 8.3-6 presents a frequency distribution of failures in a fixed population of 90 items, over a 6-hour period. To obtain a survival curve from these data, the following simplified method is used.

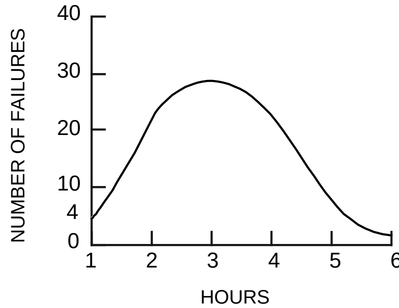
During the first period of observation, from 0 to 1 hour, 4 of the original 90 items failed. The failure rate during this period was 4/90, or 0.0444, which is equivalent to a survival rate of 1 - 0.0444, or 0.9556. In the second period of observation, 21 of the 86 remaining items failed. The failure rate was 21/86, or 0.244, and the survival rate was 1 - 0.244, or 0.756. The tabulation above Figure 8.3-7 gives the failure rates and survival rates for the remaining periods of observation. It will be noted that the failure rate increases with time.

To obtain a survival curve, which is the cumulative probability of survival with time, the probability of survival in each time period is multiplied by the survival rate in the succeeding time period. Thus, 0.9555 x 0.756 = 0.723; 0.723 x 0.538 = 0.388, etc. The probability values are plotted versus the centers of the time periods as shown at the bottom of 8.3-7.

Figure 8.3-8 presents a frequency distribution of failures for a population of 90 items in which the removal rate is constant with time. The approach described in connection with the normal curve yields the tabulation and exponential survival curve shown in Figure 8.3-9. (Note in this example, only 83 of 90 items failed in six hours).

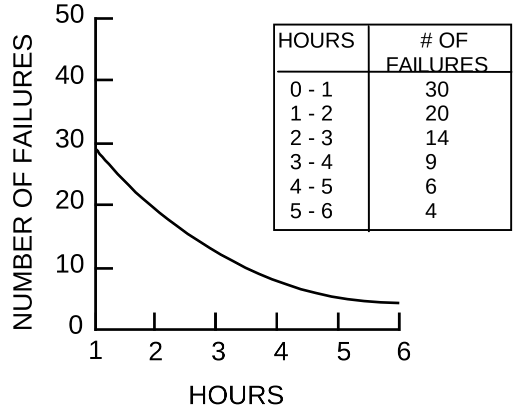
Survival curves for most electronic equipment/systems are of the exponential form. Survival curves for mechanical parts, on the other hand, are frequently of the normal or Weibull form. As parts wear out, their failure rate increases and their probability of survival decreases. A large number of such parts, all having normal or Weibull survival curves but each having a different mean life and variance, will produce a system malfunction rate which is essentially constant, since the mean lives of the parts will be randomly distributed.

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH



HOURS	# OF FAILURES
0 - 1	4
1 - 2	21
2 - 3	30
3 - 4	25
4 - 5	8
5 - 6	2

FIGURE 8.3-6: NORMAL DISTRIBUTION OF FAILURE IN TIME



HOURS	# OF FAILURES
0 - 1	30
1 - 2	20
2 - 3	14
3 - 4	9
4 - 5	6
5 - 6	4

FIGURE 8.3-8: EXPONENTIAL DISTRIBUTION OF FAILURES IN TIME

TIME	FAILURE RATE	SURVIVAL RATE	PROBABILITY OF SURVIVAL
0 - 1	0.444	0.9556	0.9555
1 - 2	0.2442	0.7558	0.7230
2 - 3	0.4615	0.5385	0.3880
3 - 4	0.7143	0.2857	0.1110
4 - 5	0.8000	0.2000	0.0220
5 - 6	1.0000	0.0000	- - -

NOTE: Population is 90 for all figures.

TIME	FAILURE RATE	SURVIVAL RATE	PROBABILITY OF SURVIVAL
0 - 1	0.333	0.667	0.667
1 - 2	0.333	0.667	0.444
2 - 3	0.350	0.650	0.289
3 - 4	0.346	0.654	0.189
4 - 5	0.353	0.647	0.122
5 - 6	0.364	0.636	0.078

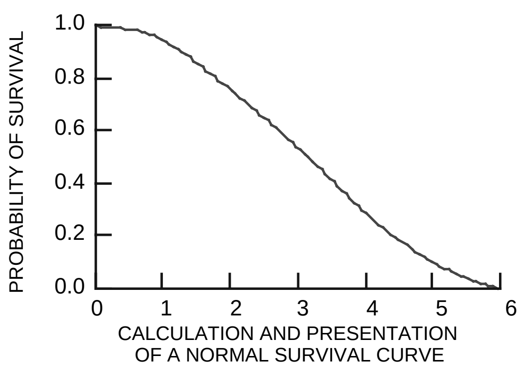


FIGURE 8.3-7: CALCULATION AND PRESENTATION OF A NORMAL SURVIVAL CURVE

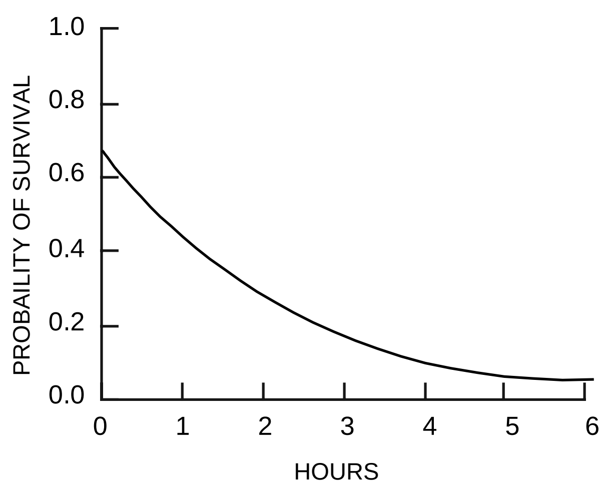


FIGURE 8.3-9: CALCULATION AND PRESENTATION OF AN EXPONENTIAL CURVE

To determine what type of population gives rise to a particular survival curve, the theoretical reliability function most closely resembling the curve is computed from sample parameters. The theoretical function is then matched to the observed curve by statistical techniques. If this procedure establishes that there is no significant difference between the observed and theoretical curves, the theoretical curve is usually employed for all additional calculations.

Figures 8.3-10 and 8.3-11 portray observed and theoretical probability of survival curves for the case of exponential and normal distributions of time to failure. Note that the mean life for the exponential case has R(t) = 0.368, whereas for the normal case, R(t) = 0.5. This is due to the symmetrical characteristic of the normal distribution, versus the skewed characteristic of the exponential.

Thus, if one can develop a mathematical expression for R(t), it can be shown that the mean time to failure is given by:

$$\text{MTTF} = \int_0^{\infty} R(t)\, dt \qquad (8.3)$$

8.3.2.3.1 Computation of Theoretical Exponential Reliability Function

When the form of the distribution is sufficiently well defined, it is possible to estimate the reliability function in terms of the parameters of the distribution. This method has the advantage of permitting utilization of all the accumulated knowledge concerning the items in the population. In addition, the reliability function can be summarized by specifying the values of the parameters, and can be compared with other reliability functions merely by comparing the values of the summarized data.

For the case of an equipment/system which is repaired upon failure, the reliability function is given by:

$$R(t) = e^{-t/\text{MTBF}} \qquad (8.4)$$

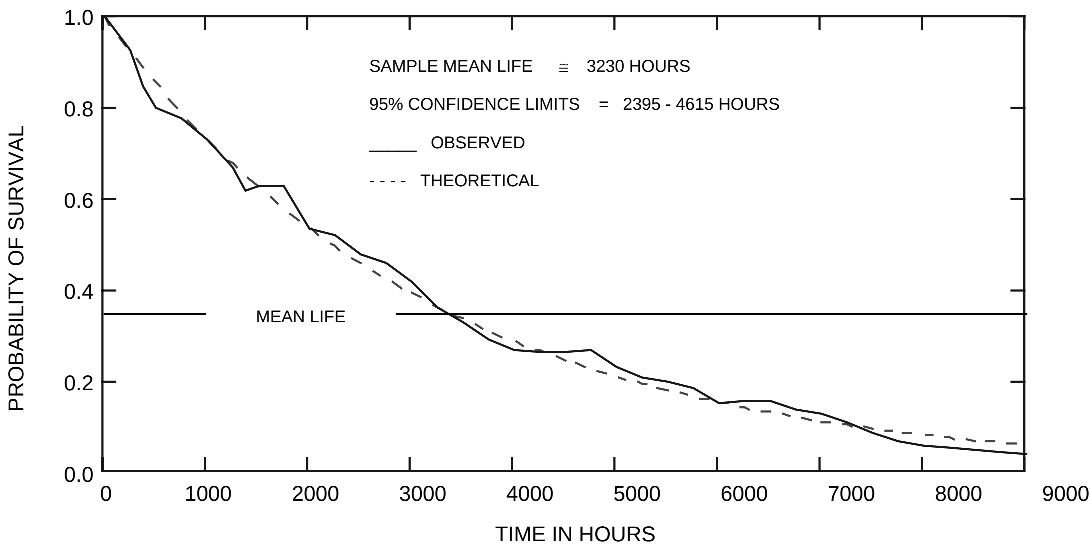


FIGURE 8.3-10: OBSERVED AND THEORETICAL EXPONENTIAL SURVIVAL CURVES

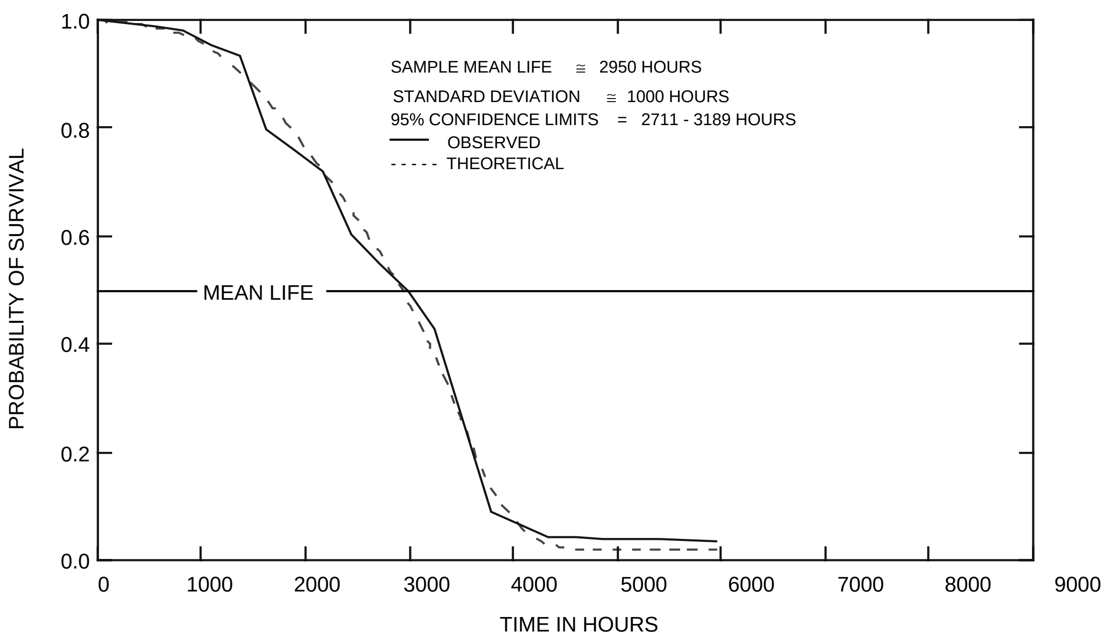


FIGURE 8.3-11: OBSERVED AND THEORETICAL NORMAL SURVIVAL CURVES

where:

t = time at which R(t) is calculated
MTBF = mean time between failures, given by

$$MTBF = \frac{nt}{r} \tag{8.5}$$

and

n = the number of equipments operated to time t
r = the number of failures, with the last failure occurring at time t

For example, assume that in a sample of twenty equipments operated for 773 hours, we observed 10 failures (each of which was repaired), with the last failure occurring at 773 hours.

Then

$$MTBF = \frac{nt}{r} = \frac{(20)(773)}{10} = 1546 \text{ hours}$$

$$R(t) = e^{-t/1546}$$

Table 8.3-7 shows the computations for R(t) for selected values of t. Figure 8.3-12 shows the actual reliability function (solid line) plotted from the data versus the theoretical exponential function from column 3 of Table 8.3-7. Determination of confidence intervals is discussed briefly in the next section.

8.3.2.3.2 Computation For Normal Reliability Function

Table 8.3-8 presents some observed failure data for a sample of twenty units tested to failure, and the failure times observed. The units were known to follow a normal distribution of time to failure.

The sample mean, $\overline{X}$, an estimate of μ, is given by:

$$\overline{X} = \sum_{i=1}^{20} X_i \,/n = \frac{39104}{20} = 1955.2 \text{ hours}$$

TABLE 8.3-7: COMPUTATION OF THEORETICAL EXPONENTIAL RELIABILITY FUNCTION FOR MTBF = 1546 HOURS

(1) t	(2) t/MTBF	(3) $e^{-t/MTBF}$
0	0	1.000
96	0.0621	0.9389
216	0.1397	0.8696
312	0.2018	0.8173
456	0.2950	0.7445
552	0.3571	0.6997
696	0.4502	0.6375
792	0.5123	0.5991
888	0.5744	0.5630
960	0.6210	0.5374
1200	0.7762	0.4602
1416	0.9159	0.4002
1546	1.0000	0.3679
1896	1.2264	0.2933
2064	1.3351	0.2631

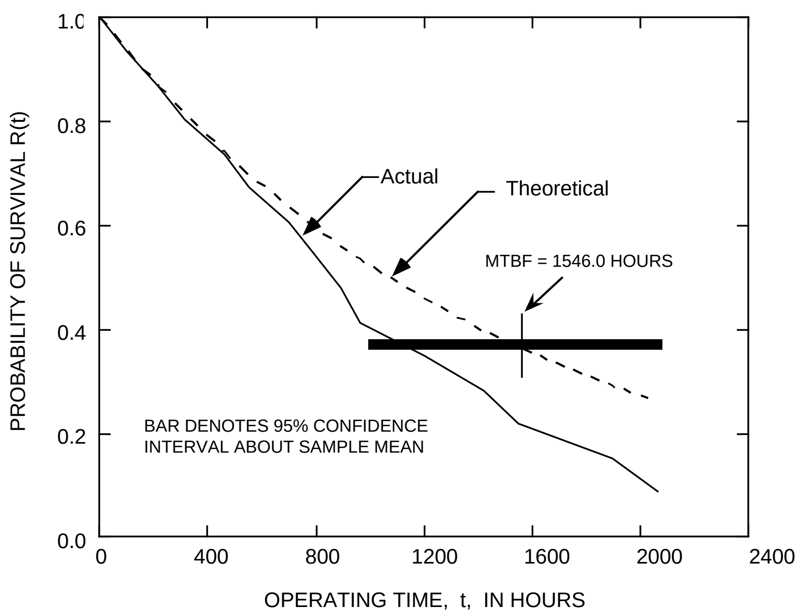


FIGURE 8.3-12: ACTUAL RELIABILITY FUNCTION AND THEORETICAL EXPONENTIAL RELIABILITY FUNCTION

TABLE 8.3-8: OBSERVED FAILURE DATA

Time	Probability of Survival, R	Time	Probability of Survival, R
175	0.95	2025	0.45
695	0.90	2115	0.40
872	0.85	2172	0.35
1250	0.80	2418	0.30
1291	0.75	2583	0.25
1402	0.70	2725	0.20
1404	0.65	2844	0.15
1713	0.60	2980	0.10
1741	0.55	3268	0.05
1893	0.50	3538	0.00

The sample standard deviation, s, an estimate of σ, is given by:

$$s = \left[\frac{\sum_{i=1}^{20} (X_i - \overline{X})^2}{n-1} \right]^{.5} = 886.6 \text{ hours}$$

where:

X_i = i^{th} failure time

n = sample size

$\overline{X}$ = sample mean

Figure 8.3-13 shows the actual or nonparametric reliability function plotted from the data versus the theoretical function calculated using the estimates of μ and σ. The theoretical values were obtained from the expression

$$R(x) = P\left(z > \frac{X - \mu}{\sigma} \right)$$

where the value of z was obtained from a table of the Standard Normal Distribution (Table 5.3.1 of Section 5).

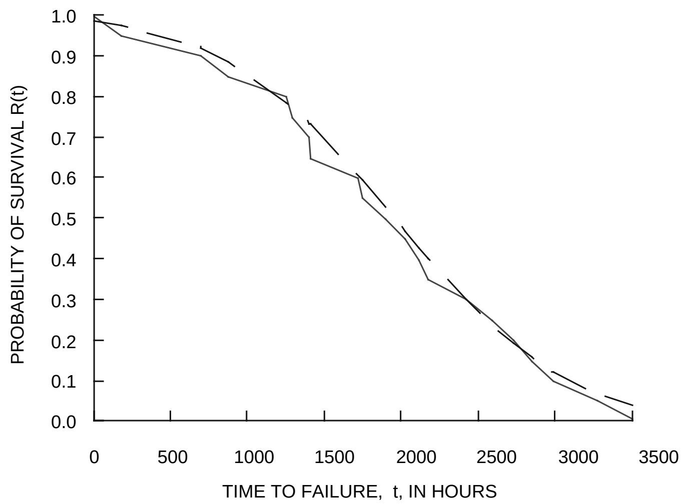


FIGURE 8.3-13: NON-PARAMETRIC AND THEORETICAL NORMAL RELIABILITY FUNCTIONS

8.3.2.4 Censored Data

If a sample contains both complete and incomplete lifetimes, the incomplete lifetimes are referred to as "censored" observations. These consist primarily of lifetimes which are too long to be observed completely ("terminated" observations) and lifetimes in which the item being observed is lost before completion of observation ("lost" observation). In the case of terminated observations, the length of observation time is controlled; in the case of lost observations, the length of observation time is not controlled. In either case, the investigator knows that the lifetime of the item exceeds the period of time during which the item was being observed. Terminated observations do not present a problem to the investigator other than to increase the complexity of the calculations, but lost observations may constitute a real problem because they maybe associated with only a portion of the population.

For example, for the case of the exponential distribution in which n items are put on test, r of them fail at time t_1, t_2 . . . t_r, with the test discontinued at t_r when the r^{th} failure occurs, the MTBF is given by

$$\text{MTBF} = \frac{\sum_{i=1}^{r} t_i + (n-r)t_r}{n} \tag{8.6}$$

where t_i is the time of each failure and (n - r) represents the number of surviving items at time t_r. In this nonreplacement case, the failed items are not repaired or replaced upon failure.

The mathematics become somewhat more difficult when analyzing censored data where distributions other than the exponential are involved, or when using nonparametric methods. These cases are treated in detail in References [1], [3], [4] and [5].

8.3.2.5 Confidence Limits and Intervals

Previously, we discussed methods of obtaining point estimates of reliability parameters, e.g., R(t), λ, MTBF, etc. For most practical applications, we are interested in the accuracy of the point estimate and the confidence which we can attach to it. We know that statistical estimates are more likely to be closer to the true value as the sample size increases. Only the impossible situation of having an infinitely large number of samples to test could give us 100 percent confidence or certainty that a measured value of a parameter coincides with the true value. For any practical situation, therefore, we must establish confidence intervals or ranges of values between which we know, with a probability determined by the finite sample size, that the true value of the parameter lies.

Confidence intervals around point estimates are defined in terms of a lower confidence limit, L, and an upper confidence limit, U. If, for example, we calculate the confidence limits for a probability of, say, 95 percent, this means that in repeated sampling, 95 percent of the calculated intervals will contain the true value of the reliability parameter. If we want to be 99 percent sure that the true value lies within certain limits for a given sample size, we must widen the interval or test a larger number of samples if we wish to maintain the same interval width. The problem, then, is reduced to one of either determining the interval within which the true parametric value lies with a given probability for a given sample size, or determining the sample size required to assure us with a specified probability that true parametric value lies within a specific interval.

Thus, we would like to be able to make assertions such as

$$P\left[\left(\hat{\theta}_L < \theta < \hat{\theta}_U\right)\right] = \eta \qquad (8.8)$$

where θ is some unknown population parameter, θ_L and θ_U are estimators associated with a random sample and η is a probability value such as 0.99, 0.95, 0.90, etc. If, for instance, $\eta = 0.95$ we refer to the interval

$$\left(\theta_L < \theta < \theta_U\right) \qquad (8.9)$$

for particular values of $\hat{\theta}_L$ and $\hat{\theta}_U$ as a 95% confidence interval. In this case we are willing to accept a 5% probability (risk) that our assertion is not, in fact, true.

Or, we may also want to make statements such as

$$P\left[\theta > \hat{\theta}_L\right] = \eta \qquad (8.10)$$

in which case we make statements like, "we are 90% confident that the true MTBF is greater than some lower confidence limit (or measured value)." Eq. (8.10) is the case of the one-sided confidence limit, versus Eq. (8.9) which is a two-sided confidence limit, or confidence interval.

To help clarify the concept of a confidence interval we can look at the situation in a geometrical way. Suppose we draw repeated samples (x_1, x_2) from a population, one of whose parameters we desire to bracket with a confidence interval. We construct a three-dimensional space with the vertical axis corresponding to θ and with the two horizontal axes corresponding to values of X_1 and X_2 (see Figure 8.3-14). The actual value of the population parameter θ is marked on the vertical axis and a horizontal plane is passed through this point. Now we take a random sample (X_1, X_2) from which we calculate the values $\hat{\theta}_U$ and $\hat{\theta}_L$ at, say, the 95% confidence level. The interval defined by $\hat{\theta}_U$ and $\hat{\theta}_L$ is plotted on the figure.

Next, we take a second sample (X'_1, X'_2) from which we calculate the value $\hat{\theta}'_U$ and $\hat{\theta}'_L$ at the 95% level. This interval is plotted on the figure. A third sample (X''_1, X''_2) yields the values $\hat{\theta}''_U$ and $\hat{\theta}''_L$, etc. In this way we can generate a large family of confidence intervals. The confidence intervals depend only on the sample values (X_1, X_2), (X'_1, X'_2), etc., and hence we can calculate these intervals without knowledge of the true value of θ. If the confidence intervals are all calculated on the basis of 95% confidence and if we have a very large family of these intervals, then 95% of them will cut the horizontal plane through θ (and thus include θ) and 5% of them will not.

The process of taking a random sample and computing from it a confidence interval is equivalent to the process of reaching into a bag containing thousands of confidence intervals and grabbing one at random. If they are all 95% intervals, our chance of choosing one that does indeed include θ will be 95%. In contrast, 5% of the time we will be unlucky and select one that does not include θ (like the interval $\hat{\theta}''_U, \hat{\theta}''_L$) in Figure 8.3-14. If a risk of 5% is judged too high, we can go to 99% intervals, for which the risk is only 1%. As we go to higher confidence levels (and lower risks) the lengths of the intervals increase until for 100% confidence levels (and lower risks) the interval includes every conceivable value of θ (I am 100% confident that the number of defective items in a population of 10,000 is somewhere between 0 and 10,000). For this reason 100% confidence intervals are of little interest.

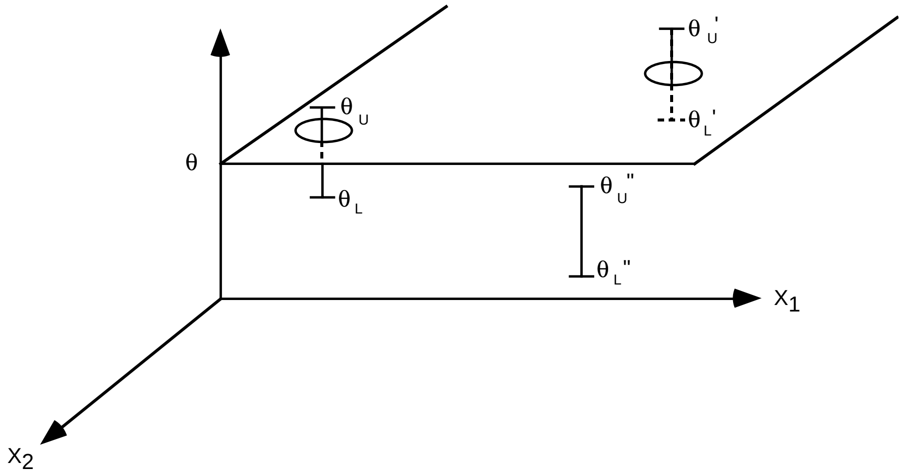


FIGURE 8.3-14: GEOMETRICAL INTERPRETATION OF THE CONCEPT OF A CONFIDENCE INTERVAL

Let us now look at some simple examples of how these concepts are applied to analyze reliability for some of the more commonly-used distributions.

8.3.2.5.1 Confidence Limits - Normal Distribution

When the lives of n components are known from a wearout test and we compute their mean, $\hat{M}$, and their standard deviation, s, and when n is large so that we can assume that $s \approx \sigma$, the upper and lower confidence limits can be readily evaluated from Table 8.3-9 for the more commonly-used confidence levels.

Strictly speaking, this procedure of assigning confidence intervals to an estimate is correct only when the true standard deviation, σ, of component wearout is known and used instead of s in Table 8.3-9. However, it can be applied in reliability work as an approximation whenever the estimate s, of σ, was obtained from a large sample, i.e., when the number of failures is at least 25, and preferably, more. In fact, it can be shown for samples of 20, $K_{\alpha/2}$ (at the 95% confidence level) is 2.09 vs. a value of 1.96 for an infinite number of samples. α is equal to 100(1 - confidence level)%.

TABLE 8.3-9: CONFIDENCE LIMITS - NORMAL DISTRIBUTION

$k_{\alpha/2}$	Two-sided Confidence intervals $M \pm K_{\alpha/2}\ s/\sqrt{n}$	Confidence levels $100(1 - \alpha)\%$
0.84	$\hat{M} \pm 0.84s/\sqrt{n}$	60.0
1.28	$\hat{M} \pm 1.28s/\sqrt{n}$	80.0
1.64	$\hat{M} \pm 1.64s/\sqrt{n}$	90.0
1.96	$\hat{M} \pm 1.96s/\sqrt{n}$	95.0
2.58	$\hat{M} \pm 2.58s/\sqrt{n}$	99.0

Figure 8.3-15 graphically illustrates what is being done. Since the normal distribution is symmetrical, we are computing the confidence interval as the area $(1 - \alpha)$ under the curve, leaving an area $\alpha/2$ in each of the left and right hand tails which is outside of the confidence interval (CI). For example, using the calculated values of $\hat{M}$ (or $\overline{X}$) and s obtained from the data in Table 8.3-10, the CI at the 95% level is

$$
\begin{aligned}
\hat{M} \pm 1.96\ s/\sqrt{n} &= 1955.2 \pm 1.96\ (886.6)/\sqrt{20} \\
&= 1955.2 \pm 388.6 \\
&= (2343.8,\ 1566.6)
\end{aligned}
$$

In other words, we can be 95% confident that the true value of the mean life (M) lies between 1566.6 and 2343.8 hours.

Actually, in reliability work, we are usually more interested in the lower confidence limit L of the mean wearout life than in the upper limit. Given a measured value of $\hat{M}$, we would like to make some statement about our confidence that the true value of M exceeds some minimum value.

When only the lower confidence limit, L, is of interest, we apply the procedure of so-called "one-sided" confidence limits, as opposed to the two-sided CI of the preceding example. The problem is to assure ourselves (or our customer) that the true mean life, M, is equal to or larger than some specified minimum value with a probability of $(1 - \alpha)$.

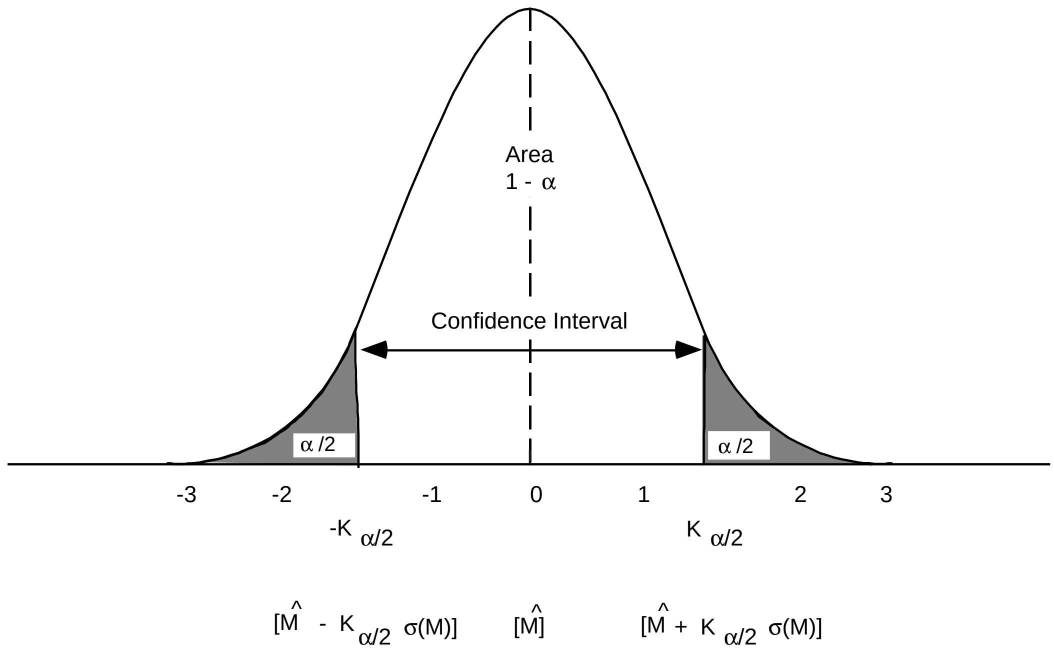


FIGURE 8.3-15: TWO-SIDED CONFIDENCE INTERVAL AND LIMITS

Whereas in the case of the two-sided confidence limits, we had an area $\alpha/2$ under the left tail of the normal curve (Figure 8.3-15), we now have an area α to the left of L and an area $(1 - \alpha)$ to the right.

Therefore, the estimate of mean life obtained from the data should be:

$$\hat{M} \geq L + K_{\alpha} \sigma/\sqrt{n} \qquad (8.11)$$

If this equation is not satisfied, the requirement that the true M must be at least L at the specified 100 $(1 - \alpha)$ percent confidence level has not been fulfilled.

Table 8.3-10, in which the assumption $s \approx \sigma$ is made, allows a quick check as to whether an estimate, $\hat{M}$, obtained from a sample of size n fulfills the requirement that the true M must not be smaller than the specified minimum L. Only the more commonly-used confidence levels are given.

TABLE 8.3-10: CONFIDENCE INTERVAL

$K_{\alpha/2}$	The estimate $\hat{M}$ must exceed: $L + K_{\alpha/2}\ s/\sqrt{n}$	Confidence levels $100\ (1 - \alpha)\%$
0.25	$L + 0.25s/\sqrt{n}$	60
0.52	$L + 0.52s/\sqrt{n}$	70
0.84	$L + 0.84s/\sqrt{n}$	80
1.28	$L + 1.28s/\sqrt{n}$	90
1.64	$L + 1.64s/\sqrt{n}$	95
2.33	$L + 2.33s/\sqrt{n}$	99

Once again, using the data and calculated values of $\hat{M}$ and s from Table 8.3-10, assume that we would like to be 95% confident that the true $M \geq 1500$ hours. The equation from Table 8.3-10 is

$$\hat{M} \geq L + 1.64\ s/\sqrt{n}$$

$$1955.2 \geq 1500 + 1.64\ (886.6)/\sqrt{20}$$

$$1955.2 \geq 1500 + 325$$

$$1955.2 \geq 1825$$

Since the inequality is satisfied, the requirement has been met.

As previously mentioned, the above procedure can be applied if the sample size n is at least 25. However, similar procedures also apply to smaller sample sizes except that now we cannot assume that $s = \sigma$, and we must use another set of equations based on Student's t distribution. Actually, all we do is replace the normal percentage points $K_{\alpha/2}$ and K_{α} in the previously developed equations by the tabulated percentage points $t_{\alpha/2;n-1}$ and $t_{\alpha;n-1}$ of the t distribution, where n-1 is called the degrees of freedom and n is the number of failures. Student's t tables are available in most standard statistical texts.

For example, for the two-sided CI example using the data from Table 8.3-10 and calculated values of $\hat{M}$ and s,

$$\hat{M} \pm t_{\alpha/2;n-1}\, s/\sqrt{n} = 1955.2 \pm 2.09\,(886.6)/\sqrt{20}$$

$$= 1955.2 \pm 414.4$$

$$= (2370,\ 1541.2)$$

which is a slightly wider CI than the case where it was assumed the s ≈ σ.

8.3.2.5.2 Confidence Limits - Exponential Distribution

Two situations have to be considered for estimating confidence intervals: one in which the test is run until a preassigned number of failures (r*) occurs, and one in which the test is stopped after a preassigned number of test hours (t*) is accumulated. The formula for the confidence interval employs the X_2 (chi-square) distribution. A short table of X^2 values is given in Table 8.3-11. The general notation used is

$$\chi^2_{p,d}$$

where p and d are two constants used to choose the correct value from the table.

The quantity p is a function of the confidence coefficient; d, known as the degrees of freedom, is a function of the number of failures. $X^2_{\alpha/2,\ 2r+2}$ for example, is the $\frac{a}{2}$ percentage point of the chi-square distribution for (2r+2) degrees of freedom.

Equations (8.12) and (8.13) are for one-sided or two-sided 100(1 - α) percent confidence intervals. For nonreplacement tests with a fixed truncation time, the limits are only approximate. Also, for non-replacement tests, only one sided intervals are possible for r = 0.

TABLE 8.3-11: DISTRIBUTION OF χ^2 (CHI-SQUARE)

DF	Probability													
	0.99	0.975	0.95	0.90	0.80	0.75	0.50	0.25	0.20	0.10	0.05	0.025	0.01	0.001
1	0.00016	0.00098	0.00393	0.0158	0.0642	0.10153	0.455	1.323	1.642	2.706	3.841	5.024	6.635	10.827
2	0.0201	0.0506	0.103	0.211	0.446	0.5753	1.386	2.772	3.219	4.605	5.991	7.377	9.210	13.815
3	0.115	0.216	0.352	0.584	1.005	1.2125	2.366	4.108	4.642	6.251	7.815	9.348	11.341	16.268
4	0.297	0.484	0.711	1.064	1.649	1.9225	3.357	5.385	5.989	7.779	9.448	11.143	13.277	18.465
5	0.554	0.831	1.145	1.610	2.343	2.674	4.351	6.625	7.289	9.236	11.070	12.832	15.086	20.517
6	0.872	1.237	1.635	2.204	3.070	3.454	5.348	7.840	8.558	10.645	12.592	14.449	16.812	22.457
7	1.239	1.689	2.167	2.833	3.822	4.254	6.346	9.037	9.803	12.017	14.067	16.013	18.475	24.322
8	1.646	2.179	2.733	3.490	4.594	5.070	7.344	10.218	11.030	13.362	15.507	17.534	20.090	26.125
9	2.088	2.700	3.325	4.168	5.380	5.898	8.343	11.388	12.242	14.684	16.919	19.023	21.666	27.877
10	2.558	3.247	3.940	4.865	6.179	6.737	9.342	12.548	13.442	15.987	18.307	20.483	23.209	29.588
11	3.053	3.816	4.575	5.578	6.989	7.584	10.341	13.701	14.631	17.275	19.675	21.920	24.725	31.264
12	3.571	4.404	5.226	6.304	7.807	8.438	11.340	14.845	15.812	18.549	21.026	23.336	26.217	32.909
13	4.107	5.008	5.892	7.042	8.634	9.299	12.340	15.984	16.985	19.812	22.362	24.735	27.688	34.528
14	4.660	5.628	6.571	7.790	9.467	10.165	13.339	17.117	18.151	21.064	23.685	26.119	29.141	36.123
15	5.229	6.262	7.261	8.547	10.307	11.036	14.339	18.245	19.311	22.307	24.996	27.488	30.578	37.697
16	5.812	6.907	7.962	9.312	11.152	11.912	15.338	19.368	20.465	23.542	26.296	28.845	32.000	39.252
17	6.408	7.564	8.672	10.085	12.002	12.791	16.338	20.488	21.615	24.769	27.587	30.191	33.409	40.790
18	7.015	8.231	9.390	10.865	12.857	13.675	17.338	21.605	22.760	25.989	28.869	31.526	34.805	42.312
19	7.633	8.906	10.117	11.651	13.716	14.562	18.338	22.717	23.900	27.204	30.144	32.852	36.191	43.820
20	8.260	9.591	10.851	12.443	14.578	15.452	19.337	23.827	25.038	28.412	31.410	34.169	37.566	45.315
21	8.897	10.283	11.591	13.240	15.445	16.344	20.337	24.935	26.171	29.615	32.671	35.479	38.932	46.797
22	9.542	10.982	12.338	14.041	16.314	17.239	21.337	26.039	27.301	30.813	33.924	36.780	40.289	48.268
23	10.196	11.688	13.091	14.848	17.187	18.137	22.337	27.141	28.429	32.007	35.172	38.075	41.638	49.728
24	10.856	12.400	13.848	15.659	18.062	19.037	23.337	28.241	29.553	33.196	36.415	39.364	42.980	51.179
25	11.524	13.119	14.611	16.473	18.940	19.939	24.337	29.339	30.675	34.382	37.652	40.646	44.314	52.620
26	12.198	13.844	15.379	17.292	19.820	20.843	25.336	30.434	31.795	35.563	38.885	41.923	45.642	54.052
27	12.879	14.573	16.151	18.114	20.703	21.749	26.336	31.528	32.912	36.741	40.113	43.194	46.963	55.476
28	13.565	15.308	16.928	18.933	21.588	22.657	27.336	32.620	34.027	37.916	41.337	44.460	48.278	56.893
29	14.256	16.047	17.708	19.768	22.475	23.566	28.336	33.711	35.139	39.087	42.557	45.722	49.588	58.302
30	14.953	16.791	18.493	20.599	23.364	24.476	29.336	34.799	36.250	40.256	43.773	46.98	50.892	59.703

Equations for Confidence Limits on Mean Life

Type of Confidence Limits	Fixed Number of Failures, r^*	Fixed Truncation Time t^*	
One Sided (Lower Limit)	$\left(\frac{2T}{\chi^2(\alpha, 2r)}, \infty\right)$	$\left(\frac{2T}{\chi^2(\alpha, 2r+2)}, \infty\right)$	(8.12)
Two Sided (Upper and Lower Limits)	$\left(\frac{2T}{\chi^2(\frac{\alpha}{2}, 2r)}, \frac{2T}{\chi^2(1-\frac{\alpha}{2}, 2r)}\right)$	$\left(\frac{2T}{\chi^2(\frac{\alpha}{2}, 2r+2)}, \frac{2T}{\chi^2(1-\frac{\alpha}{2}, 2r)}\right)$	(8.13)

The terms used are identified as follows:

- n = number of items placed on test at time $t = 0$
- t^* = time at which the life test is terminated
- θ = mean life (or MTBF for the case of replacement or repair upon failure)
- r = number of failures accumulated at time t^*
- r^* = preassigned number of failures
- α = acceptable risk of error
- $1 - \alpha$ = confidence level
- T = total test time

Note that T is computed as follows, depending on the type of test procedure.

Replacement Tests (failure replaced or repaired) $T = nt^*$ (8.14)

Non-Replacement Tests $T\sum_{i=1}^{r} t_i + (n - r)t^*$ (8.15)

where t_i = time of the i^{th} failure

Censored Items (withdrawal or loss of items which have not failed)

(a) If failures are replaced and censored items are not replaced

$$T = \sum_{j=1}^{c} t_j + (n - c)t^* \qquad (8.16)$$

where:

t_j = time of censorship

c = number of censored items

(b) If failures are not replaced

$$T = \sum_{i=1}^{r} t_i + \sum_{j=1}^{c} t_j + (n - r - c)t^* \qquad (8.17)$$

<u>Example 6</u>:

Twenty items undergo a replacement test. Testing continues until ten failures are observed. The tenth failure occurs at 80 hours. Determine (1) the mean life of the items; and (2) the one-sided and two- sided 95% confidence intervals for the MTBF.

(1) From Equation (8.4)

$$\text{MTBF} = \frac{nt^*}{r} = \frac{(20)(80)}{10} = 160 \text{ hours}$$

(2) α = 1 - Confidence Level = 1 - 0.95 = 0.05

2r = 2(number of failures) = 2(10) = 20

$$C\left[\frac{2T}{\chi^2_{(\alpha,2r)}}, \infty\right] = C\left[\frac{2(1600)}{\chi^2_{(0.05,20)}}, \infty\right] = C\left[\frac{3200}{31.41}, \infty\right] = C\,[101.88, \infty] = .95$$

That is, 101.88 hours is the lower (one-sided) 95% confidence limit, θ, of the true mean life where $\chi^2_{(0.05,20)}$ = 31.41 is from Table 8.3-11.

In other words, we are 95% confident that the true MTBF exceeds 101.88 hours.

(3) From Equation (8.13)

$$C\left(\frac{2T}{\chi^2_{\left(\frac{\alpha}{2},\,2r\right)}}, \frac{2T}{\chi^2_{\left(1-\frac{\alpha}{2},\,2r\right)}}\right) = C\left(\frac{3200}{34.17}, \frac{3200}{9.591}\right) = C(93.65, 333.65) = .95$$

That is, 93.65 hours is the lower (two-sided) 95% confidence limit for the mean life and 333.65 hours is the upper (two-sided) 95% confidence limit for that true mean. We are 95% confident that the interval between 93.65 and 333.65 hours contains the true MTBF.

Example 7:

Twenty items undergo a nonreplacement test, which is terminated at 100 hours. Failure times observed were 10, 16, 17, 25, 31, 46, and 65 hours. Calculate (1) the one-sided approximate 90% confidence interval (α = 0.10), and (2) the two-sided approximate 90% confidence limits of θ, the mean life.

(1) From Equations (8.12) and (8.15)

$$C = C\left(\frac{2\left[\left(\sum_{i=1}^{7} t_i\right) + (20-7)(100)\right]}{\chi^2(.10, 16)}, \infty \right)$$

$$= C\left(\frac{3020}{23.54}, \infty\right) = C\ (128.3, \infty) = .90$$

128.3 hours is the lower single-sided 90% confidence limit for θ, the true mean life.

(2) From Equation (8.13)

$$C\left(\frac{2T}{\chi^2\left(\frac{\alpha}{2},\ 2r+2\right)}, \frac{2T}{\chi^2\left(1-\frac{\alpha}{2},\ 2r\right)} \right) = C\left(\frac{3020}{26.30}, \frac{3020}{6.57}\right)$$

$$= C\ (114.83, 459.67) = .90$$

That is, 114.83 hours is the lower (two-sided) 90% confidence limit for θ, the true mean life, and 459.67 hours is the upper (two-sided) 90% confidence limit.

Table 8.3-12 presents the factor $2/\chi^2_{p,d}$ for one-sided and two-sided confidence limits, at six confidence levels for each. Multiplying the appropriate factor by the observed total life T gives a confidence limit on σ. Figure 8.3-16 presents a graphical technique for determining upper and lower confidence limits for tests truncated at a fixed time, when the number of failures is known.

TABLE 8.3-12: FACTORS FOR CALCULATION OF MEAN LIFE CONFIDENCE INTERVALS FROM TEST DATA (FACTORS = $2/\chi^2_{P,D}$)

(Assumption of Exponential Distribution)

d	Lower Limit: 99% Two-Sided / 99-1/2% One-Sided	Lower Limit: 98% One-Sided / 99% One-Sided	Lower Limit: 95% Two-Sided / 97-1/2% One-Sided	Lower Limit: 90% Two-Sided / 95% One-Sided	Lower Limit: 80% Two-Sided / 90% One-Sided	Lower Limit: 60% Two-Sided / 80% One-Sided	Upper Limit: 60% Two-Sided / 80% One-Sided	Upper Limit: 80% Two-Sided / 90% One-Sided	Upper Limit: 90% Two-Sided / 95% One-Sided	Upper Limit: 95% Two-Sided / 97-1/2% One-Sided	Upper Limit: 98% One-Sided / 99% One-Sided	Upper Limit: 99% Two-Sided / 99-1/2% One-Sided
2	.185	.217	.272	.333	.433	.619	4.47	9.462	19.388	39.58	100.0	200.0
4	.135	.151	.180	.210	.257	.334	1.21	1.882	2.826	4.102	6.667	10.00
6	.108	.119	.139	.159	.188	.234	.652	.909	1.221	1.613	2.3077	3.007
8	.0909	.100	.114	.129	.150	.181	.437	.573	0.733	.921	1.212	1.481
10	.0800	.0857	.0976	.109	.125	.149	.324	.411	.508	.600	.789	.909
12	.0702	.0759	.0856	.0952	.107	.126	.256	.317	.383	.454	.555	.645
14	.0635	.0690	.0765	.0843	.0948	.109	.211	.257	.305	.355	.431	.500
16	.0588	.0625	.0693	.0760	.0848	.0976	.179	.215	.251	.290	.345	.385
18	.0536	.0571	.0633	.0693	.0769	.0878	.156	.184	.213	.243	.286	.322
20	.0500	.0531	.0585	.0635	.0703	.0799	.137	.158	.184	.208	.242	.270
22	.0465	.0495	.0543	.0589	.0648	.0732	.123	.142	.162	.182	.208	.232
24	.0439	.0463	.0507	.0548	.0601	.0676	.111	.128	.144	.161	.185	.200
26	.0417	.0438	.0476	.0513	.0561	.0629	.101	.116	.130	.144	.164	.178
28	.0392	.0413	.0449	.0483	.0527	.0588	.0927	.106	.118	.131	.147	.161
30	.0373	.0393	.0425	.0456	.0496	.0551	.0856	.0971	.108	.119	.133	.145
32	.0355	.0374	.0404	.0433	.0469	.0519	.0795	.0899	.0997	.109	.122	.131
34	.0339	.0357	.0385	.0411	.0445	.0491	.0742	.0834	.0925	.101	.113	.122
36	.0325	.0342	.0367	.0392	.0423	.0466	.0696	.0781	.0899	.0939	.104	.111
38	.0311	.0327	.0351	.0375	.0404	.0443	.0656	.0732	.0804	.0874	.0971	.103
40	.0299	.0314	.0337	.0359	.0386	.0423	.0619	.0689	.0756	.0820	.0901	.0968

To Use: Multiply value shown by total part hours to get MTBF figures in hours

Note: d = 2r, except for the lower limit on tests truncated at a fixed time and where r < n. In such cases, use)d = 2(r + 1

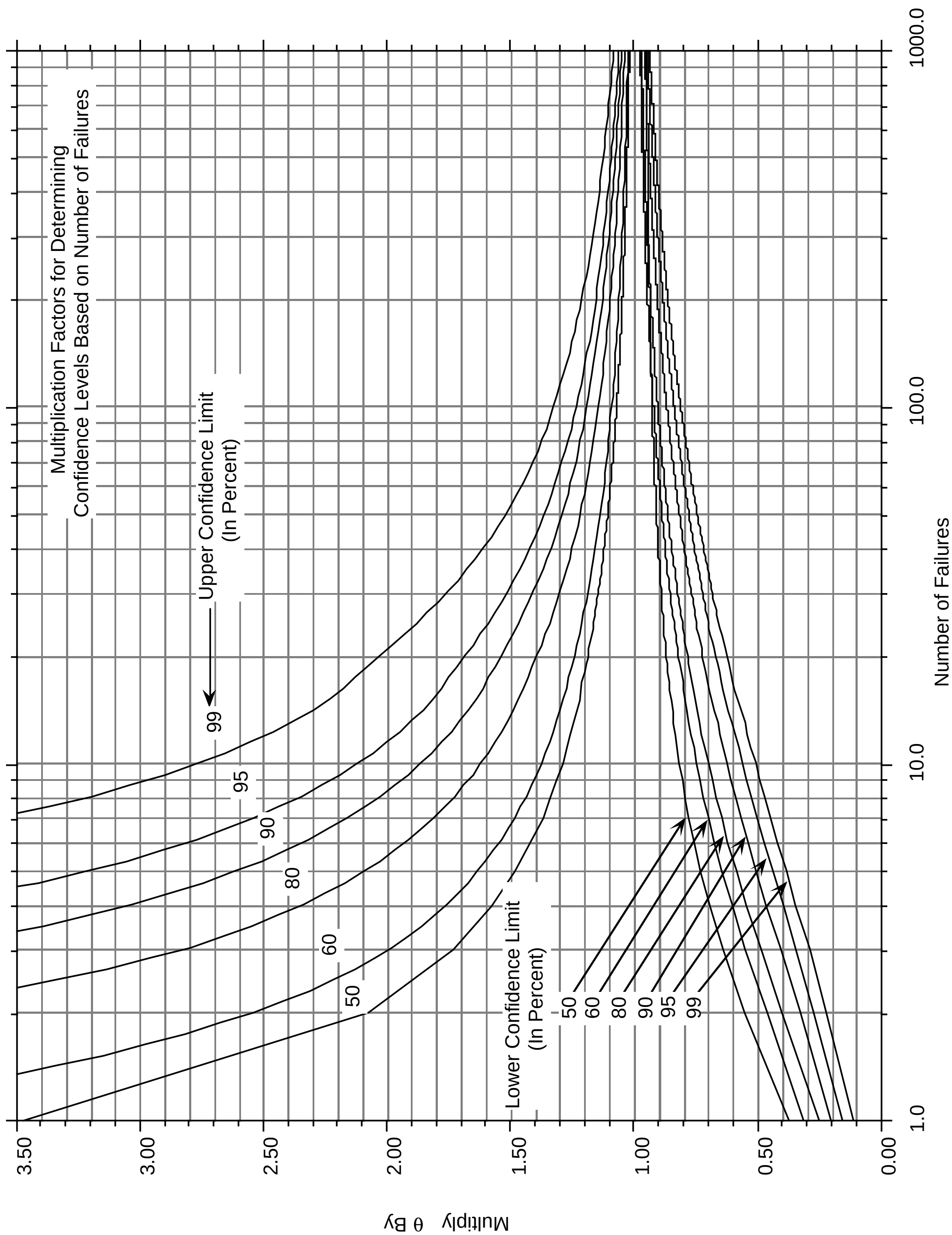


FIGURE 8.3-16: MULTIPLICATION RATIOS FOR DETERMINING UPPER AND LOWER CONFIDENCE LIMITS VS. NUMBER OF FAILURES FOR TEST TRUNCATED AT A FIXED TIME

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

Reliability Estimates (Exponential Distribution)

We know the probability of (or proportion of items) surviving t hours is found by:

$$\hat{R}(t) = e^{-t/\theta} \tag{8.18}$$

The confidence interval on R(t) is

$$C\left(e^{-t/\hat{\theta}_L} < R(t) < e^{-t/\hat{\theta}_U}\right) = 1 - \alpha$$

where:

$\hat{\theta}_L$ and $\hat{\theta}_U$ are the lower and upper confidence limits on θ.

Example 8:

Based on the data of Example 1, (1) what is the probability of an item surviving 100 hours? (2) what are the two-sided 95% confidence limits on this probability?

(1) From Equation (8.18)

$$\hat{R}(100) = e^{-100/\hat{\theta}} = e^{-100/160} = 0.535$$

(2) The two-sided confidence limits on the reliability are

$$\left(e^{-100/93.65}, e^{-100/333.65}\right) = (0.344, 0.741) = 95\%$$

8.3.2.5.3 Confidence-Interval Estimates for the Binomial Distribution

For situations where reliability is measured as a ratio of the number of successes to the total number of trials, e.g., one-shot items, missiles, etc., the confidence interval is determined by consideration of the binomial distribution. Table XI of Hald's Statistical Tables and Formulas (John Wiley and Sons, Inc., New York, 1952) and Ref. [10] gives 95% and 99% confidence limits for a wide range of values. Figure 8.3-17 allows a rough estimate to be made when the number of successes (S) and the number of trials (N) are known.

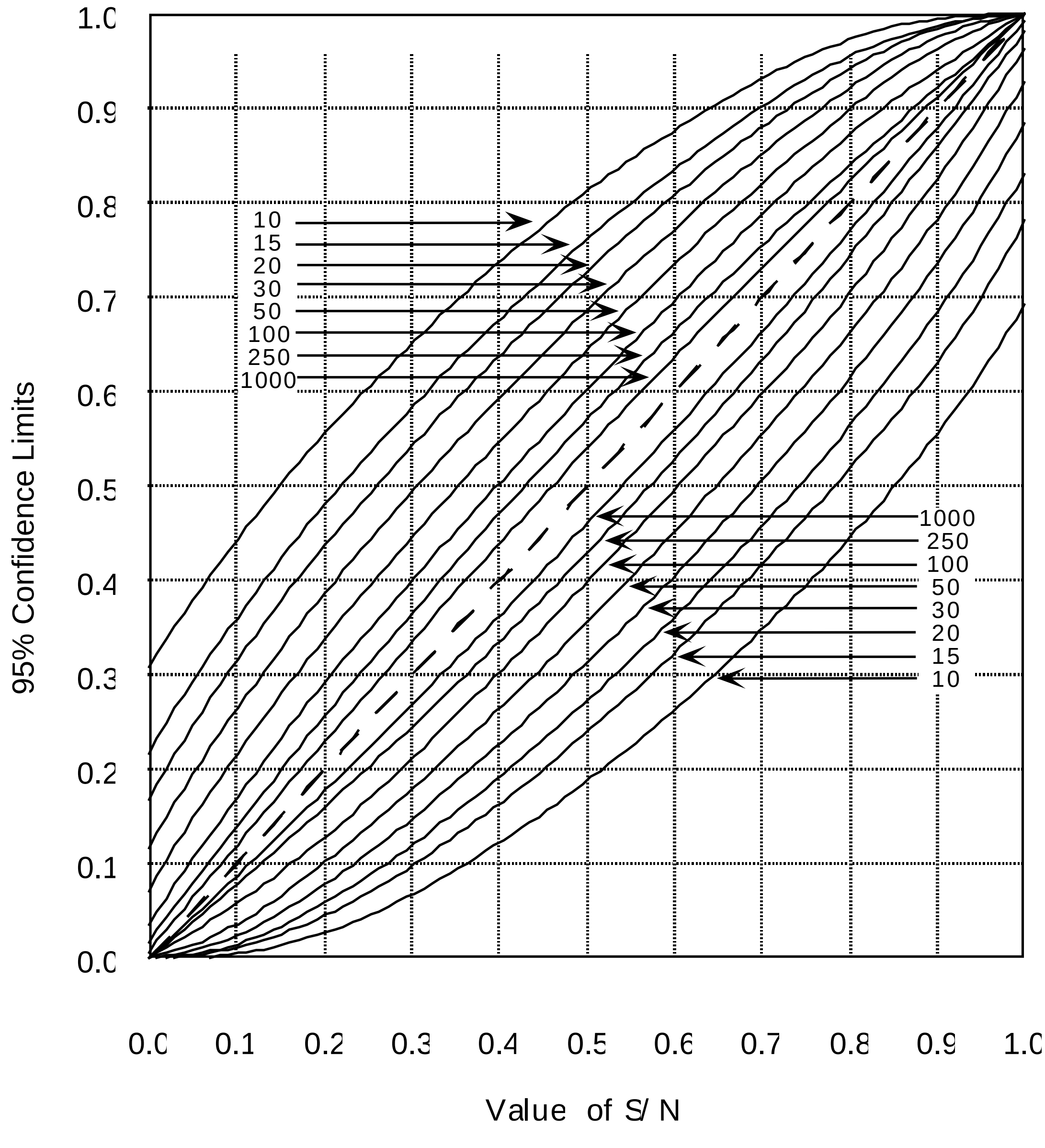


FIGURE 8.3-17: CHART FOR 95% CONFIDENCE LIMITS ON THE PROBABILITY S/N[1]

[1] From Clopper, C.J., and Pearson, E.S., "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," BIOMETRIKA, Vol. 26 (1934), p. 410. Reprinted with permission.

Example 9:

S = 8; N = 10. (a) What is the reliability estimate? (b) What are the two-sided upper and lower 95% confidence limits? Answers: (a) 0.80; (b) 0.98 and 0.43.

More detailed analyses of confidence limits and intervals, with many more examples under a variety of circumstances, and for a variety of distributions, e.g., binomial, gamma, Weibull, etc., are given in Refs. [5], [8], [9] and [10].

8.3.2.6 Tests for Validity of the Assumption Of A Theoretical Reliability Parameter Distribution

The validity of many statistical techniques used in the calculation, analysis, or prediction of reliability parameters depends on the distribution of the failure times. Many techniques are based on specific assumptions about the probability distribution and are often sensitive to departures from the assumed distributions. That is, if the actual distribution differs from that assumed, these methods sometimes yield seriously wrong results. Therefore, in order to determine whether or not certain techniques are applicable to a particular situation, some judgment must be made as to the underlying probability distribution of the failure times.

As was discussed in Section 8.3.1, some theoretical reliability functions, such as those based on the exponential, normal, lognormal, and Weibull distributions will plot as straight lines on special types of graph paper. This is the simplest procedure and should be used as a "first cut" in determining the underlying distribution. Plot the failure data on the appropriate graph paper for the assumed underlying distribution; “eyeball” it, and if it quite closely approximates a straight line, you are home free.

If it cannot be determined visually that the reliability function follows a straight line when plotted on special graph paper, then one must resort to the application of analytical “goodness-of-fit” tests.

The two goodness-of-fit tests described in this section assume a null hypothesis, i.e., the sample is from the assumed distribution. Then a statistic, evaluated from the sample data, is calculated and looked-up in a table that shows how “lucky” or “unlucky” the sample. The luck is determined by the size of the two-sided tail area. If that tail is very small (you were very unlucky if the null hypothesis is true), the null hypothesis (there is no difference between the actual and the assumed distributions) is rejected. Otherwise, the null hypothesis is accepted, i.e., the actual distribution could easily have generated that set of data (within the range of the data); the test says nothing about the behavior of the distribution outside the range of the data.

Goodness-of-fit tests are statistical tests, not engineering tests. No matter what the distribution or what the test, it is possible to take a sample small enough so that virtually no distribution will be rejected, or large enough so that virtually every distribution will be rejected.

Thus, while a method for small sample sizes is presented as well as one for large sample sizes, it is a fact of life that must be accepted that tests based on small samples are simply not very powerful (power is the probability of rejecting the null hypothesis where it, indeed, is incorrect). Therefore, the methodology is presented here for completeness, but very likely a more logical approach is to first make an assumption regarding the failure distribution based on engineering judgment or on historical data or on knowledge of the failure characteristics of similar parts. Once the failure distribution has been assumed the test can be performed for goodness-of-fit for that particular distribution. If the hypothesized distribution is shown not to fit, it is quite certain that the assumed distribution was not the one from which the samples were selected. If, however, the goodness-of-fit test shows that the data could have come from the hypothesized distribution, then it is virtually certain that tests for fit to other distributions would yield like results.

In summary then, it must be realized that the tests presented in the next two sections have limitations. The only cure for these limitations is a larger number of observations. If this proves uneconomical or not feasible from the standpoint of the test time required to generate the desired number of failures or the cost of testing , or some other practical constraint, then the only alternative is to use the results of small sample size analyses with proper discretion.

8.3.2.6.1 Kolmogorov-Smirnov (K-S) Goodness-of-Fit Test (also called “d” test)

This test is based upon the fact that the observed cumulative distribution of a sample is expected to be fairly close to the true cumulative distribution. The goodness-of-fit is measured by finding the point at which the sample and the population are farthest apart and comparing this distance with the entry in a table of critical values, Table 8.3-13, which will then indicate whether such a large distance is likely to occur. If the distance is too large, the chance that the observations actually come from a population with the specified distribution is very small. This is evidence that the specified distribution is not the correct one.

1. When to Use

When failure times from a sample have been observed and it is desired to determine the underlying distribution of failure times.

2. Conditions for Use

(a) Usually historical data or engineering judgment suggest that item failure times of interest are from a given statistical failure distribution. This test then follows the step of assuming a given failure distribution and is useful to determine if empirical data disprove this hypothesis.

(b) The Kolmogorov-Smirnov test for goodness-of-fit is distribution free and can therefore be used regardless of the failure distribution that the data are assumed to follow.

(c) The discriminating ability of the statistical test is dependent on sample size; the larger the sample size, the more reliable the results. When large sample sizes are available, the χ^2 Test for Goodness-of- Fit is more powerful but requires additional manipulation of the data. Where sample sizes are small, the Kolmogorov-Smirnov test provides limited information but is a better choice than the χ^2 alternative.

(d) Strictly speaking, this test method requires prior knowledge of the parameters. If the parameters are estimated from the sample the exact error risks are unknown.

(e) A Kolmogorov-Smirnov table is required (see Table 8.3-13).

TABLE 8.3-13: CRITICAL VALUES $d_{\alpha;n}$ OF THE MAXIMUM ABSOLUTE DIFFERENCE BETWEEN SAMPLE AND POPULATION RELIABILITY FUNCTIONS

Sample Size, N	Level of Significance, α				
	0.20	0.15	0.10	0.05	0.01
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.474	0.565	0.669
10	0.322	0.342	0.368	0.410	0.490
15	0.266	0.283	0.304	0.338	0.404
20	0.231	0.246	0.264	0.294	0.356
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
40	0.17	0.18	0.19	0.21	0.25
45	0.16	0.17	0.18	0.20	0.24
50	0.15	0.16	0.17	0.19	0.23
over 50	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

3. Graphic Method (Example Using Exponential Distribution)

Forty-eight samples of an equipment's time-to-failure are acquired. Based upon the assumption of an exponential distribution of time-to-failure, the point estimate of MTBF is calculated to be 1546 hours.

We would like to test the hypothesis that the sample came from a population where time-to-failure followed an exponential distribution with an MTBF of 1546 hours (see Figure 8.3-18).

(a) Draw the curve (dashed line) for the theoretical distribution of R(t) which is assumed to be an exponential with an MTBF = 1546 hours.

(b) Find the value, d, using Table 8.3-13 which corresponds to sample size, n = 48, and level of significance α = 0.05: d = $(1.36/\sqrt{48}$ = 0.196).

(c) Draw curves at a distance d = 0.196 above and below the theoretical curve drawn in step (a), providing upper and lower boundaries as shown in Figure 8.3-18.

(d) On the same graph draw the observed cumulative function (solid line).

(e) If the observed function falls outside the confidence band drawn in step (c), there would be a five percent chance that the sample came from an exponential population with a mean life of 1546 hours.

(f) If the observed function remains inside the band, as it does in the example, this does not prove that the assumed distribution is exactly right, but only that it might be correct and that it is not unreasonable to assume that it is.

This example could have also been solved analytically by calculating the difference between the theoretical cumulative distribution function (CDF) and the actual CDF at each data point, finding the maximum deviation and comparing it with the value derived from Table 8.3-13 (d = 0.196). If the maximum deviation is less than 0.196, we accept the hypothesis (at the .05 significance level) that the time to failure is exponentially distributed with an MTBF of 1546 hours.

4. Analytical Method	Example (Weibull Distribution)
a. Observe and record part failure times	a. Given the following 20 failure times in hours 92 640 130 700 233 710 260 770 320 830 325 1010 420 1020 430 1280 465 1330 518 1690
b. Assume a distribution of failure times based on historical information or on engineering judgment	b. Assume failure times are distributed according to the two-parameter Weibull distribution.
c. Estimate the parameters of the assumed distribution from the observed data.	c. By the graphic method or the method of least squares, find the Weibull parameters. The Weibull shape parameter β equals 1.50 and the Weibull scale parameter α equals 28400.

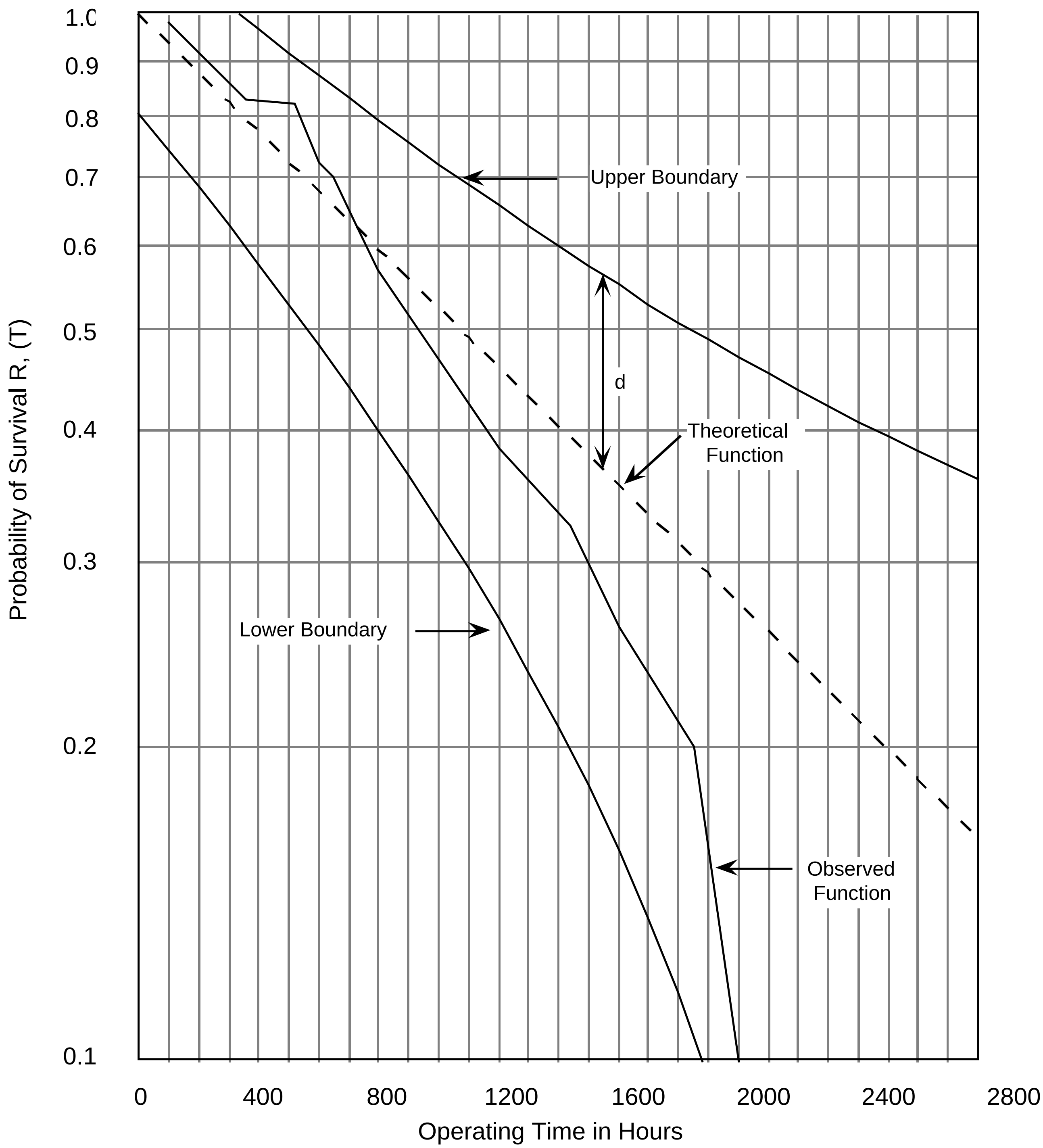


FIGURE 8.3-18: EXAMPLE OF THE APPLICATION OF THE "d" TEST

d. Calculate the probability of failure for each observation from the cumulative failure function for the assumed distribution.

d. For the Weibull distribution the cumulative failure function is

$$\hat{F}(X) = 1 - \exp\left(-\frac{X^{\beta}}{\alpha}\right)$$

where X = observed failure time,
β = 1.5 = Weibull shape parameter,
α = 28400 = Weibull scale parameter, $\hat{F}(X)$ = probability of failure at or before time X.

For the 20 observations of this example, the probability of failure at the respective times is:

X	$\hat{F}(X)$
92	.03
130	.05
233	.12
260	.14
320	.18
325	.19
420	.26
430	.27
465	.30
518	.34
640	.43
700	.48
710	.49
770	.53
830	.57
1010	.68
1020	.68
1280	.80
1330	.82
1690	.91

e. Calculate the percentile for each of (i) failure times by the relationship $F(i) = \frac{i}{n+1}$. Subtract those of Step d. above. Record the absolute value of the difference.

e. For n = 20, $\frac{i}{n+1}$ gives the following results:

$\hat{F}(x)$	$F(i)$	$\lvert \hat{F}(x)-F(i) \rvert$
.03	.05	.02
.05	.10	.05
.12	.14	.02
.14	.19	.05
.18	.24	.06
.19	.29	.10
.26	.33	.07
.27	.38	.11
.30	.43	.13
.34	.48	.14
.43	.52	.09
.48	.57	.09
.49	.62	.13
.53	.67	.14
.57	.71	.14
.68	.76	.08
.68	.81	.13
.80	.86	.06
.82	.90	.08
.91	.95	.04

f. Compare the largest difference from step e with a value at the desired significance level in the Kolmogorov-Smirnov tables to test for goodness-of-fit. If the tabled value is not exceeded then it is not possible to reject the hypothesis that the failure times are from the assumed distribution.

f. The largest difference in Step e. was .14. From the Kolmogorov-Smirnov table for a significance of .05 and for a sample of size 20 a difference of greater than .294 must be observed before it can be said that the data could not have come from a Weibull distribution with $\beta = 1.5$, $\alpha = 28400$.

8.3.2.6.2 Chi-Square Goodness-of-Fit Test

The chi-square goodness-of-fit test may be used to test the validity of any assumed distribution, discrete or continuous. The test may be summarized as follows for a continuous distribution.

(a) Determine the underlying distribution to be tested.

(b) Determine a level of significance, α, which is defined as the risk of rejecting the underlying distribution if it is, in fact, the real distribution.

(c) Divide the continuous scale into k intervals. For reliability analysis, this scale is usually time.

(d) Determine the number of sample observations falling within each interval.

(e) Using the assumed underlying distribution, determine the expected number of observations in each interval. Combining of intervals may be required because the expected number of observations in an interval must be at least 5.0. This determination may require an estimation of the distribution parameters from the sample data (w is the number of estimated parameters).

(f) Compute

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i} \tag{8.19}$$

where:

O_i = number of sample observations in the i^{th} interval

E_i = expected number of observations in the i^{th} interval

k = number of intervals

(g) Let w be the number of parameters estimated from the data and let $\chi^2_{\alpha,k-w-1}$ be the value found in Table 8.3-11.

(h) Compare the calculated χ^2 statistic with the tabled χ^2 value for the discrete level of the signature

$$\text{If } \chi^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i} > \chi^2_{\alpha,k-w-1} \tag{8.20}$$

reject the distribution under test. Otherwise, we do not have sufficient evidence to reject the assumed underlying distribution.

1. When to Use

When failure times are available from a relatively large sample and it is desired to determine the underlying distribution of failure times.

2. Conditions for Use

(a) In the statistical analysis of failure data it is common practice to assume that failure times follow a given failure distribution family. This assumption can be based on historical data or on engineering judgment. This test for goodness-of-fit is used to determine if the empirical data disproves the hypothesis of fit to the assumed distribution.

(b) The χ^2 test for goodness-of-fit is "distribution-free" and can therefore be used regardless of the failure distribution that the data are assumed to follow.

(c) This test is not directly dependent on sample size but on the number of intervals into which the scale of failure times is divided with the restriction that no interval should be so narrow that there are not at least 5 theoretical failures within the interval. Therefore, the test is only useful if a relatively large number of failures has been observed.

(d) A table of χ^2 percentage points is required (see Table 8.3-12).

3. Method (Example Using Exponential Distribution)

Consider the data in Figure 8.3-19 indicating the failure times obtained from testing a sample of 100 fuel systems. Using a significance level of $\alpha = 0.05$, test whether the assumption of an exponential distribution is reasonable. The sample mean was found to be 8.9 hours.

(a) Figure 8.3-20 is used as a means of computing

$$\sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

(b) The expected frequency, E_i, is found by multiplying the sample size by the probability of falling within the i^{th} interval if the assumed (exponential) distribution is true.

$$E_i = n\left[\exp\left(\frac{-L_i}{\hat{\theta}}\right) - \exp\left(\frac{-U_i}{\hat{\theta}}\right)\right]$$

$$= 100\left[\exp\left(\frac{-L_i}{8.9}\right) - \exp\left(\frac{-U_i}{8.9}\right)\right]$$

Interval (Hours)	Frequency
0 - 5.05	48
5.05 - 10.05	22
10.05 - 15.05	11
15.05 - 20.05	7
20.05 - 25.05	3
25.05 - 30.05	5
30.05 - 35.05	2
35.05 - 40.05	0
40.05 - 45.05	1
45.05 - 50.05	0
50.05 - 55.05	1
	100

FIGURE 8.3-19: FUEL SYSTEM FAILURE TIMES

Interval (hours) $(L_i - U_i)$	Observed Frequency (O_i)	Expected Frequency (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
0 - 5.05	48	43	5	25	.58
5.05 - 10.05	22	24	-2	4	.17
10.05 - 15.05	11	14	-3	9	.64
15.05 - 20.05	7	8	-1	1	.13
20.05 - 25.05	3	5	-2	4	.80
25.05 - 30.05	5	3	2	4	1.33
30.05 - 35.05	2				
35.05 - 40.05	0				
40.05 - 45.05	1 } 4	3	1	1	.33
45.05 - 50.05	0				
50.05 - 55.05	1				3.98

FIGURE 8.3-20: COMPUTATION

where U_i and L_i are the upper and lower limits of the i^{th} interval, $U_i = L_i + 5$, and $\theta = 8.9$ hours.

(c) Some of the original intervals were combined to satisfy the requirement that no E_i value be less than 2.5.

$$\chi^2 = \sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i} = 3.98$$

$$\chi^2_{\alpha, k-w-1} = \chi^2_{.05,7-1-1} = \chi^2_{0.5,5} = 11.070$$

(See Table 8.3-11)

Since $\chi^2 = \sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i} = 3.97 < \chi^2_{0.5,5} = 11.070,$

we do not have sufficient evidence to reject the exponential distribution as a model for these failure times.

4. Method	Example (Weibull Distribution)
a. Observe and record part failure times.	a. The following is the number of cycles to failure for a group of 50 relays on a life test:

1283	6820	16306
1887	7733	17621
1888	8025	17807
2357	8185	20747
3137	8559	21990
3606	8843	23449
3752	9305	28946
3914	9460	29254
4394	9595	30822
4398	10247	38319
4865	11492	41554
5147	12913	42870
5350	12937	62690
5353	13210	63910
5410	14833	68888
5536	14840	73473
6499	14988	

b. Assume a distribution of failure times based on historical information or on engineering judgment.	b. Assume failure times are distributed according to the two-parameter Weibull distribution.
c. Estimate the parameters of the assumed distribution from the observed data.	c. By the graphical method or method of least squares find the Weibull parameters. The Weibull shape parameter β=1.21 and the Weibull scale parameter α =127978.

d. Divide the spectrum of failure times into intervals of such a width that the theoretical number of failures in each interval will be at least five. The width of intervals need not be equal but extra care must be used in determining the expected frequencies in this case.

d. Divide the relay cycles-to-failure into the following intervals:

0 - 4000
4001 - 7200
7201 - 13000
13001 - 18000
18001 - 25000
25001 - above

e. Calculate the theoretical number of failures for each interval.

e. The expected number of failures in each interval is obtained as follows:

For the Weibull distribution the cumulative failure function is

$$F(X) = 1 - \exp\left(-\frac{\chi^{\beta}}{\alpha}\right)$$

where: X = observed failure times
β = Weibull shape parameter
α = Weibull scale parameter

Then $F(X_n) - F(X_{n-1})$ = probability that a failure time falls within the interval. Then for each interval the probability of failure in that interval, multiplied by the sample size, equals the theoretical number of failures for each interval.

(1) Upper Boundary of Interval	(2) F(X)	(3) $F(X_n) - F(X_{(n-1)})$	(4) Theoretical Failure Frequency (Col. 3x50) E_i
4000	.16	.16	8
7200	.30	.14	7
13000	.52	.22	11
18000	.66	.14	7
25000	.80	.14	7
∞	1.00	.20	10

NOTE: The theoretical frequency must not be less than 5 for any interval.

f. Calculate the χ^2 statistic by the formula

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where: k = number of intervals
O_i = observed frequency interval
E_i = theoretical frequency per interval

Upper Boundary of Interval	E_i	O_i	$\frac{(O_i - E_i)^2}{E_i}$
4000	8	8	0
7200	7	10	1.29
13000	11	12	.09
18000	7	7	0
25000	7	3	2.29
∞	10	10	0
	50	50	χ^2 = 3.67

g. Determine if the χ^2 statistic indicates that the data could have come from the hypothesized distributions using χ^2 tables (Table 8.3-11) and (k-1) - p degrees of freedom.

where: k = number of intervals
p = number of parameters estimated

g. The degrees of freedom for this example are calculated as:

d.f. = (k-1) - p
d.f. = (6-1) - 2 = 3

The value from the χ^2 table for 3 degrees of freedom at the 0.05 level of significance is 7.815. Since 3.69 does not exceed the tabled value, then the hypothesis that this data came from a Weibull distribution cannot be rejected.

8.3.2.6.3 Comparison of K-S and Chi-Square Goodness-of-Fit Tests

The K-S test is superior to χ^2 in the following ways:

(1) The K-S Test can be used to test for deviations in a given direction, while Chi-Square Test can by used only for a two-sided test.

(2) The K-S Test uses ungrouped data so that every observation represents a point of comparison, while the Chi-Square Test requires the data to be grouped into cells with arbitrary choice of interval, size, and selection in starting point. Minimum expected frequency values are required.

(3) The K-S Test can be used in a sequential test where data become available from smallest to largest, computations being continued only up to the point at which rejection occurs.

The Chi-Square Test is superior to the K-S Test in the following ways:

(1) Chi-square can be partitioned and added

(2) Chi-square can be applied to discrete populations

8.4 Reliability Demonstration

8.4.1 Introduction

The single purpose of a reliability demonstration test is to determine conformance to specified, quantitative reliability requirements as a basis for qualification or acceptance; this is to answer the question, "Does the item meet or exceed (not by how much) the specified minimum reliability requirement?"

Reliability testing involves an empirical measurement of time-to-failure during equipment operation for the purpose of determining whether an equipment meets the established reliability requirements. A reliability test is effectively a "sampling" test in the sense that it is a test involving a sample of objects selected from a population. In reliability testing, the population being measured encompasses all failures that will occur during the life span of the equipment. A test sample is drawn from this population by observing those failures occurring during a small portion of the equipment's life. In reliability testing, as in any sampling test, the sample is assumed to be representative of the population, and the mean value of the various elements of the sample (e.g., times-to-failure) is assumed to be a measure of the true mean (MTBF, etc.) of the population.

A sample in a reliability test consists of a number of times-to-failure, and the population is all the times-to-failure that could occur either from the one equipment or the more than one equipment on test. The "test" equipments (assuming more than one equipment) are considered identical and, thus, their populations are also identical. Under the assumption of an exponential failure model (***constant*** λ), a test of 10 devices for 100 hours each is mathematically equivalent to a test of 1 device for 1000 hours. If all possible samples of the same number of times-to-failure were drawn from the same or identical equipment, the resulting set of sample means would be distributed about the true MTBF (θ) of the equipment, following a normal distribution as is shown in Figure 8.4-1.

Since it is not economically feasible to test the complete population, we have to be satisfied with a sample of the population. From the data in the sample we then make some statement about the population parameter.

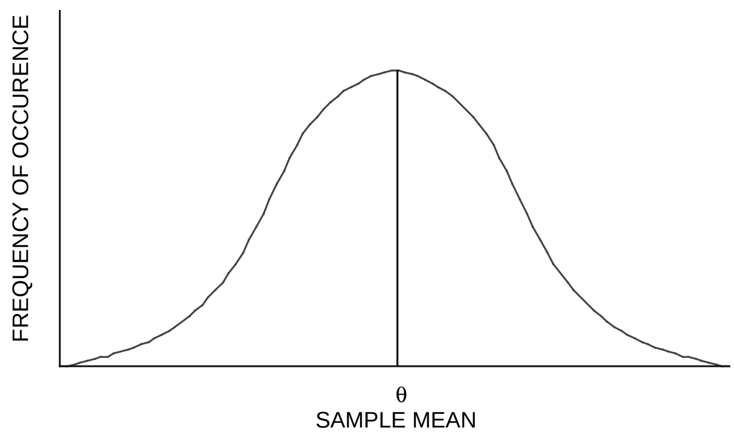


FIGURE 8.4-1: NORMAL DISTRIBUTION

What we are doing is testing a statistical hypothesis: For example, we might test

H_0: (null hypothesis) $\theta_0 \geq$ 200 hours

H_1: (alternate hypothesis) $\theta_1 \leq$ 100 hours

Based upon the test results, we either accept H_0 or reject it. In making our decision we have to keep several risks in mind.

Producer's risk (α) is the probability of rejecting H_0 when it is true (probability of rejecting good equipment)

Consumer's risk (β) is the probability of accepting H_0 when it is false (probability of accepting bad equipment)

Looking at it another way, if θ_0 and θ_1 represent the hypotheses, then the α and β errors are the hatched areas shown in Figure 8.4-2A. Of course, if we could take enough samples, then the standard deviation about each of the means would be reduced and the α and β errors would also be reduced.

However, this is usually impractical so the sample size is set as low as possible to reduce costs by specifying the maximum acceptable α and β risks that can be associated with θ_0 and the smallest acceptable θ_1. Why two values? Let's look at our decision rule, or accept/reject criteria. We would like it to look like Figure 8.4-3A.

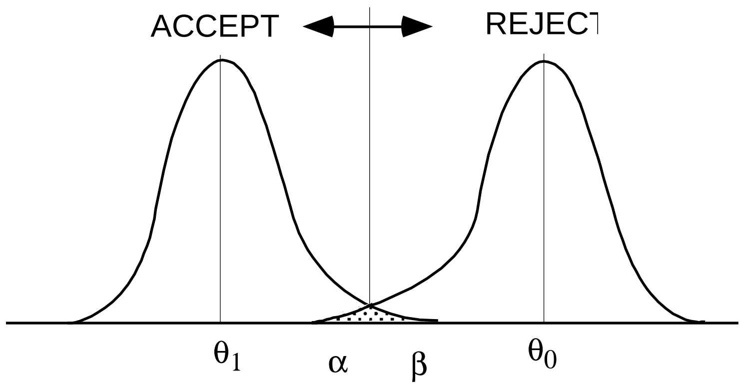


FIGURE 8.4-2A: HYPOTHESIS TEST A

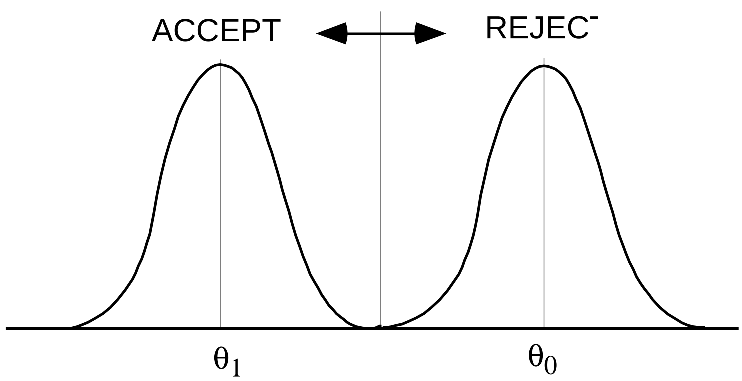


FIGURE 8.4-2B: HYPOTHESIS TEST B

This relationship between the probability of acceptance and the requirement (e.g. MTBF) is called the ***operating characteristic curve***. The ideal curve shown in Figure 8.4-2B would require an infinite number of samples. In real life we settle for something that gives a small probability of acceptance (P_A) for MTBF's below the requirement and high P_A for MTBF's above the requirement, M_0.

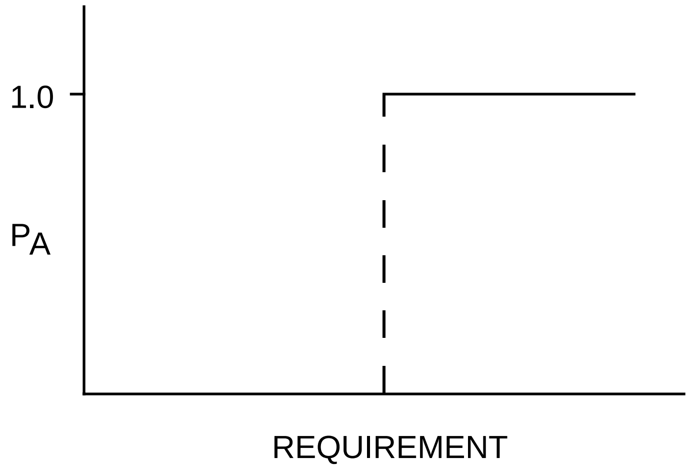


FIGURE 8.4-3A: IDEAL OPERATING CHARACTERISTIC (OC) CURVE

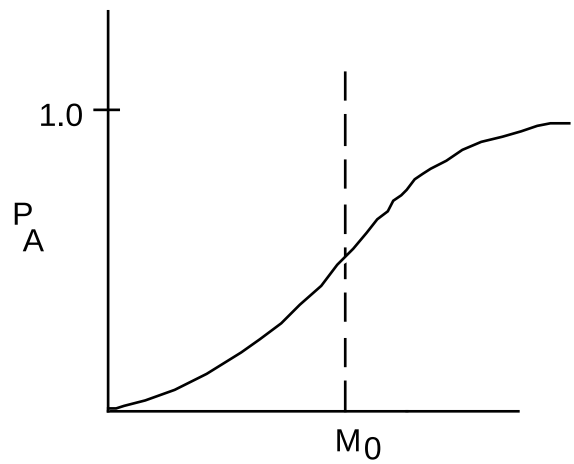


FIGURE 8.4-3B: TYPICAL OPERATING CHARACTERISTIC CURVE

For example, suppose we had an MTBF requirement of 200 hours, a demonstration test of 1000 hours, and the decision rule,

Accept H_0 if $r \leq 5$

Reject H_0 if $r > 5$

where r is the number of failures which is Poisson distributed (fixed time test) as

$$P(r) = P_R = \frac{(t/m)^r e^{-t/m}}{r!} \tag{8.21}$$

where m is the MTBF.

We plot P_A $(r \leq 5)$ for various values of m based upon the expected number of failures, as shown in Figure 8.4-4.

MTBF	t/m	$P_A(r \leq 5)$
100	10	0.067
125	8	0.191
167	6	0.446
200	5	0.616
333	3	0.916
500	2	0.983

$$PA\ (r \leq s) = \sum_{r=0}^{5} \frac{(t/n)^r e^{-t/n}}{r!}$$

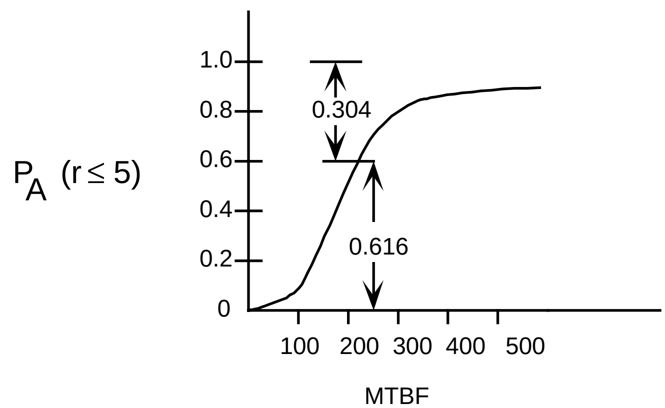


FIGURE 8.4-4: ACTUAL OPERATING CHARACTERISTIC CURVE

The decision rule "tends" to give the right decision, but won't always result in an accept decision for m > 200 or a reject decision for m < 200. Remember $P_A + P_R = 1$. Thus, we can see that we have almost a fifty-fifty chance of accepting an m of 167 hours (0.446) and a greater than 20% chance of rejecting an m = 250 hours. Neither the producer or consumer would be happy with this. Each would like a lower risk probability. But since $P_A = 1 - P_R$, if we lower P_A for $m \leq 200$ to 0.1, we raise P_R for m > 200 to 1 - 0.1 = 0.9. What do we do now?

In order to overcome this difficulty it is necessary to specify the reliability requirements, either explicitly or implicitly, in terms of two MTBF values rather than a single MTBF value. The lower value is defined as the lower test MTBF (M_m or θ_1) and the higher value is defined as the upper test MTBF (M_R or θ_0). The test plan can then be designed to give a low probability of an ***accept decision*** for equipment with an MTBF of $m < M_m$ (or θ_1) and a low probability of ***reject decision*** when $m > M_R$. P_A at $m = M_m$ (or θ_1) is the ***consumers risk*** (β); P_R at $m = M_R$ (or θ_0) is the ***producer's risk*** (α). Thus, specifying the two MTBF values $M_m(\theta_1)$ and $M_R(\theta_0)$ and the two risks (α and β) defines two points on the OC curve as shown in Figure 8.4-5.

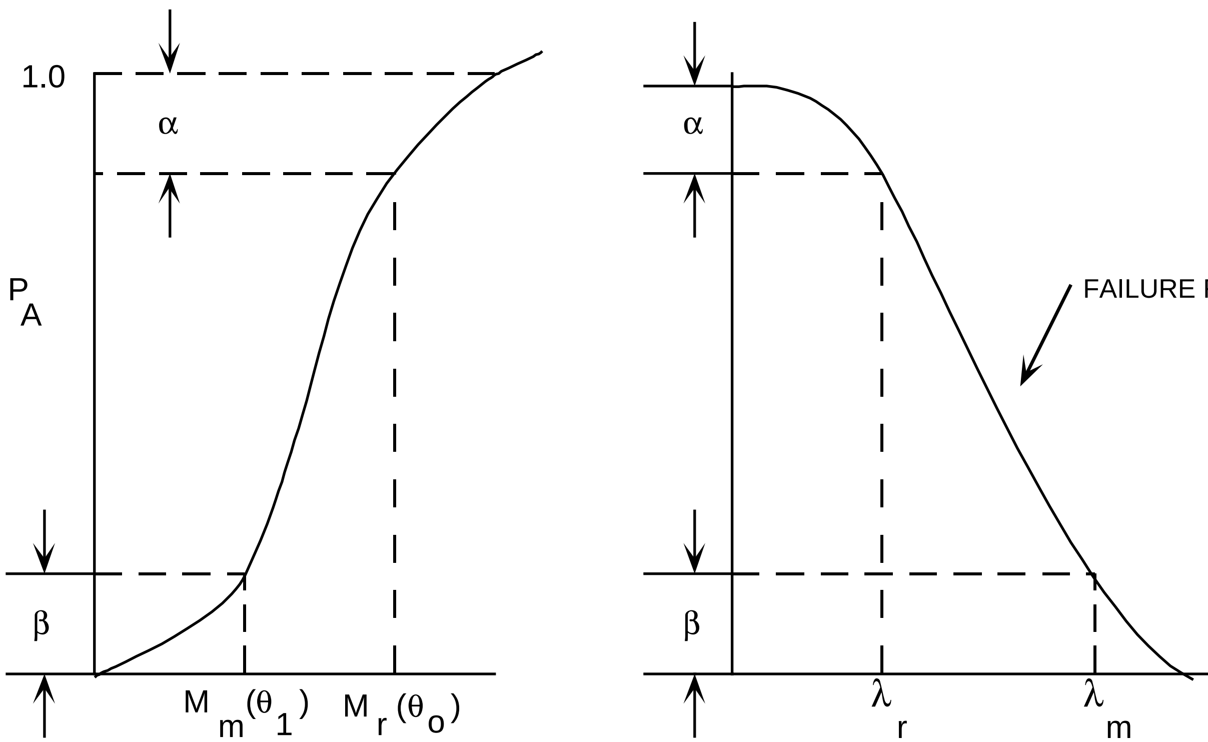


FIGURE 8.4-5: OC CURVE CHARACTERISTICS

The curve on the right is the OC curve for failure rate (α) rather than for MTBF. $\lambda_m = 1/M_m$ is the ***maximum acceptable*** failure rate. $\lambda_R = 1/M_R$ is the ***design-required*** (specified) failure rate with $\lambda_R < \lambda_m$.

The method used to design a ***fixed time*** reliability (R) demonstration test is mathematically equivalent to the method used to construct confidence limits for MTBF. Therefore, if a fixed time R demonstration involving a test time T and an accept number 0 provides a consumer risk of β with respect to a minimum acceptable MTBF (M_m or θ_1), it will be found that if the maximum allowable number of failures, r_0, actually occurs, the lower 100(1-β)% confidence limit for MTBF as calculated from the test data is exactly M_m. For this reason, the value (1-β), or 100(1-β)% is often called the ***confidence level*** of the demonstration test. Thus, a fixed time R demonstration test providing a 10% consumer risk is called "a demonstration test at a 90% confidence level," or is said to "demonstrate with 90% confidence that the lower test MTBF is achieved." This is not really correct since, technically, confidence level is used in the estimation of a parameter while an R demonstration test is testing a hypothesis about the parameter, m, rather than constructing an interval estimate for m.

There are six characteristics of any reliability demonstration test that must be specified:

(1) The reliability deemed to be acceptable, R_0, "upper test MTBF"

(2) A value of reliability deemed to be unacceptable, R_1, "lower test MTBF"

(3) Producer's risk, or α

(4) Consumer's risk, or β

(5) The probability distribution to be used for number of failures or for time-to-failure

(6) The sampling scheme

Another term frequently used in connection with reliability demonstration tests should be defined here although it is derived from two of the six characteristics. The discrimination ratio is the ratio of upper test reliability to the lower test reliability. R_0/R_1 is an additional method of specifying certain test plans.

There are, of course, an infinite number of possible values for the actual reliability. In the specification of two numerical values, R_0 and R_1, the experimenter achieves the producer's risk, α, and consumer's risk, β, only for those specific reliabilities.

For other values, the relationship is:

(a)	Probability of Acceptance	≥	1-α	for R	≥	R_0
(b)	Probability of Acceptance	≤	β	for R	≤	R_1
(c)	Probability of Acceptance	>	β	for R_1	≤	$R \le R_0$

8.4.2 Attributes and Variables

Demonstration tests are classified according to the method of assessing reliability. If each component tested is merely classified as acceptable or unacceptable, then the demonstration test is an **attributes** test. If the service life of the items under test is recorded in time units, and service life is assumed to have a specific probability distribution such as the normal or Weibull, then the test is a **variables** test. Attributes tests may be performed even if a probability distribution such as the normal or Weibull is assumed by dichotomizing the life distribution into acceptable and unacceptable time-to-failure. Attributes tests are usually simpler and cheaper to perform, but require larger sample sizes to achieve the same α and β as variables tests.

8.4.3 Fixed Sample and Sequential Tests

When R_0, R_1, α, and β have been specified, along with the probability distribution for time to failure, the test designer often has a choice of sampling schemes. To achieve the desired α and β, statistical theory will dictate the precise number of items which must be tested if a fixed sample size is desired. Alternatively, a sequential test may be selected, where the conclusion to accept or reject will be reached after an indeterminate number of observations. For reliability near R_0 or R_1, the average sample size in a sequential test will invariably be lower than in a fixed sample test, but the sample size will be unknown, and could be substantially larger in a specific case. Usually, an upper bound for sample size is known in sequential tests.

8.4.4 Determinants of Sample Size

Whether a fixed sample or sequential test is selected, the number of observations required will be related to the degree of discrimination asked for. In general,

(a) The closer R_1 is to R_0, the larger the sample size required

(b) The smaller the α specified, the larger the sample size required

(c) The smaller the β specified, the larger the sample size required

If the test is sequential, substitute "average sample size" for sample size in the above remarks.

8.4.5 Tests Designed Around Sample Size

It is possible to set the sample size (or average sample size in sequential tests) independently. For example, the sample size, N, may be limited by test facilities, cost, or time. If this is done, then one cannot specify all of the values R_0, R_1, α, and β. One of the four will be fixed when the remaining three and N are specified. The usual practice where N must be fixed is to specify R_0 and β and then to include a plot of 1 - β as a function of R_1, the corresponding probability of rejection, 1 - β. If the discriminating power is unacceptable, then R_1, α, β, or N must be altered in the direction noted in Section 8.4.4.

8.4.6 Parameterization of Reliability

In the case of variables tests, the desired reliability will be a function of the parameters of whatever probability distribution is selected. For example, if equipment mean life is normally distributed, then

$$R = \int_{T}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right] dx \qquad (8.22)$$

where:

- T = desired life
- μ = population mean
- σ = population standard deviation

Suppose that R_0 is specified at 0.995 for a service life, T, of 10,000 hours. Clearly, these specifications place numerical requirements on μ and σ to make the equation true. Therefore, the demonstration test may be performed on μ_0, (σ_0), rather than on R_0. Demonstration tests are often specified in terms of the probability distribution parameters, rather than reliabilities.

8.4.7 Instructions on the Use of Reliability Demonstration Test Plans

Instructions and examples are given for the following test plans:

(1) Attributes Demonstration Tests

- (a) Plans for Small Lots
- (b) Plans for Large Lots
- (c) Plans for Large Lots (Poisson Approximation Method)
- (d) Attributes Sampling Using ANSI/ASQC Z1.4-1993
- (e) Sequential Binomial Test Plans

(2) Variables Demonstration Tests

(a) Time Truncated Test Plans
(1) Exponential Distribution
(2) Normal Distribution
(3) Weibull Distribution

(b) Failure Truncated Tests
(1) Exponential Distribution
(2) Normal Distribution (Known)
(3) Normal Distribution (Unknown)
(4) Weibull Distribution

(c) Sequential Tests
(1) Exponential Distribution
(2) Normal Distribution

(d) Interference Demonstration Tests

(e) Bayes Sequential Tests

8.4.7.1 Attributes Demonstration Tests

8.4.7.1.1 Attributes Plans for Small Lots

1. When to Use

When testing items from a small lot where the accept/reject decision is based on attributes, the hypergeometric distribution is applicable. Attributes tests should be used when the accept/reject criterion is a go-no-go situation, when the probability distribution of times to failure is unknown, or when variables tests are found to be too expensive. The example demonstrating the method is based on a small lot and small sample size. This situation frequently characterizes the demonstration test problem associated with large systems. The sample size limits the discriminatory power of the demonstration test plan but frequently cost and time constraints force us into-larger-than desired risks.

2. Conditions for Use

The definition of successfully passing the test may be that an item survives the test. The parameter to be evaluated then is the fraction of the items in the lot that survive. The estimation of the parameter would be based on a fixed sample size and testing without repair. The selection of the criteria for success (survive, detonate on impact, time) can be derived from a requirement or, if the items being tested are known to follow a particular probability distribution, the specification of the criteria for success can be based on defining acceptable and unacceptable portions of the range of failures. If the lot size is large, say 30 or more, then the Poisson approximation may be used to make the calculation simpler.

3.

Method	Example
a. Define criterion for success/failure.	a. A missile that seeks and destroys the target. Missiles that fail to destroy the target are considered failures.
b. Define acceptable lot quality level $(1 - p_0)$.	b. Lots in which $(1 - p_0) = 90\%$ of the missiles will destroy the target are to be accepted by this demonstration test plan with high probability.
c. Specify producer's risk (α), i.e., the probability that acceptable lots will be rejected.	c. Let $\alpha = .2$. This decision is an engineering one based on the consequences of allowing good lots to be rejected and based on the time and dollar constraints associated with inspecting the lot.

3. Method	Example
d. Define unacceptable quality level $(1 - p_1)$.	d. Lots in which only $(1 - p_1) = 20\%$ of the missiles destroy the target will be accepted by the demonstrations test plan with low probability.
e. Specify the consumer's risk (β) (i.e., the probability that unacceptable quality lots will pass the demonstration test).	e. Let $\beta = .022$ (taken for convenience in calculations).
f. Now that α, β, $1 - p_0$, and $1 - p_1$ have been specified the following steps describe the calculations required to determine the sample size and accept/reject criteria which will satisfy the stated risks.	f. Given: lot size N = 10 $1 - p_0 = .9$ $1 - p_1 = .2$ $\alpha = .2$ $\beta = .022$
g. The process consists of a trial and error solution of the hyper-geometric equation using N, $1 - p_0$, $1 - p_1$ and various sample sizes until the conditions of α and β are met. The equation used is $$\Pr(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$ $x = 0, 1, 2 \ldots \min(n,r)$ where: x = number of successes in sample	g. The calculations are as follows: If N = 10 and it is assumed that the samples are taken from a lot with $1 - p_0 = .9$ then that lot contains 9 good items and 1 defective item. As the first step in the trial and error procedure assume a sample size of two. The possible outcomes are either 0, 1 or 2 good items. The probability of each outcome using the hypergeometric formula is $$\Pr(2) = \frac{\binom{9}{2}\binom{1}{0}}{\binom{10}{2}} = .8$$

3. Method	Example
g. r = number of successes in lot N = lot size n = sample size $\binom{r}{x} = \frac{r!}{x!\,(r-x)!}$	g. Pr(1) = .2 Pr(0) = 0 The same calculations for 1 - p_1 = .2 result in Pr(2) = .022 Pr(1) = .356 Pr(0) = .622
h. Find the number of successes which satisfies α and β in the calculations involving 1 - p_0 and 1 - p_1.	h. From these 2 sets of results it can be seen that if a sample size of 2 is specified, then α and β will be satisfied if the decision rule is made that if 2 successes are observed in the sample the lot is accepted and for all other outcomes the lot is rejected. If 1 - p_0 = .9, then Pr(2) = .8, therefore 1 - .8 = .2 = α. If 1 - p_1 = .2, then Pr(2) = .022 = β; NOTE: A different sample size can be traded off against different α, β, 1 - p_0 and 1 - p_1.
i. The demonstration test is then specified.	i. The test procedure is as follows: 1. Test a random sample of 2 missiles from a lot of 10 missiles. 2. If both missiles destroy the target, accept the lot. 3. If 0 or 1 successes are observed reject the lot.

4. For Further Information

There are *"Tables of the Hypergeometric Distribution"* by G.J. Lieberman and D.B. Owen, Stanford University Press, Stanford, California, 1961 to perform the mathematical calculations of Step g. Also if N becomes large (say 30) then the binomial or the Poisson distribution can be used as an approximation for the hypergeometric distribution.

8.4.7.1.2 Attributes Plans for Large Lots

1. When to Use

When testing parts from a large lot where the accept/reject decision is based on attributes, the binomial distribution is applicable. Strictly speaking, all reliability testing should follow the hypergeometric distribution as long as individual items are placed on test and tested to failure without repair. However, when the lot size is large, the binomial distribution is a good approximation for the hypergeometric and, therefore, the example presented in this section covers the use of the binomial. Attributes tests should be used when the accept/reject criterion is go/no-go, when the distribution of failure times is unknown, or when variables tests are found to be too expensive.

2. Conditions for Use

The definition of successfully passing the test may be that an item performs as specified. The parameter to be evaluated then is the fraction of the items in the lot that perform as specified. The estimation of the parameter would be based on a fixed sample size and testing without repair. The selection of the criteria for success can be derived from a requirement, or if the items being tested are known to follow a particular probability distribution, the specification of the criteria for success can be based on defining acceptable and unacceptable portions of the range of failure times. If the lot size is large, say 30 or more, then the Poisson approximation may be used to make the calculation simpler.

3. Method / Example

Method	Example
a. Define criterion for success/ failure	a. An artillery fuze that detonates on impact is considered a success. Fuzes that fail to detonate on impact are considered failures.

3. Method	Example
b. Define acceptable lot quality level $(1 - p_0)$.	b. Lots in which $1 - p_0 = .9$ (i.e., 90% of the fuzes in the lot will detonate on impact) are to be accepted by this demonstration test plan with high probability.
c. Specify producer's risk (α), (i.e., the probability that acceptable lots will be rejected).	c. Let $\alpha = .01$.
d. Define unacceptable lot quality level $(1 - p_1)$.	d. Lots with only a true fraction of acceptable parts $1 - p_1 = .5$ are to be accepted by this demonstration test plan with low probability.
e. Specify consumer's risk (β), (i.e., the probability that lots of unacceptable quality level will be accepted.)	e. Let $\beta = .12$ (selected for ease of calculation).
f. Now that α, β, $1 - p_0$, and $1 - p_1$ have been specified, the following steps describe the calculations required to determine the sample size and accept/reject criteria which will satisfy the stated risks.	f. Given: lot size N = large, say, 30 $1 - p_0 = .9$ $1 - p_1 = .5$ $\alpha = .01$ $\beta = .12$
g. The process now consists of a trial and error solution of the binomial equation using $1 - p_0$, $1 - p_1$ and various sample sizes until at a given decision point, the conditions of α and β are satisfied. The binomial equation is:	g. Assume a random sample of size n = 10 is taken from a lot whose true fraction of good parts is .9. Solve the binomial equation for the total number of consecutive outcomes whose summed probabilities equal a starting at 0 successes. The calculations for this decision point are:

3. Method	Example
$Pr(x) = \binom{n}{x}(1 - p)^x(p)^{n-x}$	$Pr(10) = \binom{10}{10}(.9)^{10}(.1)^0 = .3486$
	$Pr(9) = .387$
where:	$Pr(8) = .1935$
n = sample	$Pr(7) = .0574$
x = observed successes in sample	Pr(7 or more) = .9865
p = lot fraction defective	Then
	Pr(6 or less) = 1 - Pr (7 or more) = 1.0 - .9865 = .0135 ≈ .01 (which satisfies the risk.)
	Perform the same type of calculations assuming the true fraction defective is .5. In this instance, sum the probabilities starting at 10 successes until succeeding consecutive probabilities sum to the value of β. This yields the following results:
	$Pr(10) = \binom{10}{10}(.5)^{10}(.5)^0 = .001$
	$Pr(9) = .01$
	$Pr(8) = .045$
	$Pr(7) = .117$
	Pr(7 or more) ≈ .12 (which satisfies the β risk).
h. The demonstration test is then specified.	h. The test procedure is as follows: 1. Test a random sample of 10 fuzes. 2. If 7 or more fuzes detonate on impact accept the lot. 3. If 6 or less successes are observed, reject the lot.

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

4. For Further Information

There are several published tables for use in determining binomial probabilities in the event that the sample size makes calculations too lengthy. One of these is *"Tables of the Binomial Probability Distribution,"* National Institute of Standards and Technology, US Department of Commerce. It gives individual terms and the distribution function for p = .01 to p = .50 in graduations of .01 and n = 2 to n = 49 in graduations of 1. If n is large say, ≥ 30, the Poisson distribution can be used as an approximation for the binomial distribution.

8.4.7.2 Attributes Demonstration Test Plans for Large Lots, Using the Poisson Approximation Method

1. When to Use

In attributes demonstration test plans, if the lot size gets much above 20, the calculations required to generate a demonstration test plan become very time consuming. The Poisson distribution can be used as an approximation of both the hypergeometric and the binomial distributions if the lot size is large and if the fraction defective in the lot is small. This method can therefore be used in lieu of the previous two methods in many cases.

2. Conditions for Use

If the lot size is large and the fraction defective is small, this method is applicable. Its use is initiated by specifying a desired producer's risk, consumer's risk, acceptable lot fraction defective and unacceptable lot fraction defective. As before, it is also necessary to specify the characteristics that constitute a defective part since this is an attributes type test.

3. Method	Example
a. Define criterion for success/failure.	a. An artillery fuze that detonates on impact is considered a success. Fuzes that fail to detonate on impact are considered failures.
b. Define acceptable lot quality level $(1 - p_0)$.	b. Lots in which $1 - p_0 = .9$ (90% of the fuzes in the lot detonate on impact) are to be accepted by this demonstration test plan with high-probability.

3. Method	Example
c. Specify the producer's risk (α), (i.e., the probability that acceptable lots will be rejected).	c. Select α = .05.
d. Define unacceptable lot quality level $(1 - p_1)$.	d. Lots with only a true fraction of acceptable parts $1 - p_1 = .75$ are to be accepted by this demonstration test plan with low probability.
e. Specify the consumer's risk (β) (i.e., the probability that lots of unacceptable quality level will be accepted by this plan).	e. Select β = .02.
f. Now that α, β, $1 - p_0$, $1 - p_1$ have been specified, the Table of the Summation of Terms of Poisson's Exponential Binomial Limit* is used to determine the accept/reject criteria.	f. Given: lot size N = 1000 $1 - p_0 = .9$ $1 - p_1 = .75$ $\alpha = .05$ $\beta = .02$
g. The process now consists of a trial and error solution using Poisson Tables*, $1 - p_0$, $1 - p_1$ and various assumed sample sizes until the conditions of α and β are satisfied.	g. Assume sample size of 100. Now, calculate the expected number of failures for $1 - p_0$ and $1 - p_1$ as follows: $n(1 - p_0) = 100(.9) = 90$ $n(1 - p_1) = 100(.75) = 75$ The Poisson Tables are constructed for small values of p, so, in this case, to make calculations easier, it is necessary to work with the opposite tail of the distribution. Therefore the numbers to enter the table with are:

*See any good statistical text

3. Method | Example

$np_0 = 100(.1) = 10$

$np_1 = 100(.25) = 25$

The procedure now is to enter the column labeled c' or np' with the above numbers. Beginning with $1 - p_0 = .9$ and $np_0 = 10$, search across the np' = 10 row beginning at c or less = 1.0.

Continue to smaller values of c until the probability of c or less = $1 - \alpha$.

In this example at c = 15 or less, the probability of 15 or less is .951 which is approximately $1 - \alpha$.

The same procedure is followed in the table at $1 - p_1 = .75$ and $np_1 = 25$.

In the np' = 25 row at c = 15, the cumulative probability is .022 which is approximately equal to β.

The decision criteria is now specified as c = 15 or less failures.

h. The demonstration is then fully specified.

h. The demonstration test procedure is as follows:

1. Take a random sample of 100 fuzes from each lot of size N = 1000 and test each part.
2. If 85 or more fuzes (i.e., 15 or less defectives) detonate on impact, accept the lot.
3. If less than 85 successes are observed, reject the lot.

4. For Further Information

For additional examples using this method, refer to E. B. Grant "*Statistical Quality Control*," McGraw Hill, 1964.

8.4.7.3 Attributes Sampling Using ANSI/ASQC Z1.4-1993

ANSI/ASQC Z1.4-1993 replaced MIL-STD-105, but all applicable tables, table numbers and procedures used in MIL-STD-105 were retained.

1. When to Use

When the accept/reject criteria for a part is based on attributes decisions ANSI/ASQC Z1.4-1993 is a useful tool. These sampling plans are keyed to fixed AQL's and are expressed in lot size, sample size, AQL and acceptance number. Plans are available for single sampling, double sampling and multiple sampling. The decision as to which type to use is based on a trade-off between the average amount of inspection, the administration cost and the information yielded regarding lot quality. For example, single sampling usually results in the greatest amount of inspection, but this can be offset by the fact that it requires less training of personnel, and record keeping is simpler, and it gives a greater amount of information regarding the lot being sampled.

2. Conditions for Use

The user of a ANSI/ASQC Z1.4-1993 sampling plan must have the following information:

a. Lot Size

b. Acceptable Quality Level (AQL)

c. Sample Size

d. Acceptance Number

e. Criteria for Acceptance or Rejection

The specification of the AQL is an engineering decision based on the fraction defective that a user of parts considers acceptable. Lots with this percent defective will be accepted a high fraction of the time. Operating characteristic curves are supplied with each sampling plan and these can be used to evaluate the protection afforded by the plan for various quality levels.

ANSI/ASQC Z1.4-1993 also contains plans for normal, tightened and reduced inspection plans which can be invoked if the fraction defective of lots seems to be varying or trending.

3. Method	Example
a. Determine lot size and specify AQL and type of sampling.	a. Given a lot containing 100 parts, with an AQL of 6.5% and single sampling specified.
b. Enter the table with lot size and select the sample size code letter.	b. From Table I Sample Size Code Letters in ANSI/ASQC Z1.4-1993, find the sample size code letter for a lot of size 100. For this example and for normal sampling, the specified code number is F.
c. Enter the single sampling plan table for normal inspection with the code number from Step b.	c. Enter Table II-A Single Sampling Plans for Normal Inspection page 10 with code letter F. Under the column titled Sample Size, find the number 20 in the same row as the letter F. This is the number of parts to be randomly selected and inspected.
d. Enter the same table in the proper column for the specified AQL.	d. Find the column in Table II-A page 10 corresponding to an AQL of 6.5%.
e. Proceed horizontally along the Sample Size Code Number row until it intersects with the AQL column to obtain the acceptance number.	e. At the intersection of row R and column 6.5%, the acceptance number is 3 and the rejection number is 4.
f. The Single Sampling Plan from ANSI/ASQC Z1.4-1993 is to select a random sample of size n from a lot of size N, inspect it and accept the lot if the number of defectives in the lot is equal to or less than the Acceptance Number. If the observed number of defects is equal to or greater than the rejection number, the lot is rejected.	f. For the single sampling plan N = 100, AQL = 6.5%, select a random sample of size n = 20 and inspect it for attributes criteria. If 3 or less defectives are found in the sample accept the lot. If 4 or more defectives are found in the sample reject the lot.

4. For Further Information

In addition to the example discussed above, ANSI/ASQC Z1.4-1993 contains other plans for any lot size and for selected AQL's from .01 to 1000. Operating characteristic curves are also included.

8.4.7.4 Sequential Binomial Test Plans

1. When to Use

When the accept/reject criterion for the parts on test is based on attributes, and when the exact test time available and sample size to be used are not known or specified then this type of test plan is useful. The test procedure consists of testing parts one at a time and classifying the tested parts as good or defective. After each part is tested, calculations are made based on the test data generated to that point and the decision is made either that the test has been passed, failed, or that another observation should be made. A sequential test will result in a shorter average number of parts tested than either failure-truncated or time-truncated tests when the lot tested has a fraction defective at or close to p_0 or p_1.

2. Conditions for Use

a. The parts subjected to test will be classified as either good or defective. In other words, testing will be by attributes.

b. The acceptable fraction defective in the lot, p_0, the unacceptable fraction defective, p_1, the producer's risk α, and consumer's risk β must be specified.

c. The test procedure will be to test one part at a time. After the part fails or its test time is sufficient to classify it as a success, the decision to accept, reject or continue testing the lot will be made.

3. Method | Example

Method	Example
a. Specify p_0, p_1, α, β.	a. Given a lot of parts to be tested by attributes. Lots having only p_0 = .04 fraction defective parts are to be accepted by the demonstration test plan 95% of the time (i.e., α = .05). Lots having p_1 = .10 fraction defective are to be accepted 10% of the time (i.e., β = .10).

3. Method — Example

b. Calculate decision points from the following formula

$$\frac{1-\beta}{\alpha} \quad \text{and} \quad \frac{\beta}{1-\alpha}$$

b. The decision points are:

$$\frac{1-\beta}{\alpha} = \frac{1-.10}{.05} = 18$$

$$\frac{\beta}{1-\alpha} = \frac{.10}{1-.05} = .105$$

c. As each part is tested classify it as a part failure or a success and evaluate the following expression:

$$\left(\frac{p_1}{p_0}\right)^f \left(\frac{1-p_1}{1-p_0}\right)^s$$

where:
f = total number of failures
s = total number of successes

c. In this example, if the value of the formula

$$\left(\frac{.10}{.04}\right)^f \left(\frac{.90}{.96}\right)^s$$

1) exceeds 18, reject the lot.
2) < .105 accept the lot.
3) is between .105 and 18, the test should be continued.

d. A graphical solution for critical values of f and s is possible by solving the following equations.

1) $$ln\left(\frac{1-\beta}{\alpha}\right) = (f)\ ln\left(\frac{p_1}{p_0}\right) + (s)\ ln\left(\frac{1-p_1}{1-p_0}\right)$$

d. The equations for the graphical solution in this example are:

1) $ln\ 18 = f\,ln\ 2.5 + s\,ln\ .94$
2) $ln\ .105 = f\,ln\ 2.5 + s\,ln\ .94$

Substituting value of f and s in the equations yields the following points.

3. Method

$$2)\ ln\frac{\beta}{1-\alpha} = (f)\,ln\frac{p_1}{p_0} + g)\,ln$$

$$\frac{\beta}{1-\alpha} = (f)\,ln\frac{p_1}{p_0} + (s)\,ln\left(\frac{1-p_1}{1-p_0}\right)$$

Example

1)

f	s
0	-46.6
3.16	0
3.84	10
10	101

2)

f	s
-2.44	0
-1.78	10
0	36.4
10	184

Figure 8.4-6 shows the graphical solution for this test plan. As each good part is observed a horizontal line is drawn, and each defective part is recorded by a vertical line. When the line crosses either of the decision lines, the appropriate action is taken.

e. The Operating Characteristic Curve calculation is as follows:
Four points can be generated by observation.

p	Probability of Acceptance
p_0	$1-\alpha$
p_1	β
1	0
0	1

One additional point can be calculated with the following formula

$$p = \frac{ln\left(\frac{1-p_1}{1-p_0}\right)}{ln\left(\frac{1-p_1}{1-p_0}\right) - ln\left(\frac{p_1}{p_0}\right)}$$

$$P_r(Acc) = \frac{ln\frac{1-\beta}{\alpha}}{ln\frac{1-\beta}{\alpha} - ln\frac{\beta}{1-\alpha}}$$

where $P_r(Acc)$ = probability of acceptance

e. The OC curve for this test plan yields the following points:

p	Probability of Acceptance
.04	.95
.10	.10
1.00	0.00
0.00	1.00

The 5th point of the OC curve in the example

$$p = \frac{ln\ 0.94}{ln\ 0.94 - ln\ 2.5} = .063$$

$$P_r(Acc) = \frac{ln\ 18}{ln\ 18 - ln\ 0.105} = .562$$

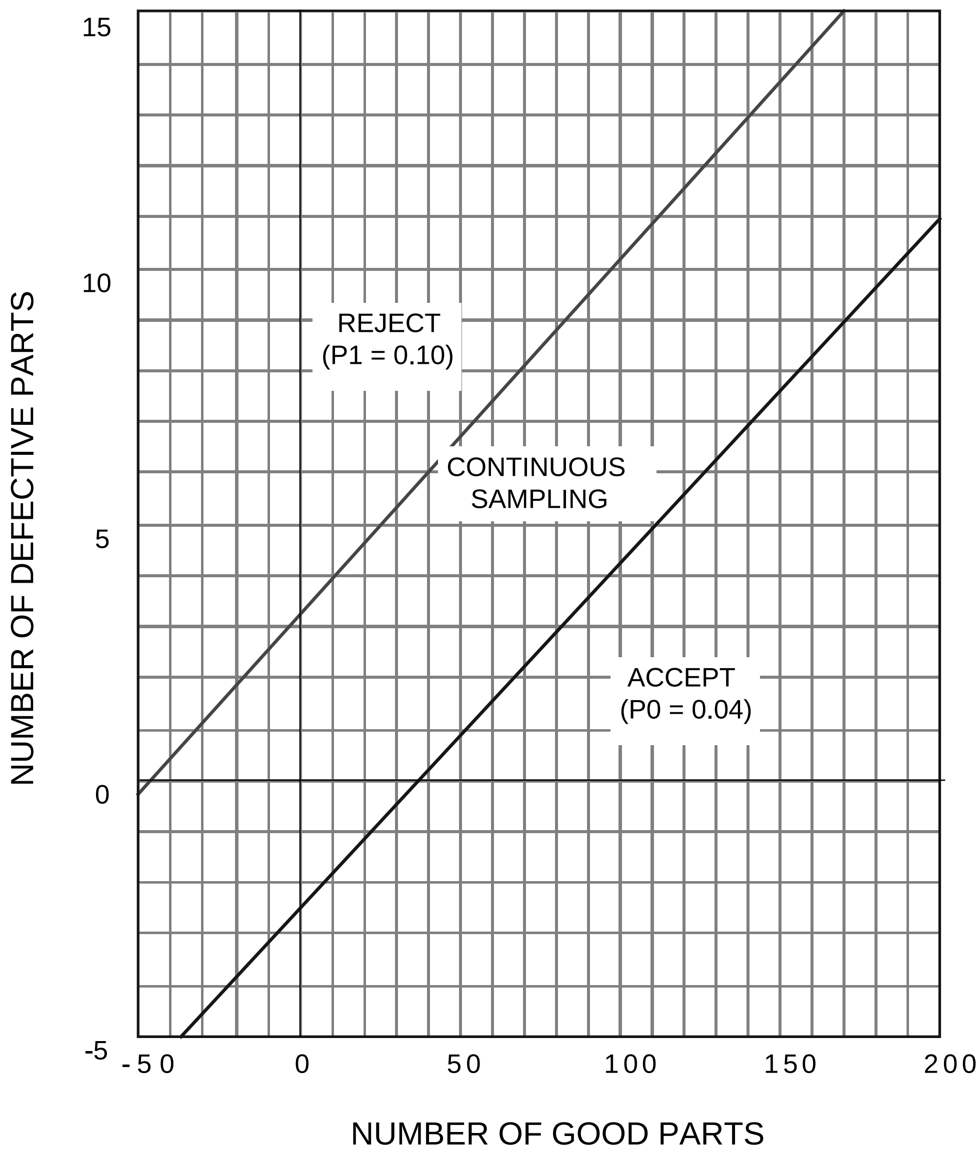


FIGURE 8.4-6: GRAPHICAL SOLUTION OF SEQUENTIAL BINOMIAL TEST

4. For Further Information

A more complete discussion of this demonstration test method is presented in *"Introduction to Statistical Analysis"* by W.J. Dixon and F.J. Massey, McGraw Hill, New York, 1951. The theory of sequential testing is presented in *"Sequential Analysis"* by A. Wald, John Wiley & Sons, 1947.

8.4.7.5 Variables Demonstration Tests

8.4.7.5.1 Time Truncated Demonstration Test Plans

8.4.7.5.1.1 Exponential Distribution (H-108)

1. When to Use

When a demonstration test program is constrained by time or schedule and testing is by variables (in this case the variable is mean life) and the distribution of failure times is known, a test plan of this type can be specified.

2. Conditions for Use

a. The failure times of the items under test must be exponentially distributed.

b. The acceptable mean life θ_0, unacceptable mean life θ_1, producer's risk (α), and consumer's risk (β), and test time (T) must be specified.

c. The decision of testing with or without replacement must be made.

3. Method / Example

Method	Example
a. Specify θ_0, θ_1, α, β.	a. Given an item type whose failure times are distributed exponentially. Specify θ_0 = 1000 hours θ_1 = 500 hours α = .10 β = .10
b. Specify a fixed test time.	b. The program plan allows time for a 200 hour test.

3. Method / Example

Method	Example
c. Specify whether testing will be with or without replacement.	c. Testing will be carried on without replacement.
d. Calculate T/θ_0.	d. $T/\theta_0 = \frac{1}{5}$
e. Calculate θ_1/θ_0.	e. $\theta_1/\theta_0 = \frac{500}{1000} = \frac{1}{2}$
f. From the appropriate table in MIL-HDBK-H108 *"Sampling Procedures and Tables for Life and Reliability Testing (Based on Exponential Distribution)"* select the sample size and number of failures which will cause rejection of the lot from which the parts were randomly selected.	f. Enter Table 2C-3 on page 2.52 of MIL-HDBK-H108 with α, β, T/θ_0 and θ_1/θ_0 and select the number of items to be placed on test (in this case 59) and the number of failures (in this example 15) which will cause failure of the demonstration test.
g. Summarize test outcome.	g. The demonstration test plan specified here has the following characteristics: 1. Lots having an MTBF of 1000 hours will be accepted 90% of the time. 2. Lots having a MTBF of 500 hours will be accepted 10% of the time. 3. Test 59 items for 200 hours each. Do not replace or repair parts as they fail. 4. If less than 15 failures occur, terminate the test at 200 hours and accept the lot. 5. If 15 or more failures occur reject the lot at the time of the fifteenth failure.

4. For Further Information

The demonstration test method and example discussed in this section are from "*Quality Control and Reliability Handbook*," *MIL-HDBK-H108*. In addition to the example presented here, the handbook has tabled sample sizes and reject numbers for testing without replacement with α = .01, .05, .10 and .25, and β = .01, .05, .10 and .25 and for all combinations thereof. The tables are also constructed for θ_1/θ_0 values of 2/3, 1/2, 1/3, 1/5 and 1/10 and T/θ_0 values of 1/3, 1/5, 1/10 and 1/20. A like set of tables is presented also for demonstration test plans for the same values of α, β, θ_1/θ_0 and T/θ_0 for testing with replacement. Tables are also provided for time truncated tests in which only α, θ_0 and T (test time) are specified (α = .01, .05, .10, .25 and .50) for plans involving testing with and without replacement. Fixed time test plans are also presented in MIL-HDBK-781.

8.4.7.5.1.2 Normal Distribution

1. When to Use

When the underlying distribution of failure times is normal and when a fixed calendar time is available for a test, this type of test plan can be specified. This test plan essentially becomes a binomial type problem since the survivors at the end of the time truncation are treated as successes. The failures regardless of their time of occurrence are utilized in specifying the accept/reject criteria.

2. Conditions for Use

a) The distribution of failure times must be normal.

b) The acceptable mean life (θ_0), unacceptable mean life (θ_1), the known or desired standard deviation of the distribution of acceptable mean lives (σ_0), the known or desired standard deviation of the distribution of unacceptable mean life (σ_1), the sample size (n), the test truncation time (T), the producer's risk (α) and the consumer's risk (β), must be specified.

c) The test should be run without replacement of failed parts.

3. Method

a. Specify θ_0, θ_1, α, β, σ_0, σ_1, n, T. If the requirements are stated in terms of reliability at some time t, it is necessary to solve the following equation.

$$z_0 = \frac{t - \theta_0}{\sigma_0}$$

where z_0 is the standard normal deviate for the desired probability of R_0, t is the desired mission time, σ_0 is the known standard deviation, and θ_0 is the acceptable mean life. The same procedure is followed to solve for θ_1 and R_1 is specified.

$$z_1 = \frac{t - \theta_1}{\sigma_1}$$

Example

a. Given an item type whose failure times are normally distributed with a known standard deviation = 50. A reliability of .95 is desired that the equipment will last 100 hours. A product with a reliability of .85 is unacceptable.

The standard normal deviate for R_0 = .95 is z_0 = -1.645 and for R_1 = .85 is z_1 = -1.04 from a table of areas under the normal curve (Table 5.3.1-1).

$$z_0 = \frac{t - \theta_0}{\sigma}$$

$$-1.645 = \frac{100 - \theta_0}{50}$$

$$\theta_0 = 182 \text{ hours}$$

$$z_1 = \frac{t - \theta_1}{\sigma}$$

$$-1.04 = \frac{100 - \theta_1}{50}$$

$$\theta_1 = 152 \text{ hours}$$

Therefore, it is possible to specify R_0 and R_1 in terms of θ_0 and θ_1.

3. Method / Example

$\theta_0 = 182$ hours
$\sigma_0 = 50$ hours
$\theta_1 = 152$ hours
$\sigma_1 = 50$ hours

The schedule and cost of testing allows 182 hours of test time with 30 samples to be placed on test. α is specified as .10 and $\beta = .05$.

b. Calculate the expected number of failures during the fixed time test if n samples are tested T hours, for samples from lots with mean lives of θ_0, σ_0 and θ_1, σ_1.

b. The $\theta_0 = 182$, $\sigma_0 = 50$, n = 30 then the expected number of failures in a test of 182 hours is 15. If $\theta_1 = 152$, $\sigma_1 = 50$, n = 30, the expected number failures in a test of 182 hours is 21.6 using a table of areas under the normal curve.

c. The problem of specifying accept/reject criterion at the end of a fixed test time, T, is now similar to the example in Attributes Plans For Large Lots. In other words, it is a binomial distribution problem since items that last T hours are listed as having successfully passed the test, while items that do not last T hours are classed as failures regardless of their exact failure times.

c. Items that exceed the fixed test time T = 182 hours are counted as successes. The remaining problem to be solved is specifying the accept/reject criterion (i.e., r or more failures out of a sample of 30 items on test for 182 hours results in failure of the demonstration test - regardless of the individual part failure times). Additionally, the test may be terminated at less than T = 182 hours if r failures are observed, in which case the demonstration test is failed.

3. Method / Example

Method

d. The accept/reject criteria can be calculated using the binomial distribution or, if the expected number of failures $\geq$ 5 the normal distribution can be used as an approximation to the binomial.

Example

d. From Step b the expected number of failures of θ_0 = 182 is 15 and the expected number of failures when θ_1 = 152 is 21.6. Therefore the normal distribution as an approximation of the binomial is used.

Method

e. Calculate the decision point based on θ_0 and α using the normal distribution.

Example

e. The decision point for θ_0 = 182, α_0 = 50, α = .10 is calculated as follows:

$$z = 1.28 \text{ for } \alpha = .10$$

$$z = \frac{x - np}{\sqrt{np(1-p)}}$$

$$1.28 = \frac{x - 15}{\sqrt{15(.5)}}$$

$$x = 18.5 \text{ failures}$$

The demonstration test plan procedure is now stated as follows:

Take a random sample of 30 items, test them for 182 hours. If, 18.5 or less failures are observed the test is passed.

Method

f. Adjust the decision point to a whole number, thus adjusting α slightly.

Example

f. Either 18 or 19 failures can be set as the rejection number without affecting α too severely. For this example, assume that 19 failures will be allowed and still accepted. α now becomes

3. Method | Example

$$z = \frac{19 - 15}{\sqrt{15(.5)}} = 1.46$$

From a Table of Areas under the Normal Curve the probability of exceeding z = 1.46 is .09. Therefore, α = .09.

g. Calculate β based on the accept/reject criteria established in Step f.

NOTE: The OC curve for this demonstration test plan can be constructed by assuming different values of θ and performing similar calculations to those of this step. Note that np and 1 - p will change for each new value of θ.

g. If θ_1 = 152 hours, σ_1 = 50, T = 182 hours, n = 30, and the decision rule for passing the test is 19 or less failures, then β is calculated as:

$$z = \frac{x - np}{\sqrt{np(1-p)}} = \frac{18 - 21.6}{\sqrt{21.6(.28)}}$$

$$z = -1.46$$

The area under the normal curve not exceeding a z value of -1.46 is .07. Therefore, β = .07.

h. Summarize the characteristics of the demonstration test plan.

h. Test a random sample of 30 items for 182 hours. If 19 or less failures are observed, the test has been passed. If 19 or more failures are observed the test is failed. If the 19th failure occurs before 182 hours, stop testing when it occurs, as the test is failed.

This test plan will reject lots with an average mean life of 182 hours and standard deviation of 50 hours approximately 9% of the time. It will accept lots with an average mean life of 152 hours and a standard deviation of 50 hours approximately 7% of the time.

4. For Further Information

Additional examples describing this method are presented in most books on elementary statistics.

8.4.7.5.1.3 Weibull Distribution (TR-3, TR-4, TR-6)

1. When to Use

When the distribution of failure times is Weibull and when only a given calendar time is available for a demonstration test, then this type of test plan is useful. Test plans covering this situation have been generated by Kao and Goode and published as a series of Quality Control and Reliability Technical Reports (TR-3, TR-4, TR-6) titled *"Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution"* by the Office of the Assistant Secretary of Defense (Installations and Logistics), September 1961, February 1962 and February 1963. (Refs. [13], [14], [15]). The plans are based on the user of the test plans specifying his reliability parameter of interest in terms of mean life, hazard rate, or reliable life (life at given failure %). The plans were generated based on the assumption of a known shape parameter and give protection against a certain fraction of items in a lot not meeting the acceptance criterion. The test procedure essentially states that a sample of n items should be tested t hours. Those surviving the fixed time are classed as successes, while those not surviving are considered failures regardless of the exact time of failure. From this definition of failure it can be seen that these plans are based on the binomial distribution. Tables of the cumulative binomial distribution can be used to generate the OC curves for specific test plans. Each set of test plans features a set of conversion factors relating to ANSI/ASQC Z1.4-1993 Sampling Plans. Tabled test plans are presented for values of the Weibull shape parameter of .33, .5, 1, 1.67, 2.5, 3.33, 4 and 5.

2. Conditions for Use

a. The failure times of the items being evaluated follow the Weibull distribution with known or assumed shape parameter β.

b. The acceptable mean life μ_0, unacceptable mean life μ_1, producer's risk α, consumer's risk β (care must be taken to differentiate this quantity from the Weibull shape parameter which is also symbolized by β) and the test time t, must be specified.

c. Testing is without replacement.

d. It is also possible to select test plans be specifying the fraction defective allowable in a lot having an acceptable quality level.

3. Method	Example
a. Specify μ_0, μ_1, α, β (consumer's risk), β (Weibull shape parameter) and test time t.	a. Given a lot of items whose failure times follow the Weibull distribution. Historical failure data on the item indicates the Weibull shape parameter β is approximately 2.0. The program schedule allows 2500 hours of reliability demonstration testing. Lots having a mean life μ_0 of 10,000 hours are to pass the demonstration test 95% of the time (i.e. α = .05). Lots having a mean life μ_1 of 5,000 hours are to be accepted by this test plan only 10% of the time (i.e., consumer's risk β = .10).
b. Determine the sample size and acceptance number for a plan that will give the protection specified in Step a.	b. Enter Table 3e on page 32 on TR-3 *"Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution"* which is for sampling plans for the case of the Weibull shape parameter β = 2.0. The quantity that is used to enter the table is $$t/\mu_1 \times 100 = \frac{2500}{5000} \times 100 = 50$$ Search the column headed by 50 for the parenthesized value in the body of the table corresponding to $$t/\mu_0 \times 100 = \frac{2500}{10000} \times 100 = 25$$

3. Method	Example
	b. The table contains values for t/μ_0 x 100 of 24 and 26. To assure greater protection (i.e., a smaller α) the larger value should be used.
	The t/μ_0 x 100 = 26 row specifies a sample size of 50 with an acceptance number of 5.
c. Summarize the test procedure.	c. The test procedure is as follows: 1) Select a random sample of 50 items (from a large lot). 2) Test the items for 2500 hours. 3) If the number of failures observed during the test is 5 or less accept the lot. 4) If there are 6 or more failures is reject the lot. 5) If the 6th failure occurs before 2500 hours, the test may be discontinued at that point and the lot rejected.

4. For Further Information

Frequently, the exact test desired is not covered in the tabled values in which case it is possible to interpolate to some degree at the expense of changing the risks slightly. Operating characteristic curves can be generated using a table of binomial probabilities.

Each of the Technical Reports contains an extensive bibliography describing other publications in which the details leading to these sampling plans were presented by Professors Goode and Kao.

8.4.7.5.2 Failure Truncated Tests

8.4.7.5.2.1 Exponential Distribution (MIL-HDBK-H108)

1. When to Use

When tests designed to demonstrate life characteristics of items whose failure times are exponentially distributed are to be performed wherein the test will be terminated after a preassigned number of failures then a test plan of this type can be specified. Plans of this type are available in MIL-HDBK-H108,"*Sampling Procedure and Tables for Life and Reliability Testing (Based on Exponential Distribution),*" also known as "*Quality Control and Reliability Handbook.*" Plans are presented for testing with and without replacement. Test criteria are tabled for specified values of α and β equal to .01, .05, .1, and .25 and for all combinations thereof, and for values of θ_1/θ_0 of 2/3, 1/2, 1/3, 1/5 and 1/10. A set of tables is also presented for cases in which α and θ_0 only are specified for various values of termination number r. Since a major factor in specifying a demonstration test plan of this type is the expected waiting time before a decision is made (i.e., a given number of failures occur) there is also included a set of tables for calculating this statistic for various sample sizes and termination numbers. Operating characteristic curves are presented for many of the demonstration test plans to enable the assessment of risk for values of mean life other than θ_0 and θ_1.

2. Conditions for Use

a. The failure times of the items placed on test must be exponentially distributed.

b. The acceptable mean life θ_0, unacceptable mean life θ_1, producer's risk α, and consumer's risk β should be specified.

c. The decision of whether testing will be with or without replacement must be made.

d. An estimate may be made regarding the time available for the test as this will affect the number of items placed on test.

3. Method	Example
a. Specify θ_0, θ_1, α, β.	a. Given an item type whose failure times are distributed exponentially. Specify θ_0 = 1000 hours θ_1 = 500 hours α = .10 β = .10
b. Specify whether testing will be with or without replacement.	b. Testing will be without replacement.
c. Calculate θ_1/θ_0.	c. $\theta_1/\theta_0 = \frac{500}{1000} = \frac{1}{2}$
d. Enter the appropriate table in MIL-HDBK-H108 and select a termination number and acceptability constant.	d. Enter Table 2B-5 on page 2.41 of MIL-HDBK-H108 with α = .10, β = .10, and $\theta_1/\theta_0 = \frac{1}{2}$. The termination number is 15 and the acceptability constant is .687.
e. Establish test procedure.	e. The specified demonstration test has the following characteristics: 1) Items with a mean life of 1000 hours will be accepted by this test plan 90% of the time. 2) Items with a mean life of only 500 hours will be accepted by this test plan only 10% of the time. 3) Select a random sample of 15 or more items and test until 15 failures are observed. 4) Multiply the acceptability constant by θ_0 (in this example 1000) = .687.

3. Method Example

e.

5) After 15 failures have been observed stop the test and sum the hours of operating time accumulated on all items that have been on test (both failed and unfailed). Divide the total

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

2. Conditions for Use

a. The distribution of failure times must be normal.

b. The standard deviation of failure times must be assumed known.

c. The acceptable mean life θ_0, the standard deviation σ_0 of the distribution of acceptable mean life, the standard deviation σ_1 of unacceptable mean life, the sample size n to be tested to failure, the producer's risk α must be specified.

d. Note that unacceptable mean life θ_1 is not specified in this example. If it were desirable to specify a θ_1, it could be done but one of the other four test plan parameters θ_1, α, β, or sample size n would change. In other words, any four of these quantities can be specified but then the fifth is automatically constrained by the selection of the 4.

e. There is also a tradeoff between the sample size and the accept/reject decision point. In the following example, the sample size to be tested has been specified, but it would be possible to specify a mean life which, if the observed average failure time did not exceed, would result in failure of the lot to pass the demonstration test. With this critical mean life specified, it would be necessary to solve for the sample size to be tested.

f. Testing should be without replacement.

3. Method	Example
a. Specify θ_0, σ_0, σ_1, β and n.	a. Given a lot whose item failure times are normally distributed as follows: θ_0 = 200 hours σ_0 = 50 hours α = .01 σ_1 = 50 hours β = .05 n = 25

3. Method | Example

b. Solve for the accept/reject decision point.

b. The accept/reject point is calculated as follows:

$$z_0 = \frac{\bar{x} - \theta_0}{\sigma_0/\sqrt{n}}$$

$$-2.33 = \frac{\bar{x} - 200}{50/\sqrt{25}}$$

$$\bar{x} = 176.7$$

c. Solve for θ_1.

c. Using the result from Step (b) and the specified β = .05

$$z_1 = \frac{\bar{x} - \theta_1}{\sigma_1/\sqrt{n}} +$$

$$1.645 = \frac{176.7 - \theta_1}{50/\sqrt{25}}$$

$$\theta_1 = 160.25$$

NOTE: The z values are from a table of "Areas Under the Normal Curve."

d. Summarize the characteristics of the demonstration test plan.

d. The demonstration test procedure is as follows:

1) Take a random sample of 25 items from a population whose distribution of failure times is normal.

3. Method | Example

2) Test until all items have failed, recording the exact failure time of each.
3) Take the arithmetic mean of the 25 failures and compare it with the decision point 176.7 hours. If the observed mean equals or exceeds 176.7 hours the demonstration test is passed. If it is less than 176.7 the demonstration test is failed.
4) The demonstration test shown in this example will:

   - accept lots with a mean life of 200 hours and a standard deviation of 50 hours 99% of the time.
   - accept lots with a mean life of 160.25 hours and standard deviation of 50 hours 5% of the time.

e. Construct the operating characteristic curve.

e. This is done by assuming values of θ other than θ_0 and θ_1 and solving for the probability of acceptance of a lot with that θ. Assume

$$\theta = 175, \sigma = 50$$

$$z = \frac{176.7 - 175}{50/\sqrt{25}} = \frac{1.7}{10} = .17$$

From a table of Areas Under the Normal Curve the probability of acceptance of a lot with a mean life of 175 hours, σ = 50 is approximately .43.

3. Method — Example

f. Calculate the expected waiting time for a decision.

f. The expected waiting time for a decision is the expected failure time of the last order statistic. In this ex-ample and sample size $n = 25, \alpha = 50$ and $\mu = 200$. These values are used with Table 10A.1, page 186 of the book *"Contributions to Order Statistics"* edited by A.E. Sarhan and B.G. Greenberg, published by John Wiley & Sons, New York, 1962.

Table 10A.1 give a z = 1.965 for the last order statistic in a sample of n = 25. Applying the formula

$$z = \frac{x - \mu}{\sigma}$$

$$1.965 = \frac{x - 200}{50}$$

$$x = 298 \text{ hours}$$

Therefore the expected waiting time for a decision of $\theta_0 = 200$, and 25 items are tested to failure, is 298 hours.

4. For Further Information

MIL-STD-414 Section D yields a series of variables demonstration test plans for the normal distribution with σ known. The tests are constructed to assure protection in the form of percent defective of the lot from which the sample was drawn, whereas, the example presented here is based on mean life.

8.4.7.5.2.3 Normal Distribution, σ Unknown (MIL-STD-414)

1. When to Use

When the distribution of failure times is normal, with unknown standard deviation and the criterion for acceptance is a variable (in the case, hours of life expectancy) with the protection desired stated in terms of percent defective in the lot from which the sample was drawn, then this type of demonstration test is useful. This procedure basically is an application of MIL-STD-414, *"Sampling Procedures and Tables for Inspection by Variables for Percent Defective."* It contains plans for both single and double specification limits. The criteria for acceptance can either be stated in terms of an acceptability constant, k, stated in standard normal deviates or as a maximum allowable percent defective, M. MIL-STD-414 also presents plans based on the calculation of an estimate of the standard deviation from sample data and also presents the range method. In the range method, the sample is segmented and the range of each sub-sample is used to estimate variability. It also contains test plans for the case when the standard deviation is known.

2. Conditions for Use

a. The distribution of failure times must be normal.

b. The standard deviation is unknown and must be assumed equal for both acceptable and unacceptable lots (when it is known, see previous example).

c. Failure is measured in hours or cycles of operation.

d. All items in the sample will be tested to failure.

e. The lot size, acceptable quality level AQL, specification limit or limits, and inspection level must be stated.

f. Testing is performed without replacement of failed items.

3. Method	Example
a. Specify the lot size from which the sample is to be randomly drawn, AQL (the percent defective of accept-able lots), the specification limit, and the method to be used (standard deviation or range method) to measure variability.	a. Given an item type whose failure times are normally distributed. The lot to be evaluated contains 100 items with an unknown standard deviation. An AQL of 4% represents an acceptable level of defectives in a lot. The normal inspection level in MIL-STD-414 is IV. The standard deviation method is to be used for determining compliance with the accepability criterion. The minimum life (L) for items of this type is 300 hours.
b. Determine the sample size to be tested.	b. Enter Table A-2 on page 4 of MIL-STD-414 with the lot size = 100. It is found that for Inspection Level IV, sample size code letter F applies. On page 39 in Table B-1 sample size code letter F calls for a sample size of 10.
c. Determine the acceptability constant k.	c. From Table B-1 enter Row F and the column headed by AQL = 4.00. This yields an acceptability constant k = 1.23.

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

3. Method | Example

d. Draw a random sample from the lot and test until all items fail recording exact failure times.

d. Ten failure times are re-corded as follows:

Failure Time (Hours)

275
310
315
370
400
425
450
515
625
630

e. Calculate the sample mean and standard deviation from the observed test data.

e. Using standard statistical calculations

$\bar{x}$ = 432 hours
s = 119 hours

f. Calculate the quantity

$$\frac{\bar{x} - L}{s}$$

where L = the specified minimum life.

f. $\dfrac{\bar{x} - L}{s} = \dfrac{432 - 300}{119} = 1.10$

g. Compare $\dfrac{\bar{x} - L}{s}$ with k.

g. From Step c, the acceptability constant is k = 1.23. From Step f, $\dfrac{x - L}{s}$ = 1.10 Since 1.10 < 1.23, reject the lot.

4. For Further Information

MIL-STD-414 also presents test plans for cases where the standard deviation is known. Operating characteristic curves are presented in Section A of MIL-STD-414 to enable assessment of the risk at all quality levels. All lot sizes can be accommodated, but only certain values of AQL are covered by test plans. MIL-STD-414 also covers tightened and reduced sampling. A discussion of the methodology of the development of this type of sampling plan is presented in *"Quality Control and Statistics"* by A. J. Duncan, published by Richard D. Irwin, Homewood, Illinois, 1959.

8.4.7.5.2.4 Weibull Distribution

1. When to Use

When the underlying distribution of failure time is Weibull, with the shape parameter, β, known or assumed, and the test must be truncated after a specified number of failures has occurred. The ordered failure times are required, along with the number of items on test.

2. Conditions for Use

a. The two-parameter Weibull distribution must be assumed for failure times.

b. The parameter, β, must be known and be the same under the null and alternative hypothesis concerning the population mean.

c. The acceptable mean life, μ_0, the unacceptable mean life, μ_1, and the producer's risk must be specified. If the number of failures at which the test is truncated is specified, then the consumer's risk will be determined, and cannot be set arbitrarily.

3. Method | Example

a. The method involves replacement of the original failure times x_1, . . . x_r by a new variable defined as $y_i = x_i^{\beta}$

This variable has an exponential distribution with mean α. Hence, the previous method developed for failure-truncated exponential life-distributions may be used (See Section Exponential Distribution (MIL-HDBK-H108).

a. With producer's risk .05 and consumer's risk .10, test the hypothesis that μ_0 = 800 hours against μ_1 = 400 hours. Assume a Weibull distribution with parameter β = 1.5. Twenty specimens were placed on test, and the test was concluded after the fourth failure, the observed failure times being 600, 750, 1000, and 1220 hours.

b. To perform a Weibull demonstration test with parameters μ_0, μ_1, β. Solve the following equations:

$$\mu_0 = \alpha_0^{1/\beta}\ \Gamma\left(\frac{1}{\beta} + 1\right)$$

$$\mu_1 = \alpha_1^{1/\beta}\ \Gamma\left(\frac{1}{\beta} + 1\right)$$

for α_0 and α_1.

b. $$\alpha_0 = \left[\frac{\mu_0}{\Gamma(\frac{1}{\beta} + 1)}\right]^{\beta}$$

$$= \left(\frac{800}{\Gamma(1.67)}\right)^{1.5}$$

$$= \left(\frac{800}{.903}\right)^{1.5}$$

$$= 24600$$

$$\alpha_1 = \left(\frac{400}{.903}\right)^{1.5}$$

$$= 9400$$

3. Method	Example
c. Perform the demonstration test in Section Exponential Distribution (MIL-HDBK-H108) on the observations $y_1, y_2, \ldots y_K$ from the exponential distribution with	c. $y_1 = 600^{1.5} = 14,700$
$\theta_0 = \alpha_0$ $\theta_1 = \alpha_1$	$y_2 = 750^{1.5} = 20,500$ $y_3 = 1000^{1.5} = 31620$ $y_4 = 1220^{1.5} = 42,600$
The test is described in MIL-HDBK-H108.	$\hat{\theta} = \frac{1}{4}$ [14,700 + 20500 + 31620 + 42600 + 16(42600)]
On page 2.26 of MIL-HDBK-H108, the formula for $\hat{\theta}$ is	$\hat{\theta} = 197755$
$\hat{\theta} = \left[\frac{1}{r}\sum_{i=1}^{r} y_i + (n-r)y_r\right]$	$\theta_0 = 26400$
This is compared with acceptability constant, C, given on page 2.28 of MIL-HDBK-H108. The acceptance region is	C/θ_0 = .342 for producer's risk .05 and 4 failures (Table 2B-1) (MIL-HDBK-H108) Critical Value $= \frac{26400}{.342}$ $= 77200$
$\hat{\theta} \geq \theta_0/(C/\theta_0)$	Since 197755 > 77200, accept the value, μ_0, for the Weibull population mean
d. The consumer's risk may be estimated from OC curves provided in the referenced document. Compute θ_1/θ_0 and read the value of the β error from Table 2A-2.	d. $\frac{\theta_1}{\theta_0} = \frac{9400}{26400} = 0.36$ β = 0.38 from Table 2A-2 (MIL-HDBK-H108)

3. Method

The larger the value of θ, the smaller the value of β error. To achieve a β error of 0.1, for example, it would be necessary (Table 2-A-2) to continue testing until 9 failures had occurred.

4. For Further Information

Tables of the Gamma Function are presented on page 497 of the *"Handbook of Tables for Probability and Statistics"* edited by W. H. Beyer, Chemical Rubber Company, 1966.

8.4.7.5.3 Sequential Tests

8.4.7.5.3.1 Exponential Distribution (MIL-HDBK-781)

1. When to Use

When the demonstration test is to be based upon time-to-failure data and the underlying probability distribution is exponential, the sequential test is an alternate for the fixed sample size or fixed time tests discussed in Sections Time Truncated Demonstration Test Plans and Failure Truncated Tests. The sequential test leads to a shorter average number of part hours of exposure than either fixed sample or fixed time tests if the lot tested is near θ_0 or θ_1. Sequential tests should not be used where the exact length, or cost, of the test must be known before-hand, or is specified.

2. Conditions for Use

a. The failure distribution must be exponential.

b. The upper test MTBF, θ_0, lower test MTBF, θ_1, producer's risk, α, and consumer's risk, β, must be specified.

c. The test may be run either with or without replacement of failed items, since the pertinent statistic is "total item-hours" of test time.

d. The producer's risk, α, and consumer's risk, β, are always equal in these test plans.

3. Method / Example

Method

a. Specify θ_0, θ_1, α, β. If the requirements are stated in terms of reliability at a time T_0, this will involve solution of the equation.

$$\exp\left[-(T_0/\theta)\right] = R$$

for θ. The solution is

$$\theta = -\frac{T_0}{\ln R}$$

b. Compute θ_0/θ_1

Example

a. Given equipment type whose failure times are distributed exponentially. A reliability of 0.95 is desired for 150 hours of operation. A product with a reliability of 0.9 or lower is unacceptable. We specify that α = 0.10 for 0.95 reliability and β = 0.10 for 0.90 reliability.

We have

$$\theta_0 = -\frac{150}{\ln .95}$$

$$\theta_0 = 2924 \text{ hours}$$

$$\theta_1 = -\frac{150}{\ln .90}$$

$$\theta_1 = 1424 \text{ hours}$$

b. $\theta_0/\theta_1 = \frac{2924}{1424} = 2.05$

3. Method | Example

c. Tests in MIL-HDBK-781 are classified by θ_0/θ_1, α and β. Find the Test Plan which most nearly fits the three values, and record the acceptance and rejection criteria. These are given in terms of θ_1, and must be multiplied by θ_1 to convert to "equipment hours" criteria.

c. For $\alpha = \beta = .10$ the nearest test in MIL-HDBK-781 is Test Plan IIID. The criteria given for acceptance and rejection are:

No. of Failures	Equipment Reject	Hours Accept
0	N/A	4.4
1	N/A	5.79
2	N/A	7.18
3	0.7	8.56
4	2.08	9.94

After multiplying by θ_1, or 1424 hours, we obtain

No. of Failures	Equipment Reject	Hours Accept
0	-	6266
1	-	8245
2	-	10224
3	997	12189
4	2962	14155

For example, if 3 failures are encountered prior to 997 equipment hours, reject the equipment as unsatisfactory.

d. The OC curve of each sequential test is given as multiples of θ_0 and θ_1. The document supplies for each Test Plan the expected length and the OC curve.

d. The expected number of equipment hours to reach a decision, when θ_0 is the population parameter, and the OC curve are given in MIL-HDBK-781A (Ref. [18]).

4. For Further Information

The material presented herein is from MIL-HDBK-781, *"Reliability Test Methods, Plans and Environments for Engineering Development, Qualification and Production."* The theory of sequential testing is developed in *"Sequential Analysis"* by A. Wald, John Wiley and Sons, Inc., 1947. Examples of sequential exponential demonstration tests are given in an article by Benjamin Epstein and Milton Sobel, *"Sequential Life Tests in the Exponential Case,"* Annals of Mathematical Statistics, Vol. 25, 1955, pp. 82-93.

8.4.7.5.3.2 Normal Distribution

1. When to Use

When the underlying failure distribution is assumed to be normal, and random sample observations are gathered sequentially. This method does not apply to ordered sample observations such as are usually obtained in life testing. It is useful where the cost of a single test is high, testing is done one unit at a time, and it is desired to minimize expected sample size.

As an example, consider the destructive testing of an aluminum alloy exhaust fan, where the component is rotated in a "whirl pit" at increasing velocity until a tensile failure occurs. In service, the component will rotate at a maximum velocity v_0, and the purpose of the demonstration test is to assure that the population mean velocity at failure is sufficiently high to provide satisfactory reliability at v_0.

2. Conditions for Use

a. The distribution of failures must be normal.

b. The acceptable population mean, μ_0, unacceptable mean, μ_1, must be specified, along with the known or assumed population standard deviations, σ_0 and σ_1, the producer's risk, α, and consumer's risk, β. If α is unknown, and the test involves a strength distribution, α is often assumed to be 5% of the mean, in accordance with the discussion of normal distribution estimation in Section 5 of this handbook.

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

3. Method | Example

Method

a. Specify μ_0, μ_1, σ_0, σ_1, α, and β. Compute

$$A = \frac{1 - \beta}{\alpha}$$

$$B = \frac{\beta}{1 - \alpha}$$

Example

a. $\mu_0 = 1000$
$\mu_1 = 800$
$\sigma_0 = \sigma_1 = 100$
$\alpha = \beta = .05$

$$A = \frac{.95}{.05} = 19.0$$

$$B = \frac{.05}{.95} = .053$$

Method

b. Compute, as each new observation is obtained, the corresponding unit normal deviates

$$z_{oi} = \frac{x_i - \mu_0}{\sigma_0}$$

$$z_{1i} = \frac{x_i - \mu_1}{\sigma_1}$$

and the corresponding probability density from a table of the normal distribution ordinates (Table 5.3.1-2).

Note that it is not the usual areas under the normal curve but the ordinates that are required.

Example

b. The first sample observation was found to be

$x_1 = 1020$, hence

$$z_{01} = \frac{1020 - 1000}{100} = 0.2$$

$$z_{11} = \frac{1020 - 800}{100} = 2.2$$

The ordinate in the normal table corresponding to 0.2 is 0.3900 while the ordinate corresponding to 2.2 is 0.0355.

3. Method | Example

c. Form the product of ordinates

$$L_0 = \sum_{i=1}^{k} f(z_{0i})$$

and

$$L_1 = \sum_{i=1}^{k} f(z_{1i})$$

Determine, as each new sample is received, the ratio,

$$\frac{L_1}{L_0}$$

If

$$B < \frac{L_1}{L_0} < A$$

continue testing. If

$$\frac{L_1}{L_0} < B, \text{ accept } \mu_0$$

$$\frac{L_1}{L_0} > A, \text{ accept } \mu_1$$

c. $L_0 = .3900$

$L_1 = .0355$

$$\frac{L_1}{L_0} = \frac{.0355}{.3900} = .091$$

Since this is between B and A, continue testing. The second observation was

$x_2 = 904.$

Calculating as before,

$z_{02} = .96$

Ordinate = .2516

$z_{12} = 1.04$

Ordinate = .2323

$$\frac{L_1}{L_0} = .091\left(\frac{.2323}{.2516}\right) = .084$$

Therefore, continue testing.

We observe

$x_3 = 1050$

$z_{03} = 0.5$

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

3. Method / Example

Ordinate = .3521

z_{13} = 2.5

Ordinate = .0175

$$\frac{L_1}{L_0} = .084\left(\frac{.0175}{.3521}\right) = .004$$

Since this is less than B, accept μ_0 as population mean.

d. The expected sample size (assuming that the true parameter is μ_0) may be obtained from the formula

E(N) =

$$\frac{(1 - \alpha)\, ln\, B + \alpha\, ln\, A}{\frac{1}{2\sigma^2}\left[2(\mu_1 - \mu_0)\mu_0 + \mu_0^2 - \mu_1^2\right]}$$

d. For this test, the expected number of observations was

E(N) =

$$\frac{.95\, ln\, .053 - .05\, ln\, 19.0}{\frac{1}{20000}\left[2 - (200)(1000) + 1x10^6 - 6.4x10\right.}$$

$\approx$ 2

(Note: sample size must be an integer)

4. For Further Information

See *"Sequential Analysis"* by Abraham Wald, John Wiley and Sons, N.Y., 1947, p. 77 and p. 53.

8.4.7.6 Interference Demonstration Tests

1. When to Use

Interference demonstration testing is applicable to mechanical systems where a strength distribution and a stress distribution overlap, or interfere. See Section 7 for several detailed examples. In the case of demonstration testing, both the strength and stress distribution must be assumed to be normal. We distinguish four cases:

Case 1: The mean of the stress distribution is assumed to be known, and the standard deviation of the stress distribution is assumed to be zero. See the discussion in Section 7 for conditions where these assumptions are valid. In this case, the interference problem becomes identical to life testing of the normal distribution described in Section 8.4.7.5.2.2, Normal Distribution σ Known. The specified stress level plays the role of the specified life. The strength distribution plays the role of the life distribution, and the demonstration procedure follows the example in Section 8.4.7.5.2.2.

Case 2: The mean of the stress distribution is assumed to be known, along with its standard deviation (often assumed to be 5% of the mean). The standard deviation of the strength distribution is assumed to be known, and its mean unknown. This may be translated to a demonstration test on strength and solved by the methods of Section 8.4.7.5.2.2. An example will be given.

Case 3: The mean of the stress distribution and the mean of the strength distribution are unknown, but their standard deviations are assumed known. In this instance, sampling data will be required from both stress and strength. It is rare that a sample size for each may be specified ahead of testing. Therefore, it is unlikely that the consumer's risk may be set for this test. β will be a function of N and α. An example will be given.

Case 4: The means and standard deviations of the strength and stress distributions are unknown. This case cannot be subjected to a demonstration test using standard statistical methods.

2. Conditions for Use

a. The strength distribution and stress distribution must be stochastically independent.

b. The strength distribution and stress distribution must be normal.

c. A random sample of strength and stress observations must be obtained.

3. Method

a. If the strength distribution has normal parameters μ_x , σ_x and the stress distribution has nor-mal parameters μ_y , σ_y , then the statistic

$$w = x - y$$

is normally distributed with parameters

$$w = \mu_x - \mu_y$$

$$\sigma_w = \sqrt{\sigma_x^2 + \sigma_y^2}$$

and the reliability is defined as the probability that w exceeds zero. Clearly, specifying a particular reliability is the equivalent of requiring the unit normal deviate

$$z = \frac{(\mu_x - \mu_y) - 0}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

to correspond to this reliability in the right tail of the unit normal.

Example

1. Stress has a specified value of 30 KSI* with standard deviation 1.5 KSI. Strength is expected to be in the vicinity of 40 KSI but the mean is unknown. The standard deviation is assumed to be 2.0 KSI. A reliability of 0.99 is acceptable while a reliability of 0.90 is unacceptable. The producer's risk is .05 and the consumer's risk .10.

Solution:

$$\sigma_w = \sqrt{2^2 + (1.5)^2}$$

$$= 2.5 \text{ KSI}$$

The unit normal deviates corresponding to 0.99 and 0.90 reliability are 2.33 and 1.28 respectively.

Therefore,

$$2.33 = \frac{(\mu_0 - 30) - 0}{2.5}$$

$$1.28 = \frac{(\mu_1 - 30) - 0}{2.5}$$

* KSI = 1000 lbs/sq. in

3. Method Example

and the requirements on the strength distribution are

$\mu_0 = 35.9$

$\mu_1 = 33.2$

with a known $\sigma = 2.0$, $\alpha = .05, \beta = .10$. The methods of Section 8.4.7.5.2.2, may now be used.

2. If we retain the data of example 1, and delete the information concerning the mean of the stress distribution, then,

$\sigma_x = 2.0 \quad \mu_0 - \mu_X$

$= 35.9 - 30 = 5.9$

$\sigma_y = 1.5 \quad \mu_1 - \mu_X$

$= 33.2 - 30 = 3.2$

$\alpha = .05$

$\beta = .10$

If N_x observations of strength and N_y observations of stress are obtained, the appropriate statistic is

$$z = \frac{(\overline{x} - \overline{y}) - 5.9}{\sqrt{\frac{\sigma_x^2}{N_x} + \frac{\sigma_x^2}{N_y}}}$$

3. Method | Example

Hence, the critical value of $(\bar{x} - \bar{y})$ is

$$z_\alpha \sqrt{\frac{\sigma_x^2}{N_x} + \frac{\sigma_y^2}{N_y}} + 5.9$$

For example, ten observations of strength and four observations of stress are available.

For 0.99 reliability, we have from the previous example, $\mu_x - \mu_y = 5.9$, and $z_\alpha = z_{.95} = -1.65$

$$-1.65\sqrt{\frac{4.0}{10} + \frac{2.25}{4}} + 5.9$$

$$= +4.21$$

as the critical value of the statistic $(\bar{x} - \bar{y})$. Accept if

$$\bar{x} - \bar{y} \geq 4.21$$

Otherwise, reject. The β risk for this example would be

$$z = \frac{4.21 - 3.2}{\sqrt{\frac{4.0}{10} + \frac{2.25}{4}}} = +1.03$$

$$\beta = 0.15$$

A larger sample size for either stress or strength will reduce β

8.4.7.7 Bayes Sequential Tests

1. When to Use

A test plan of this type can be specified if mean life θ is the parameter of interest and if a prior distribution on θ is known. The use of a test plan of this type results in a smaller sample size than most other test plans described in this section.

2. Conditions of Use

a. The lot of items being evaluated must have a known prior distribution on the mean life.

b. The parameters of the prior distribution must be specified as well as θ_1, the minimum acceptable mean life. It is necessary to specify two other terms K_2 and K_1 as criteria for terminating the test. K_2 is a probability such that if $Pr(\theta \geq \theta_1/\theta_n) \geq K_2$ the test is deemed passed. It is usually specified at .90, .95 or .99 and is the probability associated with a lower bound at θ_1. K_1 is usually specified as .01, .05, or .10 and $1 - K_1$ is the probability associated with an upper bound at θ_1. $K_2 + K_1$ need not equal 1.

c. In this demonstration test procedure it is possible to pass or fail without testing. If testing is called for, one item is tested at a time and a decision is made after each failure to either accept, reject, or continue testing.

3. Method	Example
a. Specify the prior distribution form, its parameters, and the quantities θ_1, K_1 and K_2	a. It has been found that a given item type has a prior distribution on its mean life θ that is inverted gamma with a shape parameter $\lambda = 3$, a scale parameter $\alpha = 100$, a minimum acceptable mean life $\theta_1 = 60$, $K_1 = .10$ and $K_2 = .90$.
b. Compute P_0 to determine if testing should be performed: if $P_0 \geq K_2$, accept and do not test if $P_0 \leq K_1$, reject and do not test if $K_1 < P_0 < K_2$, place an item on test	b. To solve for P_0 use the Tables of Percentage Points of the X^2 distribution for 2λ degrees of freedom (d.f.). In this case use 6 d.f. Next solve the equation $X^2 = \frac{2\alpha}{\theta_1} = \frac{2(100)}{60} = 3.33$ In the X^2 Table for 6 d.f. $X^2 = 3.33$ corresponds to a percentage point (P_0 in this problem) of approximately .23. Therefore, $K_1 < P_0 < K_2 = .10 < .23 < .90$ resulting in the instruction to begin testing.

3. Method

c. Construct a table of decision points for each failure time. This is done by solving for

$$\hat{\theta}_n * = \frac{\theta_1 X^2_{K\,2,\,2(n+\lambda)} - 2\alpha}{2n}$$

where n = # of failures

and

$$\hat{\theta}\, n* = \frac{\theta_1 X^2_{K1,\,2(n+\lambda)} - 2\alpha}{2n}$$

Example

c. For 1 failure the following decision points are calculated

$$\hat{\theta}_1 * = \frac{60X^2_{(.90,8)} - 2(100)}{2(1)}$$

$$\hat{\theta}_1 * = \frac{60(13.36) - 200}{2} = 301$$

$$\hat{\theta}_1 * = \frac{60X^2_{(.10,\,8)} - 2(100)}{2(1)}$$

$$\hat{\theta}_1 * = \frac{60(3.49) - 200}{2} = 4.7$$

The following table gives the accept/reject mean lives for additional failures.

n	$\hat{\theta}_n*$ Accept if θ_n $\hat{\theta}_n*$	$\hat{\theta}_n*$ Reject if θ_n $\le$ $\hat{\theta}_n*$
1	301	4.7
2	190	23.5
3	152	29.7
4	133	33.4
5	-	-

3. Method | Example

$\hat{\theta}$ n* and $\hat{\theta}_{n*}$ eventually terminate at some n. Therefore, the test could not continue indefinitely.

$$\text{The } \theta n = \frac{\sum_{i=1}^{n} t_i}{n}$$

where:
t = failure time

n = number of failures

d. Test the first part and make the decision to accept, reject or continue testing.

d. Test the first item. If its failure time is:

1) 4.7 hours or less, reject the product.
2) 301 hours or more, accept the product.
3) greater than 4.7 and less than 301, test another sample to failure compare again to the accept/reject criteria of Step c.

4. For Further Information

The theoretical development of this method is presented in “A Sequential Bayes Procedure for Reliability Demonstration,” by R.E. Schafer and N.D. Singpurwalla, Naval Research Logistics Quarterly, March 1970.

The methodology of fitting prior distributions is developed in RADC-TR-69-389 “Bayesian Reliability Demonstration - Phase I - Data for A Prior Distribution.” Further details are provided in RADC-TR-76-296, Vols. I through V, “Reliability Acceptance Sampling Plans Based Upon Prior Distribution,” and in RADC-TR-81-106, “Bayesian Reliability Tests Made Practical.”

8.4.8 Reliability Demonstration Summary

MIL-HDBK-781 covers the detailed requirements for development and production reliability tests for equipment that experiences a distribution of time-to-failure that is exponential. MIL-HDBK-781 contain: test conditions, procedures, and various fixed length and sequential test plans with respective accept/reject criteria. Refs. [5] and [12] provide additional guidance and details on reliability measurement. The reliability test plan should contain, as a minimum, the following information:

(1) How the equipment/system will be tested

- The specified test conditions, e.g., environmental conditions, test measures, length of test, equipment operating conditions, accept/reject criteria, test reporting requirements, etc.

(2) Who will perform the tests

- Contractor, Government, independent organization

(3) When the tests will be performed

- Development, production, field operation

(4) Where the tests will be performed

- Contractor's plant, Government organization

Section 8.4.7 presented step-by-step instructions on the use of various types of reliability demonstration test plans. Instructions and examples are given for the following test plans:

(1) Attributes Demonstration Tests
  - (a) Plans for Small Lots
  - (b) Plans for Large Lots
  - (c) Plans for Large Lots (Poisson Approximation Method)
  - (d) Attributes Sampling Using ANSI/ASQC Z1.4-1993
  - (e) Sequential Binomial Test Plans

(2) Variables Demonstration Tests
  - (a) Time Truncated Test Plans
    - (1) Exponential Distribution
    - (2) Normal Distribution
    - (3) Weibull Distribution

(b) Failure Truncated Tests
  (1) Exponential Distribution
  (2) Normal Distribution (Known)
  (3) Normal Distribution (Unknown)
  (4) Weibull Distribution

(c) Sequential Tests
  (1) Exponential Distribution
  (2) Normal Distribution

(d) Interference Demonstration Tests

(e) Bayes Sequential Tests

8.5 Reliability Growth

Experience has shown that programs which rely simply on a demonstration test by itself to determine compliance with the specified reliability requirements generally do not achieve the reliability objectives with the allocated resources. This is particularly true of complex systems. Generally, these systems require new technologies and represent a challenge to the state of the art. Moreover, the requirements for reliability, maintainability and other performance parameters are usually highly demanding. Consequently, striving to meet these requirements represents a significant portion of the entire acquisition process and, as a result, the setting of priorities and the allocation and reallocation of resources such as funds, manpower and time are often formidable management tasks.

In order to help ensure that the equipment/system will meet the required operational reliability requirement, the concept of reliability growth testing and management has been developed for equipment/system development programs.

8.5.1 Reliability Growth Concept

Reliability growth is defined as the positive improvement of the reliability of an equipment through the systematic and permanent removal of failure mechanisms. Achievement of reliability growth is dependent upon the extent to which testing and other improvement techniques have been used during development and production to "force out" design and fabrication flaws, and on the rigor with which these flaws are analyzed and corrected.

Figure 8.5-1 suggests an ideal growth process. The initial reliability of the prototype starts at some level that might be considered the state-of-the-art at the beginning of development. Through the development effort, reliability grows up to the pilot production stage. At that time, some loss of growth occurs due to the introduction of manufacturing problems. During the pilot production, corrective actions are continuing that cause resumption of growth. At the beginning of full scale production, some loss in the achieved level of reliability occurs because of the effects of mass production. However, growth will resume as these problems are eliminated. And, at a time when the equipment is released to the field it should have achieved the specified level or, under ideal conditions, the inherent or predicted level. The slope of this curve is affected by many variables and these will be discussed later. Thus, reliability growth is the result of an iterative design process. As the design matures, it is investigated to identify actual (via testing) or potential (via analysis) sources of failures. Further design effort is then spent on correcting these problem areas. The design effort can be applied to either product design or manufacturing process design. There are three essential elements involved in achieving reliability growth:

(1) Detection of failure sources (by analysis and test)

(2) Feedback of problems identified

(3) Effective redesign effort based on problems identified

The rate at which reliability grows is therefore dependent on how rapidly activities in this iterative loop can be accomplished, how real the identified problems are, and how well the redesign effort solves the identified problems. It is important to realize that some activities may act as a bottleneck. The bottleneck activities may vary from one development program to the next. Even within a single program they may vary from one stage of development to the next. In most cases, however, failure sources are detected through testing, and the testing process effectively controls the rate of growth. As a consequence, the reliability growth process becomes familiarly known as one of test, analyze, and fix (TAAF). However, the reliability achieved as a result of the growth process only becomes meaningful when the necessary changes developed and proven during TAAF to achieve that reliability are properly and fully incorporated in configuration-control documentation for production hardware.

Reliability growth testing (RGT) is only one aspect of a total reliability growth program. It must be accompanied by a reliability growth management program. This involves setting interim reliability goals to be met during the development testing program and the necessary allocation and reallocation of resources to attain these goals. A comprehensive approach to reliability growth management throughout the development program consists of planning, evaluating and controlling the growth process.

Note that RGT or TAAF, is intended neither to replace a sound design approach and thorough analytical effort nor compensate for a poor design. RGT should never be used or viewed as a "trial and error" approach to designing a reliable product.

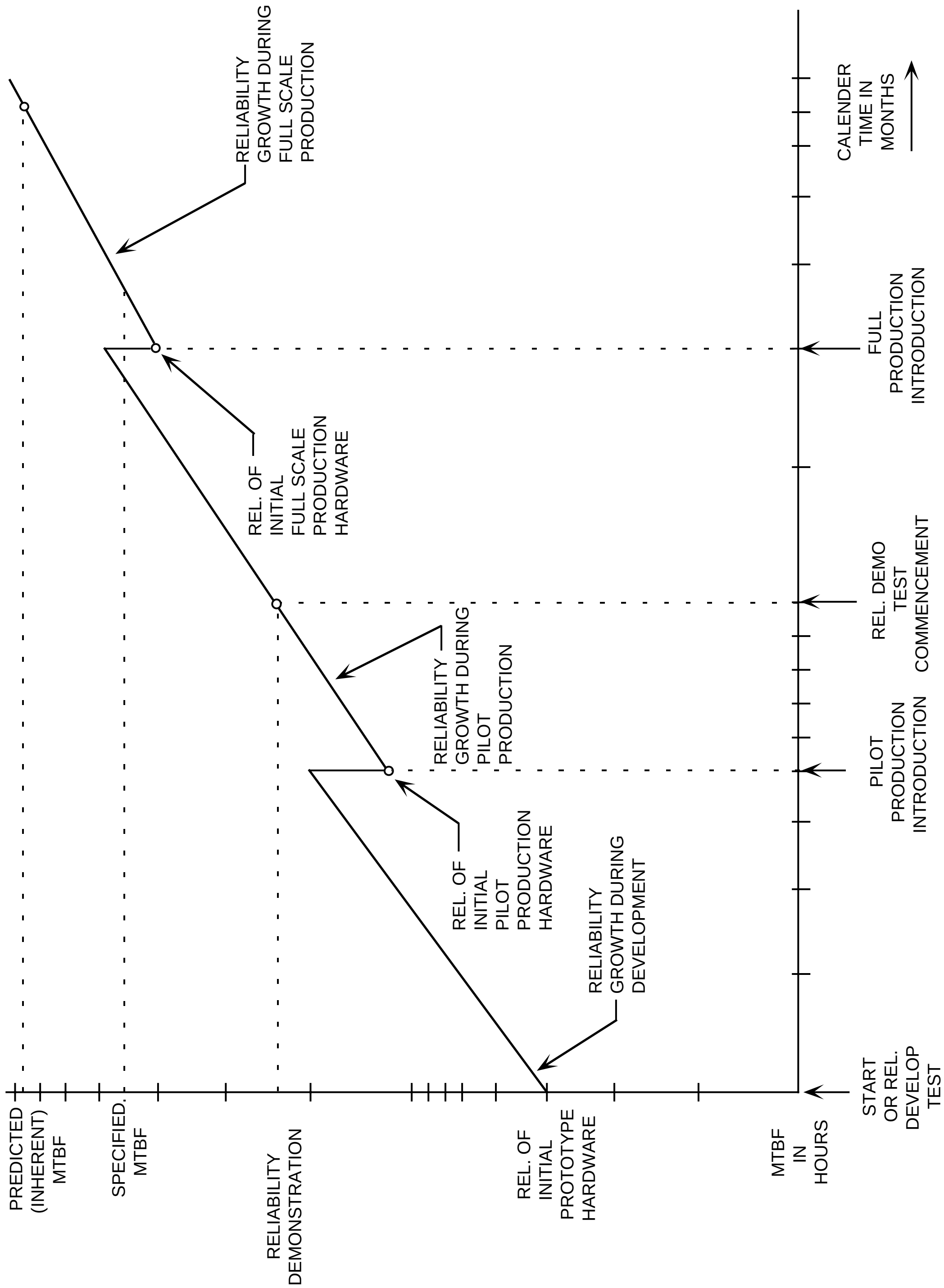


FIGURE 8.5-1: RELIABILITY GROWTH PROCESS

Reliability growth planning addresses program schedules, amount of testing, resources available and the realism of the test program in achieving the requirements. The planning is qualified and reflected in the construction of a reliability growth program plan curve. This curve establishes interim reliability goals throughout the program. To achieve these goals it is important that the program manager be aware of reliability problems during the conduct of the program so that he can effect whatever changes are necessary, e.g., increased reliability emphasis. It is, therefore, essential that periodic assessments of reliability be made during the test program (e.g., at the end of a test phase) and compared to the planned reliability growth values. These assessments provide visibility of achievements and focus on deficiencies in time to affect the system design. By making appropriate decisions in regard to the timely incorporation of effective fixes into the system commensurately with attaining the milestones and requirements, management can control the growth process.

8.5.2 Reliability Growth Modeling

For complex electronic/electro-mechanical avionic systems, the model used most often for reliability growth processes, and in particular reliability growth testing, is one originally published by J. T. Duane. (Ref. [16]). Essentially, this model provides a deterministic approach to reliability growth such that the system MTBF versus operating hours falls along a straight line when plotted on log-log paper. That is, the change in MTBF during development is proportional to T where T is the cumulative operating time and α is the rate of growth corresponding to the rapidity with which faults are found and changes made to permanently eliminate the basic causes of the faults observed.

The model is shown graphically in Figure 8.5-2, with each of the growth lines having different slopes, depending upon the emphasis given to the reliability growth program.

Duane's postulate was that as long as reliability improvement efforts continue, the following mathematical expression would hold:

$$\lambda_{\Sigma} = \frac{F}{H} = K H^{-\alpha} \qquad (8.23)$$

where:

- λ_{Σ} = cumulative failure rate
- H = total test hours
- F = number of failures, during time H
- K = constant determined by circumstances
- α = growth rate

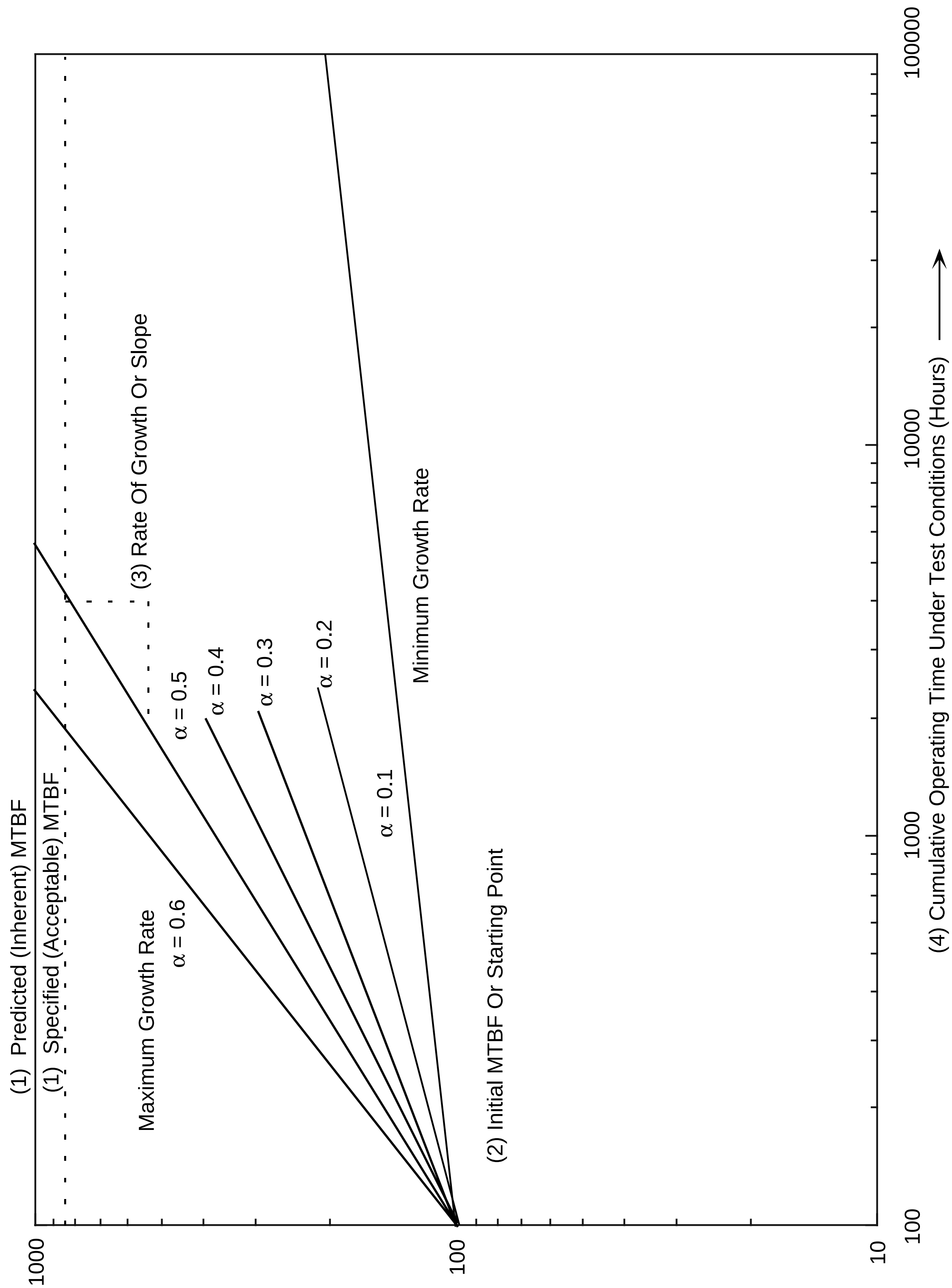


FIGURE 8.5-2: RELIABILITY GROWTH PLOTS

The original mathematical model was expressed in terms of cumulative failure rate; but currently, since equipment reliability is generally expressed in terms of MTBF, the following expression is used,

$$M_R = M_I \left(\frac{T_t}{t_i}\right)^{\alpha} \tag{8.24}$$

where:

M_R = required MTBF

M_I = initial MTBF

t_i = time at which initial data point is plotted (preconditioning time)

T_t = time at which the instantaneous MTBF of the equipment under test will reach the MTBF requirement

α = growth rate

Differentiating Eq. (8.23) with respect to time

Since $\frac{F}{H} = KH^{-\alpha}$

then $F = KH^{(1-\alpha)}$

The instantaneous failure rate is found by differentiating with respect to H (i.e., time).

$$\lambda_{\text{instantaneous}} = \frac{dF}{dH} = \frac{d(KH^{(1-\alpha 1)})}{dH} \tag{8.25}$$

$$= \frac{Kd(H^{(1-\alpha)})}{dH} = (1-\alpha)KH^{-\alpha}$$

so that the "instantaneous" or current failure rate is $(1-\alpha)$ times the cumulative failure rate, or the "instantaneous MTBF" is $\frac{1}{1-\alpha}$ times the cumulative MTBF. An adequate interpretation of "instantaneous MTBF" is ***The MTBF that the equipment currently on test would exhibit if we stopped the reliability growth and continued testing***.

Thus the "instantaneous" or current MTBF curves are straight lines displaced from the cumulative plot by a factor $\frac{1}{1-\alpha}$, which shows up as a fixed distance on a logarithmic plot, as shown in Figure 8.5-3.

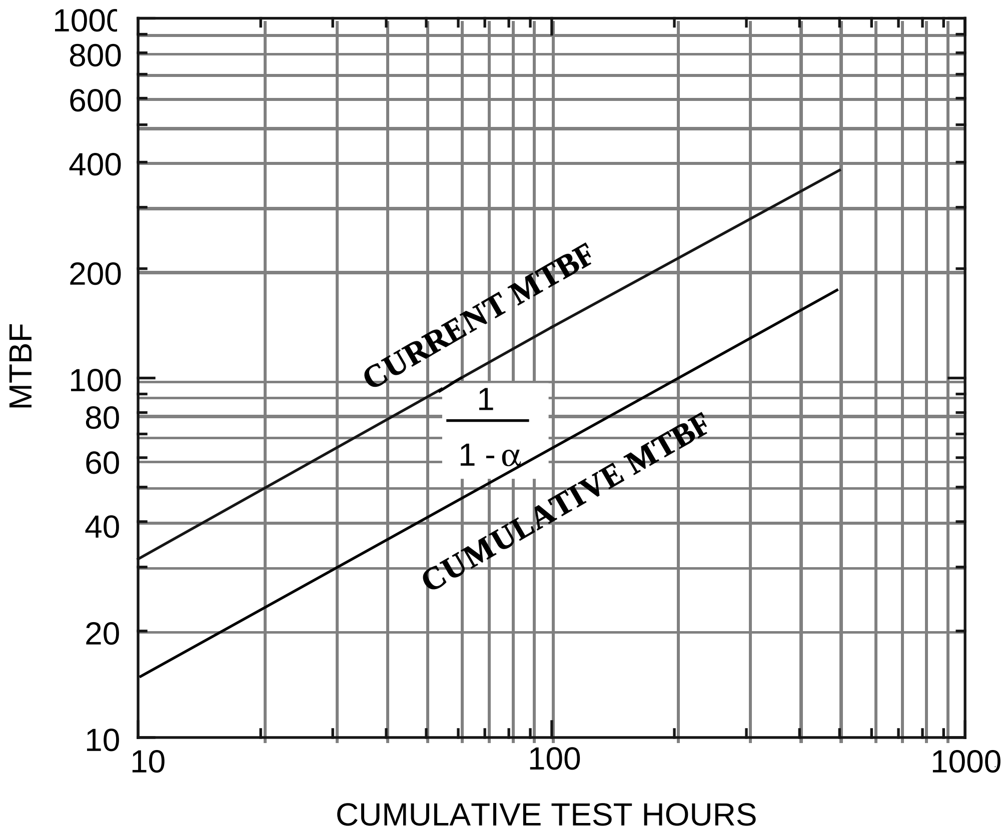


FIGURE 8.5-3: UP-IS-GOOD DUANE CHART WITH PLOT OF CURRENT MTBF

Normally, the cumulative MTBF (M_c) is measured in test and converted to instantaneous (or current) MTBF (M_I) by dividing by $1 - \alpha$, that is,

$$M_I = \frac{M_c}{1 - \alpha} \tag{8.26}$$

The cumulative MTBF is plotted versus cumulative test time, a straight line is fitted to the data and its slope, α, is measured. The current MTBF line is then drawn parallel to the cumulative line but displaced upward by an offset equal to $\frac{1}{1 - \alpha}$. The corresponding test time at which this line reaches the required MTBF is the expected duration of the growth test. Much evidence has been accumulated since Duane's original report that verifies the adequacy of the Duane Model in representing the real world of reliability growth testing.

In fact, recently the Duane Model has been successfully applied to software growth modeling (Ref. [18]).

Crow presents a formal mathematical development of the growth model. He showed that the failure rate during development follows the Weibull failure rate curve. The development which follows is similar to that given by Crow (Ref. [17]).

Mathematically, this model may be expressed by the equation

$$F(t) = \lambda t^{-\alpha} \qquad \lambda^{*} > 0; \quad 0 < \alpha < 1 \tag{8.27}$$

*λ is used here as a parameter of the Weibull distribution - it is not a failure rate.

where F(t) is the cumulative failure rate of the system at time t, λ and α are parameters. By definition, therefore, it follows that the cumulative failure rate is

$$F(t) = \frac{E(t)}{t} \tag{8.28}$$

where E(t) is the expected number of failures experienced by the system during t time units of development testing. Thus, from the above two equations

$$E(t) = \lambda t^{1-\alpha} \tag{8.29}$$

The instantaneous failure rate, r(t), is of the most interest for applications. It is defined as the change in the expected number of failures per unit time. For a nonexponential system, it varies with time while for an exponential system the failure rate is constant.

Differentiating E(t) with respect to time gives the instantaneous failure rate r(t) as follows:

$$r(t) = \frac{dE(t)}{dt} = (1-\alpha)\,\lambda\, t^{-\alpha} \tag{8.30}$$

By substituting in the previous equations

$$\beta = 1 - \alpha$$

one gets

$$r(t) = \lambda \beta t^{\beta - 1} \tag{8.31}$$

which is the Weibull failure rate function for a repairable system, i.e., for a non-homogeneous

Poisson process with a Weibull intensity function.

Thus, if one plans to use the Crow's Model, called the AMSAA Growth Model, during a development program, the Weibull failure rate function can be used to determine the failure rate at a particular development time t. The values of λ and β are estimated from test data. Since λ is only a multiplier and β determines how much the failure rate changes with the development time, β is referred to as the growth parameter. For the systems studied by Duane, a β of approximately 0.5 was estimated.

To gain further insight into the AMSAA Growth Model, consider Figure 8.5-4 which is a plot of the Weibull failure rate versus development time for β = 0.5 and λ = 0.4. In the early stages of development the failure rate decreases rather rapidly due to more failures and more rework going on during this time. As the development progresses, the rate of decrease of the failure rate drops off considerably. The AMSAA Model assumes that at some time t_0 which corresponds to about the time that development ends and production starts, the failure rate levels off to a fairly constant value. When the failure rate becomes constant, the time between failures can be described by the exponential distribution with a mean time between failure of

$$\text{MTBF}(t_0) = \left[\lambda\beta t_0^{\beta-1}\right]^{-1} \qquad (8.32)$$

Crow (Ref. [22]) has developed the maximum likelihood estimates (MLE) of β and λ and also a goodness-of-fit test to determine if the AMSAA Model fits a particular set of data. The MLE estimate for β is

$$\hat{\beta} = \frac{N}{\sum_{r=1}^{k}\sum_{i=1}^{n} N_r(t)\, ln \dfrac{T}{X_{ir}}} \qquad (8.33)$$

where:

k = number of different subsystems,

T = the operating time for each of the k subsystems,

$N_r(T)$ = number of failures observed for the r^{th} subsystem during T time,

X_{ir} = the age of the r^{th} subsystem at the i^{th} failure (initially at the beginning of development)

$N = \sum_{i=1}^{k} N_r(t)$ (Number of failures)

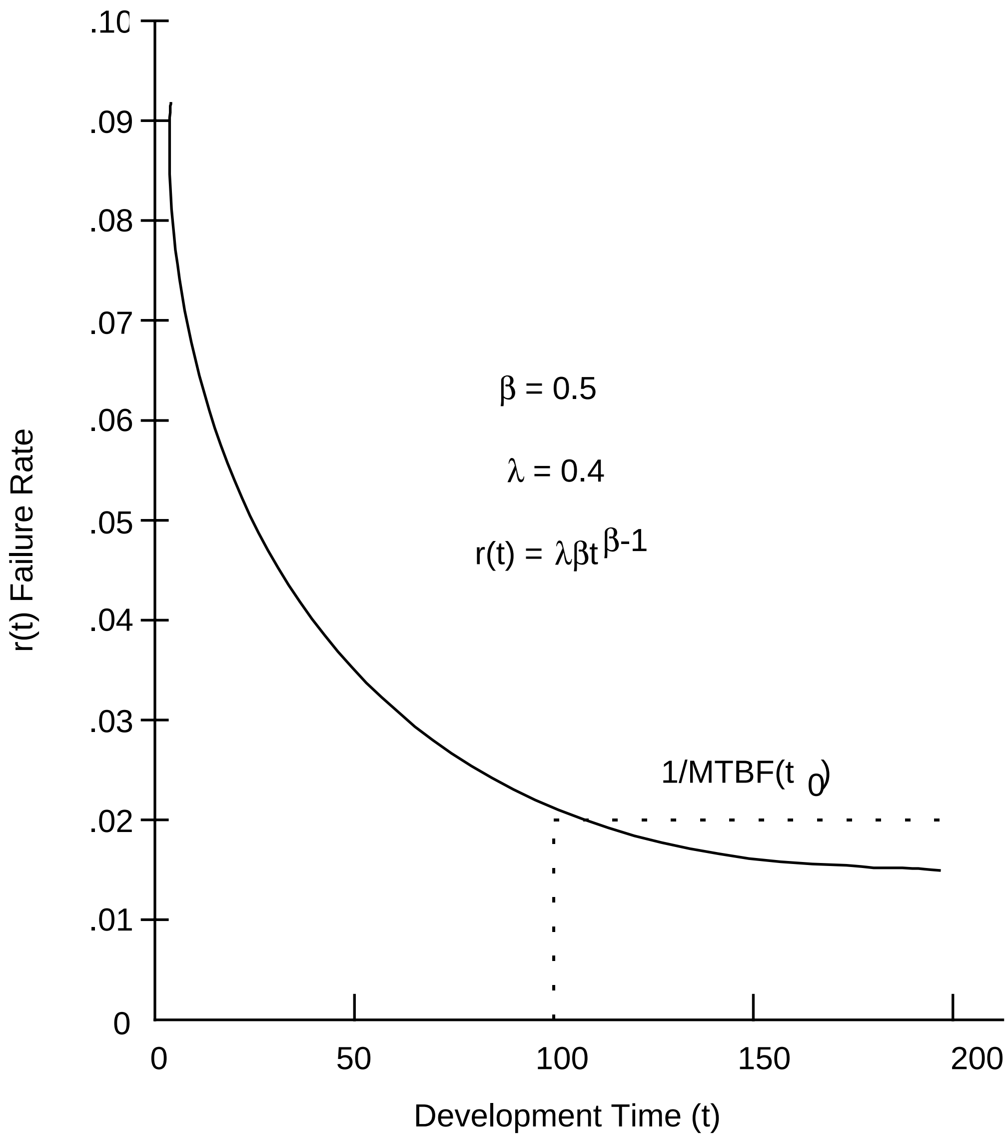


FIGURE 8.5-4: FAILURE RATE VS. DEVELOPMENT TIME
FOR WEIBULL FAILURE RATE

The previous MLE estimate of $\hat{\beta}$ is biased. The unbiased estimate is obtained by using

$$\bar{\beta} = \frac{N-1}{N}\hat{\beta} \tag{8.34}$$

The MLE of λ is

$$\hat{\lambda} = \frac{N}{kT^{\hat{\beta}}} \tag{8.35}$$

The chi-square goodness-of-fit test can be used to determine if the observed data fits the AMSAA Model. The chi-square statistic is calculated using

$$\chi_c^2 = \sum_{i=1}^{c} \frac{\left(O_i - E_i\right)^2}{E_i} \tag{8.36}$$

To compute the statistic the development time is divided into c intervals. The observed number of failures in the i-th interval, O_i, is obtained from the observed data. The expected number of failures in the i-th interval, E_i, is obtained using

$$E_i = \frac{N\left(t_i^{\bar{\beta}} - t_{i-1}^{\bar{\beta}}\right)}{T^{\bar{\beta}}} \tag{8.37}$$

where t_{i-1} and T_i are the beginning and ending times for the i^{th} interval. The χ_c^2 is compared with the tabled value of chi-square, χ_T^2 with degrees of freedom equal to c - 1 and the specified level of significance. If $\chi_c^2 < \chi_T^2$ then it can be concluded that the data fits the AMSAA Model.

8.5.2.1 Application Example

An engine system was analyzed for reliability growth using the AMSAA Model. The data available for analysis were based on 8063 hours of development testing. During this time there were 40 failures and the time of each failure was recorded. The average rates for this system during each interval of 1000 hours are shown in Figure 8.5-5.
Using this data the MLE's of λ and β, using equations 8.34 and 8.35, respectively, were

computed to be

$$\hat{\lambda} = 0.1279$$

$$\hat{\beta} = 0.6387$$

The unbiased estimate of β, using equation 8.35, is

$$\bar{\beta} = 0.6227$$

The chi-square goodness-of-fit statistic was calculated next using equation 8.36 and six intervals. The result was

$$\chi_c^2 = 10.09$$

Using a 1% level of significance and degrees of freedom of 6 - 1 = 5, the tabled value of chi-square is

$$\chi_T^2 = 15.086$$

Thus it can be concluded that the AMSAA Model fits the data.

Using the Eq. (8.31), the estimated failure rate for the engine becomes

$$r(t) = .128(.623)\, t^{.623-1}$$

$$= .08\, t^{-.377}$$

A plot of this failure rate curve is given in Figure 8.5-5. Notice the curve is beginning to flatten out. In fact it would take 100,000 hours of development time to get the failure rate down to .001 failures/hour.

Failure Times:

Interval 1 (0 to 1344 hours)	1, 43, 43, 171, 234, 274, 377, 530, 533, 941, 1074 ,1188, 1248
Interval 2 (1345 to 2688 hours)	2298, 2347, 2347, 2381, 2456, 2456, 2500
Interval 3 (2689 to 4032 hours)	2913, 3022, 3038, 3728, 3873
Interval 4 (4033 to 5376 hours)	4724, 5147, 5179
Interval 5 (5377 to 6720 hours)	5587, 5626
Interval 6 (6721 to 8064 hours)	6824, 6983, 7106, 7106, 7568, 7568, 7593, 7642, 7928, 8063

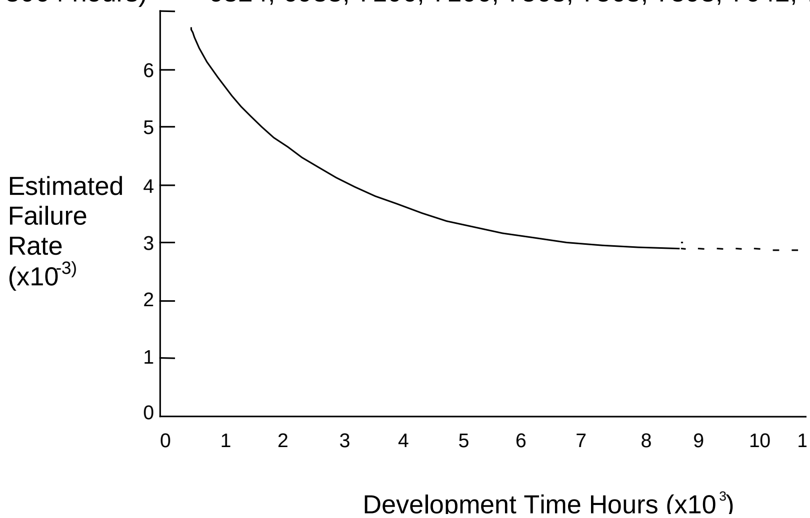


FIGURE 8.5-5: FAILURE RATE VS. DEVELOPMENT TEST TIME FOR WEIBULL FAILURE RATE

8.5.3 Comparison of the Duane and AMSAA Growth Models

The Duane Model and the Army Material Systems Analysis Activity (AMSAA) Model, developed by Dr. L. H. Crow in 1972 are the two most widely-used growth models. The Duane Model is based on an empirical relationship that holds as long as the MTBF is growing:

$$MTBF_{cum} = \frac{1}{K} T^{\alpha}$$

where:

$MTBF_{cum}$	=	Cumulative MTBF
K	=	Constant determined by the initial MTBF
α	=	Growth rate (the slope of the log-log plot of $MTBF_{cum}$ vs Test Time)
T	=	Cumulative test time

Typically the log-log plot of cumulative failures vs. test time will result in a linear relationship if the system reliability is improving. The test-analyze-and-fix (TAAF) procedure improves system reliability by the incorporation of design changes. If the slope of the best fit line of such a plot is positive, the system is said to be growing in reliability as time progresses.

The Duane Model assumes that the design fixes are 100% effective and that they are implemented immediately.

The instantaneous MTBF essentially estimates the projected field failure rate by accounting for fixes without purging the failure data.

The Duane Model assumes that growth is a deterministic process, while the AMSAA Model views the process of reliability growth as a probabilistic process. The AMSAA Model is based on the empirical relationship developed by Duane and is equivalent to a non-homogeneous Poisson process model with a Weibull intensity function. A typical AMSAA Model plot is shown in Figure 8.5-6. The AMSAA Model is

$$r_c(t) = \lambda t^{\beta-1}$$

where:

- $r_c(t)$ = The cumulative failure rate at time t
- t = Total test time
- β = Estimate of the time value of the growth parameter
- λ = Scale parameter

The instantaneous failure rate, $r_i(t)$, at time t is the incremental change in number of failures (F) with respect to the change in time.

$$\frac{F}{t} = r_c(t) = \lambda t^{\beta-1} \tag{8.38}$$

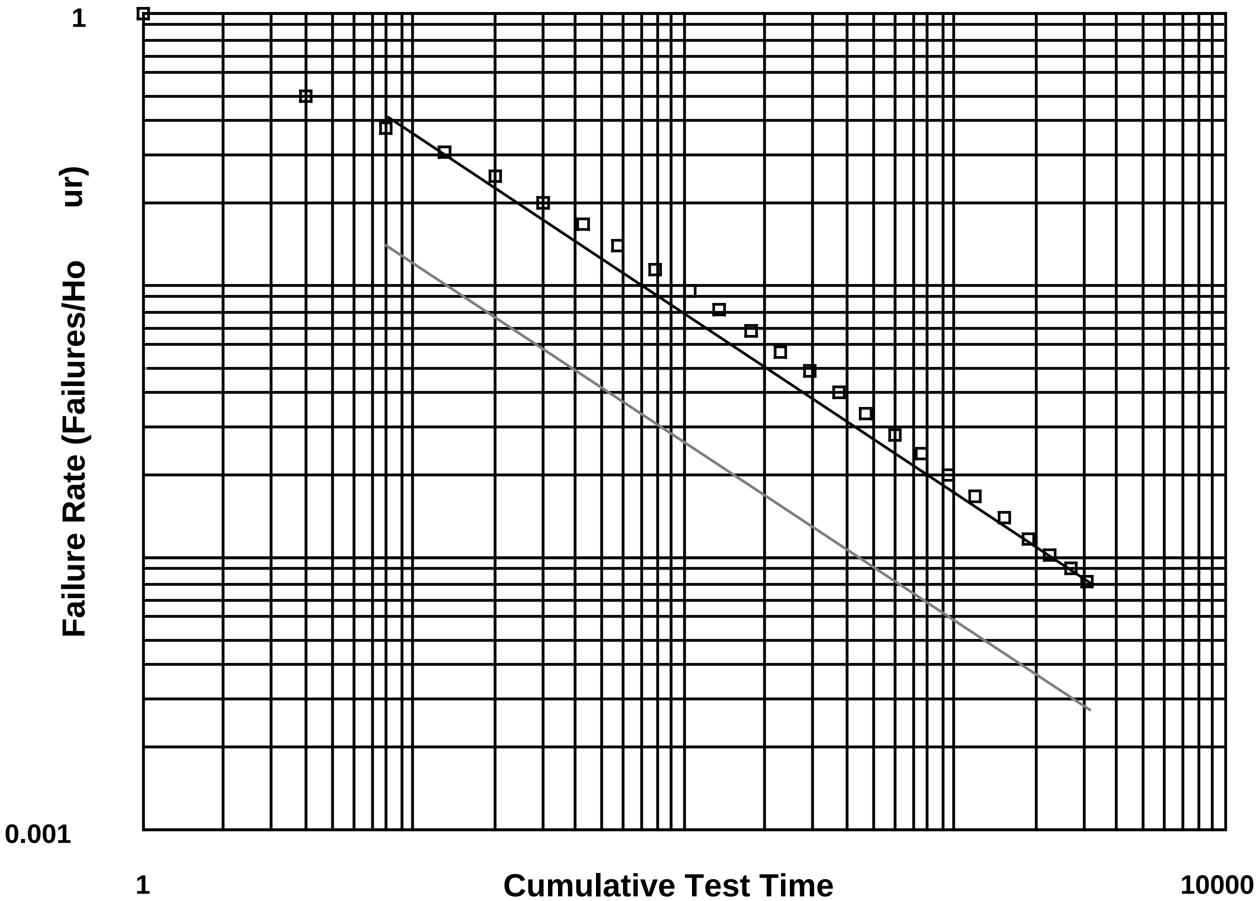
$$F = \lambda t^{\beta} \tag{8.39}$$

$$\frac{dF}{dt} = \lambda\beta t^{\beta-1} = r_i(t) \tag{8.40}$$

Therefore

$$\beta\, r_c(t) = r_i(t) \tag{8.41}$$

It can be seen that the parameter α used in the Duane Model is equivalent to $(1-\beta)$ of the AMSAA Model.



Growth Rate (β - 1)

β - 1 = -0.671

Instantaneous Failure Rate & MTBF

λ inst = 0.003 (F/Hr.)
MTBF inst = 3,777.164 (Hrs.)

FIGURE 8.5-6: RELIABILITY GROWTH ANALYSIS (AMSAA MODEL)

The Duane plot uses a least squares estimate of where the plot would fall while the AMSAA Model takes into account the exponential relationship between each data plot. Therefore in reliability growth plotting, the AMSAA Model tends to give a more accurate representation of the reduction in failure rate with respect to time. However, the Duane Model is typically used for program planning purposes even by proponents of the AMSAA Model because of its inherent simplicity.

8.5.3.1 Other Growth Models

Parametric models imply that there is a pattern to the growth, while nonparametric models allow the growth curve to "fall where it will." Because of this, only the parametric models are useful for mathematical descriptions of the generic or budgeted growth. Also, the nonparametric models generally do not allow growth projections to be made. However, either parametric or nonparametric models can be effectively used for controlling reliability growth.

Another consideration is the type of failure distribution that the growth model assumes. Many of the models treat the failure distribution in a nonparametric fashion. However, some models are based specifically on the assumption that the failure distribution is exponential.

Finally, although some of the models utilize a continuous time scale, others utilize a discrete scale, implying that the testing is performed in stages.

Although the Duane and the AMSAA reliability growth models have been the most widely used, a number of other models, both discrete and continuous, have been proposed in the literature.

8.5.4 Reliability Growth Testing

Reliability growth testing is the formal process of testing an equipment under natural and induced environmental conditions to discover and identify latent failure modes and mechanisms whose recurrence can be prevented through implementation of corrective action, thus causing the growth of equipment reliability.

These tests are conducted during the development phase on samples which have completed environmental tests prior to production commitment and do not replace other tests described in the contract or equipment specification. MIL-HDBK-781 contains the details on reliability growth test requirements, methods and procedures for application to electronic equipment.

8.5.4.1 When Reliability Growth Testing is Performed

The formal reliability growth test is usually performed near the conclusion of full scale development, concurrent with or after successful completion of environmental qualification testing and prior to reliability qualification (demonstration) testing. Although all testing should be viewed and planned as contributing to reliability growth, the formal test program dedicated to reliability growth is normally deferred until after environmental qualification, when the design of the equipment reflects the anticipated configuration and manufacturing processes to be used in production, but prior to commitment to production. The hardware to be tested should have all significant fixes required as a result of environmental qualification testing incorporated before initiating the reliability growth test. The reliability growth test must be successfully concluded, and all significant fixes incorporated in the test hardware prior to initiating the reliability qualification (demonstration) test. The reliability growth test is for the purpose of detecting reliability problems after all performance design and environmental problems have been resolved. The reliability qualification (demonstration) test discussed in Section 8 is for the purpose of proving reliability.

8.5.4.2 Reliability Growth Approach

The MIL-HDBK-781A (Ref. [18]) approach to reliability growth is patterned after the Duane and the AMSAA Models. With the Duane Model the change in MTBF during development is proportional to T^{α} where T is the cumulative operating time and "α" is the rate of growth corresponding to the rapidity with which faults are found, and changes are made to permanently eliminate the basic causes of the faults observed.

In order to structure a growth test program (based on the Duane Model) for a newly designed system, a detailed test plan is necessary. This plan should describe the test-analyze-fix concept, and show how it will be applied to the system under development. The plan should incorporate the following:

(a) Values for specified and predicted (inherent) reliabilities. Methods for predicting reliability (model, data base, etc.) should also be described.

(b) Criteria for reliability starting points, i.e., criteria for estimating the reliability of initially fabricated hardware, should be determined. For avionics systems, the initial reliability for newly fabricated systems has been found to vary between 10% and 30% of their predicted (inherent) values.

(c) The reliability growth rate (or rates) should be defined. To support the selected growth rate, the rigor with which the test-analyze-fix conditions are structured should be completely defined.

(d) Calendar time efficiency factors, which define the relationship of test time, corrective action time and repair time to calendar time, should be determined.

Note that each of the factors listed above impacts the total time (or resources) which should be scheduled to grow reliability to the specified value. Figure 8.5-2 (repeated here as Figure 8.5-7) illustrates the concepts described above.

In addition, Figure 8.5-7 graphically depicts the four elements needed to structure and plan a growth test program such as is described above. These four elements are further described as follows:

(a) Inherent Reliability: Represents the value of design reliability estimated during prediction studies, which may correspond to the value above that specified in procurement documents. Ordinarily, the contract specified value of reliability is somewhat less than the inherent value. The relationship of the inherent (or specified) reliability to the starting point greatly influences the total test time.

(b) Starting Point: Represents an initial value of reliability for the newly manufactured hardware. This usually falls within the range of 10% to 30% of the inherent or predicted reliability. Estimates of the starting point can be derived from prior experience or are based on percentages of the estimated inherent reliability. Starting points should take into account the amount of reliability control exercised during the design program and the relationship of the system under development to the state-of-the-art. Higher starting points, when justified, minimize test time.

Determination of the starting point is often difficult, with little documented guidance available. The following prioritized list provides the recommended procedures for establishing the starting point.

(1) Use actual data on early design

(2) Use the results of past reliability growth test and reliability prediction results

(3) Compute the default ratio (i.e., 10%) of the initial MTBF divided by the MTBF prediction

The first option is to use actual reliability data (i.e., failures, test time) recorded on the system during its early life. The design team necessarily tests the early design as a natural part of the design/development process. This testing is often informal with little standardized or documented reliability reports/data. Nevertheless, this type of data typically exists and it is most indicative of the actual MTBF of the system prior to reliability growth testing. The initial MTBF is computed as the cumulative amount of test time divided by the cumulative number of failures. To obtain this type of data and apply it to develop a planned reliability growth curve, requires a high degree of cooperation and sharing of information between the various engineering disciplines at an organization.

In many instances, this first option is not viable because the requisite data simply cannot be retrieved or the planned growth curve is needed as part of a proposal or early design document before any design activities take place.

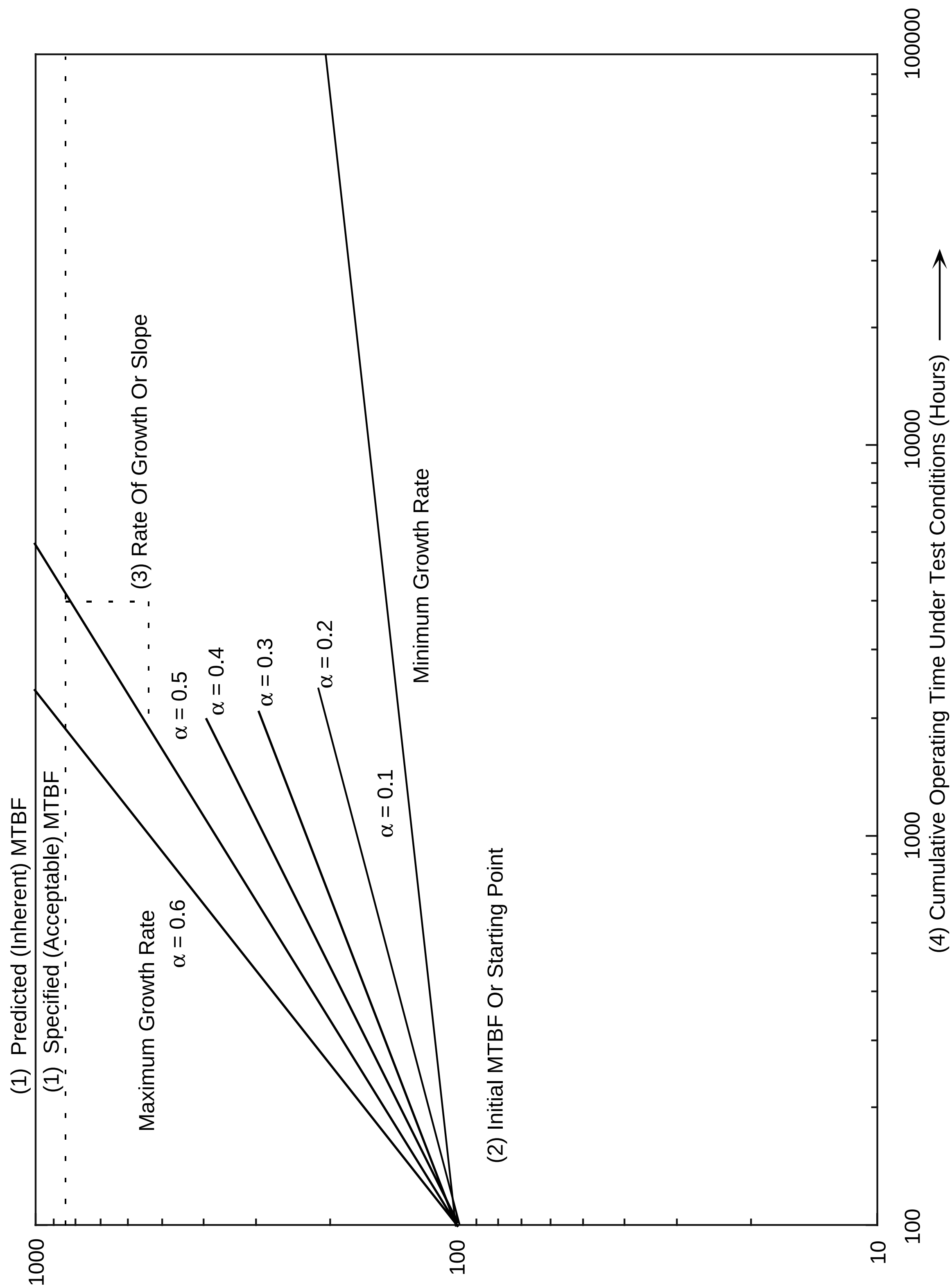


FIGURE 8.5-7: RELIABILITY GROWTH PLOTS

The second recommended option involves the use of results of prior reliability growth tests. These results must be from the same organization and indicative of the TAAF philosophy to be enacted. The degree of reliability growth or growth potential is not an inherent design parameter but is dependent on the management philosophy adopted for reliability growth testing. An aggressive management philosophy which is dedicated to seeking out the root-cause of failure and determining effective design fixes will be much more successful than testing programs with a less aggressive approach.

The following example indicates how to use this past data on reliability testing. A 250 hour pre-conditioning period was assumed to determine the actual starting point. It is important to distinguish between the planned and the actual MTBF starting point. Once the test has been conducted and the actual data are available, an actual starting point can be computed, which may differ from what was planned.

MTBF Prediction (MIL-HDBK-217)	Final Test MTBF ($MTBF_{inst}$ at Test Conclusion)	Initial MTBF (at 250 hours)	Initial MTBF/ MTBF Prediction
2,200	2,000	410	.19
800	610	84	.11
1,000	1,100	90	.09
920	830	220	.24
1,550	1,400	310	.20

It is necessary to compute the ratio of the initial MTBF (at the assumed 250 hour pre-conditioning period) divided by the MTBF prediction per MIL-HDBK-217. In the example, the ratio ranges from .09 to .24. In practice, it has been found that these ratios typically range from .10 to .30. In the example, the average ratio is .17.

The next step is to multiply the computed ratio by the MTBF prediction. If the equipment to undergo the reliability growth test has an MTBF prediction of 5,000 hours, then the estimated starting point would be,

$$MTBF_{\text{starting point}} = (.17)(5{,}000) = 850 \text{ hours}$$

The final and least preferred option is to apply a default ratio of .10. It has been found that use of this ratio yields a conservative estimate of the starting point. It needs to be recognized that this estimate is not precise; however, it provides a starting point if no other approach is viable. Again, using the 5,000 hour MTBF estimate, the starting point would be,

$$MTBF_{\text{starting point}} = (.10)(5{,}000) = 500 \text{ hours}$$

(c) Rate of Growth: Is depicted by the slope of the growth curve. This, in turn, is governed by the amount of control, rigor, and efficiency by which failures are discovered, analyzed, and corrected through design and quality action. Test programs which foster the discovery of failures, coupled with management supported analysis and timely corrective action, will result in a faster growth rate and consequently less total test time.

(d) Calendar Time/Test Time: Represents the efficiency factors associated with the growth test program. Efficiency factors include repair time, and operating/nonoperating time as they relate to calendar time. Lengthy delays for failure analysis, subsequent design changes, implementation of corrective action or short operating periods will extend the growth test period.

Figure 8.5-7 shows that the value of the parameter "α" can vary between 0.1 and 0.6. A growth rate of 0.1 can be expected in those programs where no specific consideration is given to reliability. In those cases, growth is largely due to solution of problems impacting production, and from corrective action taken as a result of user experience. A growth rate of 0.6 can be realized if an aggressive, hard-hitting reliability program with management support is implemented. This type of program must include a formal stress-oriented test program designed to aggravate and force defects and vigorous corrective action.

Figure 8.5-7 also shows the requisite hours of operating and/or test time and the continuous effort required for reliability growth. It shows the dramatic effect that the rate of growth has on the cumulative operating time required to achieve a predetermined reliability level. For example, Figure 8.5-7 shows, for an item product whose MTBF potential is 100 hours, that 100,000 hours of cumulative operating time is required to achieve an MTBF of 200 hours when the growth rate is 0.1. And, as previously stated, a 0.1 rate is expected when no specific attention is given to reliability growth. However, if the growth rate can be accelerated to 0.6 (by growth testing and formal failure analysis activities) then only 300 hours of cumulative operating time is required to achieve an MTBF of 200 hours.

Some general guidance on reliability growth test time is as follows:

> Fixed-length test times of 10 to 25 multiples of the specified MTBF will generally provide a test length sufficient to achieve the desired reliability growth for equipment in the 50 to 2000 hour MTBF range. For equipments with specified MTBFs over 2000 hours, test lengths should be based on equipment complexity and the needs of the program, but as a minimum, should be one multiple of the specified MTBF. In any event, the test length should not be less than 2000 hours or more than 10,000 hours.

Where time is not an appropriate measurement parameter for the particular hardware, the Duane Model is adaptable to other measurement parameters such as cycles, events, rounds, etc.

8.5.4.3 Economics of Reliability Growth Testing

The purpose of reliability growth testing is simple: to save money during the planned service life of the equipment. Experience has shown that an investment in assuring that specified reliability is, in fact, achieved prior to production will result in significantly-reduced life-cycle costs over the planned service life of the equipment due to savings realized by fewer maintenance actions, fewer required spares, and less handling damage, among others. This relationship is illustrated in Figure 8.5-8.

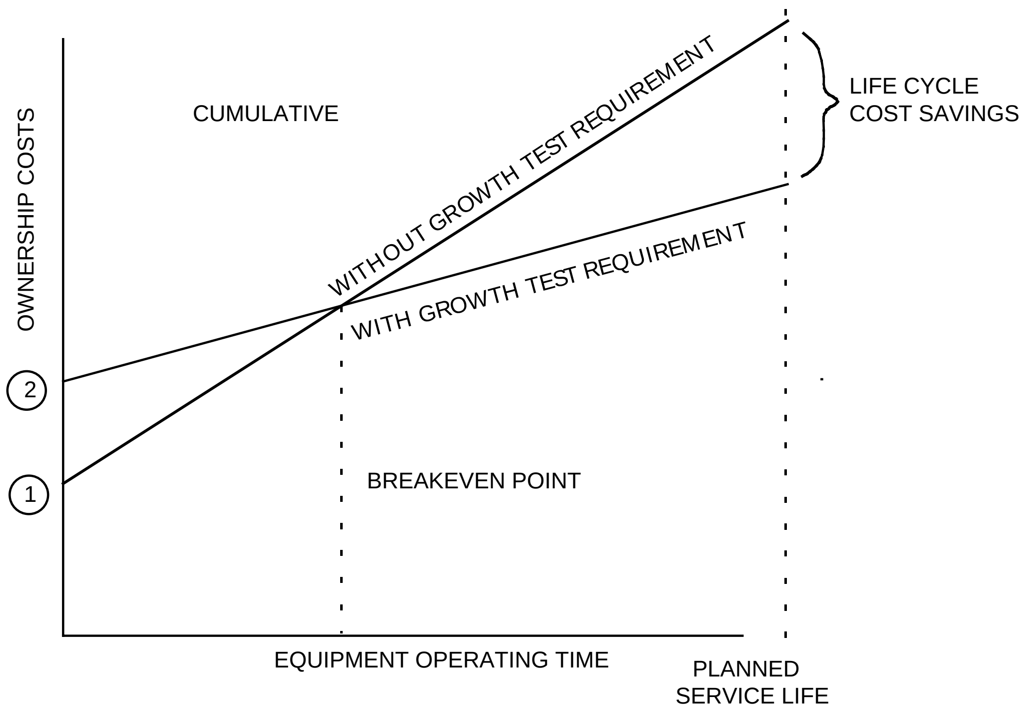


FIGURE 8.5-8: COMPARISON OF CUMULATIVE LIFE CYCLE COSTS WITH AND WITHOUT SPECIFIED RELIABILITY GROWTH TEST REQUIREMENTS

Point (1) represents the acquisition cost of an equipment without a reliability growth test requirement and a delivered MTBF (based on post-production experience) considerably less than the specified MTBF for that equipment. The cumulative cost of ownership rises with equipment operating time to account for equipment repairs and spares support over the life of the equipment.

Point (2) represents the acquisition cost of the same equipment, with the added cost of the reliability growth test program to achieve specified MTBF as a delivered MTBF. The cumulative cost of ownership with equipment operating time increases at a slower rate than the previous case due to less frequent repairs and reduced spares support requirements until a breakeven point is reached. At this point the growth test program has paid for itself and the difference in costs due to the reliability growth program represents a life-cycle cost savings.

8.5.5 Reliability Growth Management

Reliability growth management is the systematic planning for reliability achievement as a function of time and other resources and is used for controlling the ongoing rate of achievement by reallocation of resources based on comparisons between planned and assessed reliability values.

Reliability growth management is part of the system engineering process. It does not take the place of the other basic reliability program activities such as predictions, apportionment, failure mode and effect analysis, and stress analysis. Instead, reliability growth management provides a means of viewing all the reliability program activities in an integrated manner.

It is imperative to recognize that a total reliability program is needed for effective reliability growth management. While it is generally recognized that reliability will grow in the presence of a reliability program, reliability growth planning provides an objective yardstick and an orderly means of measuring progress and directing resources so that reliability requirements may be achieved in a timely and cost effective manner. A good reliability growth plan can greatly improve the chances of achieving total reliability program objectives. However, it is not intended to be the total reliability program.

MIL-HDBK-189 provides procuring activities and development contractors with an understanding of the concepts and principles of reliability growth, advantages of managing reliability growth, and guidelines and procedures to be used in managing reliability growth. It should be noted that this Handbook is not intended to serve as a reliability growth plan to be applied to a program without any tailoring. The Handbook, when used with knowledge of the system and its development program, will allow the development of a reliability growth management plan that will aid in developing a final system that meets its requirements and lowers the life cycle cost of the fielded systems.

8.5.5.1 Management of the Reliability Growth Process

There are innumerable ways in which reliability can grow during development. There are, of course, only a finite number of reliability growth models available. Consequently, acquisition managers cannot conduct their development programs in just any fashion, and have an existing reliability growth model available for estimation and prediction purposes. The manner in which the development program is managed and the choice of the reliability growth model are, therefore, dependent. Essentially, there are two ways by which acquisition managers can evaluate the reliability growth process.

(a) They may monitor the various reliability oriented activities (FMEA's, stress analysis, etc.) in the growth process to assure themselves that the activities are being accomplished in a timely manner and that the level of effort and quality of work is appropriate. This is a qualitative approach.

(b) They may utilize assessments (quantitative evaluations of the current reliability status) that are based on information from the detection of failure sources.

The assessment approach is, preferable in that it is results-oriented, in the form of quantitative estimates of planned and achieved reliability as the program progresses.

Figure 8.5-9 illustrates how assessments may be used in controlling the growth process. One of the more important points to emphasize is that assessments have been a way of life in reliability work for many years, as have the resultant decisions.

What, then, is new about reliability growth management? What is new is a formal standard against which the assessment may be compared. The fact that managers in the past have made decisions based on assessments implies that they had at least a subjective standard of acceptable reliability growth against which to make comparison. A formal, objective standard has the advantage of remaining constant, unless formally changed, rather than bending in the hope that "tomorrow will be better."

Figure 8.5-10 illustrates an example of a reliability growth curve, showing both the budgeted (planned) reliability growth and assessments. A comparison between the assessment and the budgeted value will suggest whether the program is progressing as planned, better than planned, or not as well as planned. Based upon the first two data points of assessed growth, the decision would probably be made to continue development with no changes. If reliability progress is falling short, as the two subsequent assessed data points indicate, new strategies should be developed. These strategies will probably involve the reassignment of resources to work on identified problem areas. They may, as a last resort, result in adjustment of the time frame, or relaxation of the original requirement.

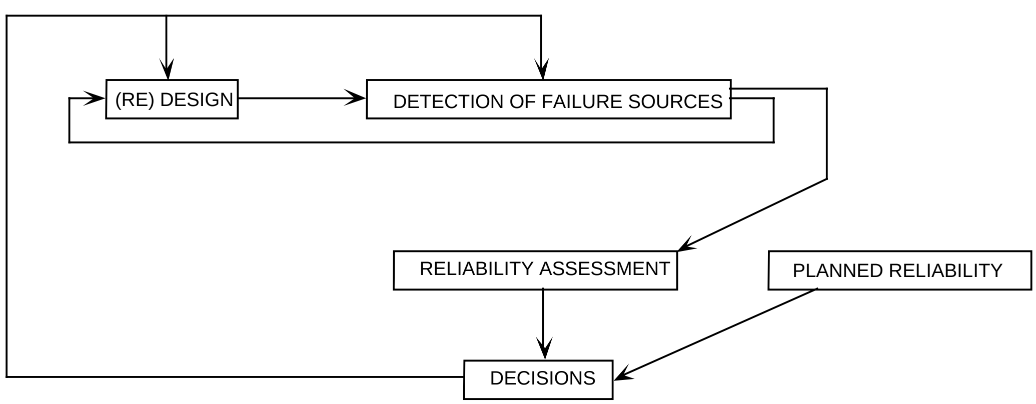


FIGURE 8.5-9: RELIABILITY GROWTH MANAGEMENT MODEL (ASSESSMENT)

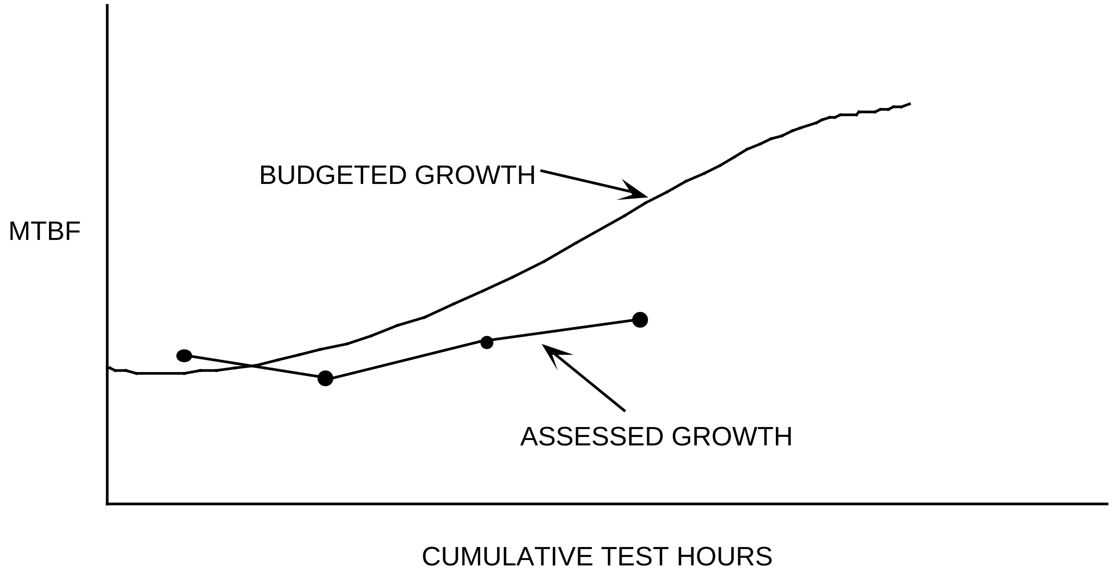


FIGURE 8.5-10: EXAMPLE OF A RELIABILITY GROWTH CURVE

8.5.5.2 Information Sources That Initiate Reliability Growth

The detection of failure sources is the activity that effectively initiates the growth process by pointing the way for redesign. Because the information sources that are used for detecting failure sources are so varied and because they can be relied on at different times during the life cycle, great program flexibility is possible. Although the total number of information sources that can be used to initiate reliability growth is rather large, they can be grouped into five categories: external experience, analysis, tests, production experience, and operational experience.

(a) External Experience. This is information generated outside the specific development program which has applicability within the program. Examples of this type of information are historical data, publications, technical experience of personnel, and information from currently operating systems.

(b) Analysis. This is information generated within the specific development program, excluding the test of hardware. Examples are feasibility studies, probabilistic reliability design, failure mode and effect analysis, and design reviews.

(c) Tests. Although this source of information is self-explanatory, the various ways in which testing is performed are important considerations. The hardware may be in any level of maturity, ranging from breadboard to final production configurations. Various levels of assembly may be tested, ranging from components to system level. Finally, the environmental conditions can vary all the way from testing under ambient conditions to overstress or accelerated testing. Testing is the most common source of information for initiating growth; it is the source usually modeled because it yields objective measurements.

(d) Production Experience. The production process itself may identify weak areas in the design.

(e) Operational Experience. The use of fielded systems will identify design deficiencies which point the way toward reliability growth.

8.5.5.3 Relationships Among Growth Information Sources

The chronological relationship of these information sources is illustrated in Figure 8.5-11. This figure illustrates that growth is at least possible at any point in the life cycle. However, what are the relative merits of growing reliability at these various points? To a large extent, this question can only be answered with respect to a specific development program. But there are two fundamental considerations that must be made. First, changes can be accomplished very economically early in the life cycle. The example usually given is that a change which would cost $1 on the drawing board will end up costing about $100 if it is made after the equipment is fielded. Therefore, it is desirable to grow reliability as early as possible. However, the information upon which early changes are based tends to contain many unknown factors, such as operational conditions and component interactions. Second, changes which are made later in the life cycle tend to be better directed, as there are fewer unknowns in the information as hardware maturity nears. The two desired characteristics will be referred to as "timeliness" and "credibility."

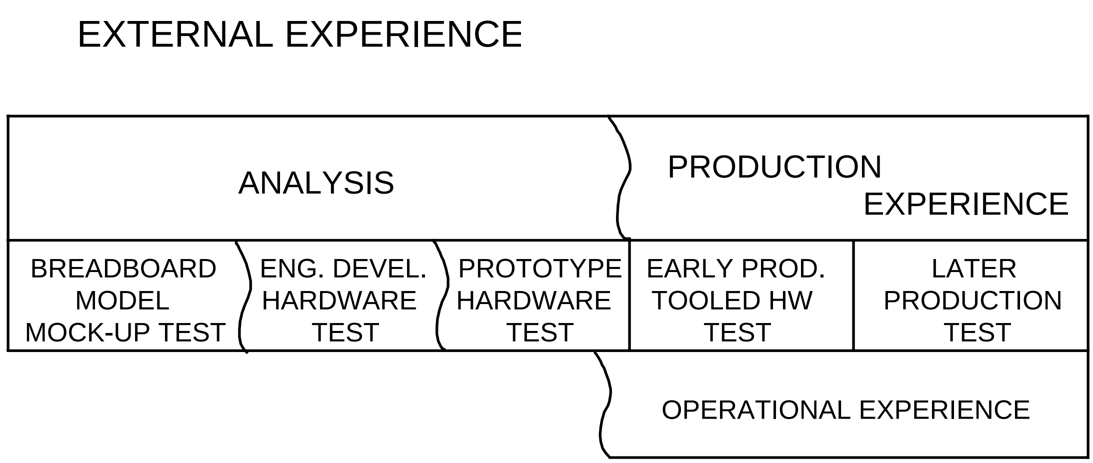


FIGURE 8.5-11: INFORMATION SOURCES THAT INITIATE RELIABILITY GROWTH

Depending on the characteristics of the specific program and system, it may be desirable to place particular emphasis on certain combinations of these information sources. In effect, we would like to achieve a reasonable combination of timeliness, credibility, and economy. The following paragraphs give some suggestions about when it may be desirable to place emphasis on various types of information sources. The rationale that is given here could serve as a basis for a more formal economic model for specific applications. The suggestions that are given here are intended to point out those information sources which have the strongest potential under varying situations. A good program would probably utilize all of the information sources to some degree, but the mix and emphasis will vary from one program to the next.

(a) Reliability Growth Through External Experience. The strongest feature of external experience is that it may be available at the very beginning of the life cycle, thus emphasizing timeliness. This, of course, assumes that appropriate external experience is available.

(b) Reliability Growth Through Analysis. Analysis becomes particularly valuable when the system reliability is high, mainly because the next best alternative, testing, will tend to be time-consuming and, therefore, expensive. However, in order to be able to rely heavily on analysis, much detailed knowledge is necessary. The operation of the system must be well understood. This implies that the development must be reasonably within the state-of-the-art. There must be good, detailed knowledge of the environment and use conditions. Finally, appropriate design analysis techniques must either be available or specially developed and there must be a good information base to support these techniques. Many reliability programs put too little emphasis on analysis and the associated information base. One problem with a reliance on analysis is that the effects cannot be measured objectively.

(c) Reliability Growth Through Testing. Reliability growth models are generally based on test results. Therefore, testing is a very important information source for initiating reliability growth. Testing will have the greatest payoff if many failures are encountered which can be thoroughly analyzed. Therefore, a low system reliability and an inability to perform failed part analysis suggest strong emphasis be placed on testing. One other factor which must be considered is the cost of testing itself. High test costs may discourage strong reliance on testing to achieve growth. However, generally there is no valid substitute for a good test program in the reliability growth process.

(d) Reliability Growth Through Production Experience. The production process and its quality controls are major contributors to reliability. In fact, a drop in reliability during the transition from development to production is a common phenomenon. It then becomes necessary to grow reliability based on manufacturing process redesign and/or better quality controls. Many process and control problems can be eliminated during the production phase through the use of process capability studies, worst-case analyses, and similar producibility-related techniques. Moreover, it is unlikely that all process and control problems could be eliminated during pre-production; and almost certainly, the payoff from these techniques, expressed as a function of effort, would show a diminishing-returns pattern. It is almost inevitable that some problems can be more cost-effectively eliminated after production starts, particularly when the production run is relatively long and the tooling is relatively inexpensive.

(e) Reliability Growth Through Operational Experience. Although some reliability growth through operational experience is inevitable, this method is the least desirable of the five sources listed. Improving reliability through retrofitting of fielded systems often costs up to a hundred times as much as the same change made on the drawing board.

8.6 Summary of the Differences Between Reliability Growth Testing and Reliability Demonstration Testing

Reliability growth is the result of an iterative design process. As the design matures, it is investigated to identify actual (via testing) or potential (via analysis) sources of failures. Further design effort is then spent on correcting these problem areas. The design effort can be applied to either product design or manufacturing process design. There are three essential elements involved in achieving reliability growth:

(1) Detection of failure sources (by analysis and test)

(2) Feedback of problems identified

(3) Effective redesign effort based on problems identified

Reliability demonstration tests, on the other hand, are designed for the purpose of proving, with statistical confidence, a specific reliability requirement; not specifically to detect problems, or to grow reliability. The test takes place after the design is frozen and its configuration is not allowed to change. However, in practice, some reliability growth may occur because of the deferred correction of failures observed during the test.

Reliability demonstration is specified in most military system procurement contracts and involves, in many instances, formal testing. Demonstration tests are normally conducted after development has been completed but before high rate production has been initiated. Demonstration tests are normally conducted after growth tests in the development cycle using initial production hardware.

As previously indicated, reliability demonstration testing, carries with it certain statistical confidence levels, and the more demonstration testing, the more confidence. The more reliability growth testing that is performed, the higher the actual reliability. Depending on program funding and other constraints, system testing may follow one of two options. The first option maximizes growth testing and minimizes demonstration testing resulting in a high MTBF at a low confidence. Option two minimizes reliability growth testing with a resultant lower MTBF at higher confidence. These concepts are shown graphically in Figure 8.6-1.

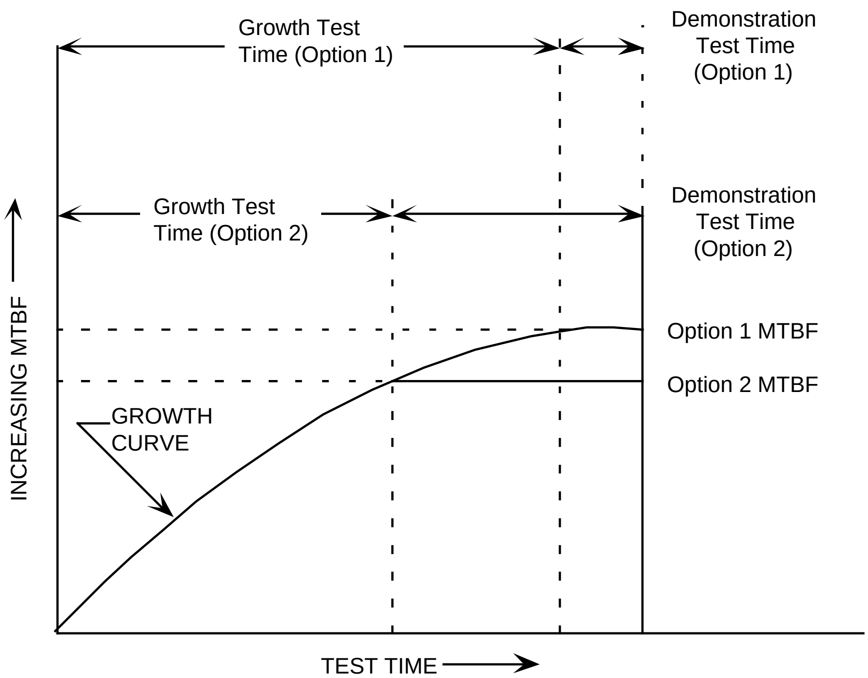


FIGURE 8.6-1: RELIABILITY TESTING OPTIONS

8.7 Accelerated Testing

Although accelerated testing is commonly used today, it frequently means different things to different people. There are potentially two main reasons for performing an accelerated test. These are: a) life estimation or b) problem/weakness identification (or confirmation) and correction. The difference between these reasons, although subtle, can have a very significant impact upon the underlying assumptions upon which the test is based, the models utilized in constructing the test, the test equipment and chambers used, the way in which the test itself is conducted, and the manner in which the resulting data is analyzed and interpreted.

Accelerated Life Testing is the means by which length of life can be determined. Here the primary focus is on estimating the life of an item under "normal" operating conditions, based upon data obtained under much more severe conditions. In this case, the failure mechanism is usually well documented and understood; thus, problem identification and correction is of secondary importance.

Accelerated Stress Testing is used to identify problems and weaknesses inherent in the design, the parts used, or the manufacturing process so that they can be subsequently fixed. This is done by changes in: the design itself, the parts used, or the manufacturing processes employed. A thorough understanding, or at least a workable knowledge, of the basic failure mechanisms is the focus of attention here, estimation of item life may, or may not, be a concern.

Accelerated testing attempts to get more reliability information from a given test time using a test environment that is more severe than that experienced during normal equipment use, however:

Accelerated testing must always be approached with due caution. There are basic limitations to the technique. Every accelerated test application is unique. Subtle differences in the application can totally invalidate the data recorded during the test or the conclusions reached by the test.

This unfortunate outcome can occur, for example, if the operating range of validity for a given model is exceeded; or if the underlying test/modeling assumptions, while true for most applications, are not valid for a given specific application. Therefore, it is frequently necessary to first perform a preliminary accelerated test to validate the theory for a given application and then determine the applicable relationship (if not already available in the literature) between the applied stress and the resulting acceleration of the associated degradation. This preliminary accelerated test could also be viewed as a "sanity check."

Given these caveats, accelerating factors which may be used, either singly or in combination, include:

- More frequent power cycling
- Higher temperatures
- More severe temperature cycling
- Higher vibration levels
- Higher humidity

A second very important confounding factor in accelerated testing is the equipment level at which the test is performed. Some accelerating techniques are appropriate only for part level testing, while others can be used only for higher levels of assembly, and a very few techniques may be applicable for both part level and assembly level. The underlying assumptions and modeling approaches which may be perfectly legitimate at the part level may be totally invalid for tests performed on higher level equipment and vise-versa.

In addition to the primary purposes of accelerated testing, it also may be useful for:

- Identifying reliability problems in a chosen design
- Comparing the reliability of competing designs
- Acceptance testing
- Environmental Stress Screening
- Verifying the elimination of a given problem, etc.

8.7.1 Accelerated Life Testing

Accelerated life testing requires the use of a model relating the reliability (or life) measured under high stress conditions to that which is expected under normal operation. These tests require: (a) an understanding of the anticipated failure mechanism(s) and (b) a knowledge of the magnitude of the acceleration of this failure mechanism, as a function of the accelerating stress. In most cases appropriate acceleration factors can be obtained from a study of the literature, but in some cases new models may have to be developed. This will probably involve a significant investment of time and money.

It is very important, however, that the range of validity of a given acceleration model not be exceeded and that the accelerating stress change only the rate of failure and not the type of failure experienced. If an accelerated test introduces a new failure mechanism that will never be experienced in normal use, it may lead to false conclusions and possibly to unnecessary design changes. For this reason it is very beneficial to continue the accelerated life test until at least a minimum number of failures have occurred. Post mortem analysis will verify that the anticipated failure mechanism is indeed occurring, and that no new, different failure mechanisms have been introduced.

8.7.2 Accelerated Stress Testing

The main objective of a stress test is to convert latent defects or design weaknesses into actual failures, that is, to identify design, part and manufacturing process problems which could cause subsequent failures in the field. Time compression can frequently be realized by accelerating the environmental stress applied during the test, just as time compression is obtained during accelerated life testing. This same approach may be used both during development tests and during Environmental Stress Screening (ESS).

8.7.3 Equipment Level Accelerated Tests

Accelerated testing of equipment is usually quite limited. Creating a valid model relating the rate of equipment failures at a high stress - to that at normal operating conditions - is extremely difficult. Likewise it is very difficult to formulate stress conditions that do not change the failure mechanisms occurring within the equipment.

One example of an accelerated test that can be used effectively on equipment is that of increasing the duty cycle. Take for example an equipment normally operated at some given duty cycle, e.g., running only during one shift, or avionics equipment operating only a few hours before and during a flight. In such cases a higher duty cycle could easily be used during the test. The system undergoing test could be operated continuously for three shifts a day or the avionics equipment might be cycled continuously, with only enough time between simulated flights to permit the temperature within the equipment to stabilize during non-operating conditions. Although the failure rate per operating hour does not change, the number of failures accrued per day is increased.

This type of accelerated testing is commonly done in reliability qualification test, and although it is not usually recognized as such, this is actually a form of accelerated testing.

Another example of equipment level accelerated testing is ESS. In this case equipment is often subjected to higher stresses, particularly thermal cycling and vibration, as part of the ESS program. Here the purpose of the stresses are to detect defects induced into the equipment during the manufacturing process, e.g., weak solder joints, etc. Assuming that each defect is removed when it is discovered, with ESS there is no need of a model to correlate the rate of failure under stress to the rate of failure under normal operation.

Given these specific exceptions, accelerated testing is seldom applied at the equipment level. However, accelerated testing is an extremely important concept for component testing.

8.7.4 Component Level Accelerated Test

Components (parts) tend to have many fewer failure modes than equipment. Thus it is far easier to identify a stress which can be effectively accelerate the rate of failure without seriously changing the failure mechanism.

There is usually one or more dominant failure mechanisms accelerated by a given stress, e.g., dielectric breakdown of capacitors as a function of voltage, or corrosion as a function of humidity. In this case it is usually relatively easy to find an acceleration model relating failure rate as a function of operating stress. For this reason accelerated life testing is used extensively for components and the technique is highly recommended for most types of parts and for most part applications.

8.7.5 Accelerated Test Models

Accelerated test models relate the failure rate or the life of a component to a given stress such that measurements taken during accelerated testing can then be related back to the expected performance under normal operating conditions. The implicit working assumption here is that the stress will not change the shape of the failure distribution.

Three of the most commonly used acceleration models are:

1. Inverse Power Law
2. Arrhenius Acceleration Model
3. Miner's Rule

These are not the only models that exist, there are other models as well. The most important factor of concern is the correct choice of the model. The model chosen must be one that accurately models the reliability or life under the accelerated conditions to reliability or life under normal operating conditions. Great care is essential in choosing the most appropriate model and in selecting the appropriate range of validity for the chosen model in a specific application. Documenting the rationale for these choices is important.

8.7.5.1 The Inverse Power Law Acceleration Model

The inverse power law states that component life is inversely related to a power of the dominant stress.

$$\frac{\text{Life at normal stress}}{\text{Life at accelerated stress}} = \left(\frac{\text{Accelerated stress}}{\text{Normal stress}}\right)^N \qquad (8.42)$$

where N is the acceleration factor.

Assuming that an application is within the valid operating range of the model and that the shape of the failure distribution does not change under accelerated conditions, the inverse power law model can be used to solve such problems as the following.

Example: Suppose the mean life of a population of automobile tires was 20,000 miles when driven at 50 miles per hour. Through testing it has been determined that the mean life of these tires is 10,000 miles at 70 miles per hour. Thus:

$$\frac{20,000}{10,000} = \left(\frac{70}{50}\right)^N \qquad \text{Hence: } N = 2.06$$

From this knowledge, we want to use life data collected at 70 mph to show that there is a 90% probability that a tire will last 10,000 miles at 50 mph.
To solve this problem, use the life test data at 70 mph to demonstrate, with a 90% probability, that a tire will last 10,000 miles at 50 mph.

$$\text{Given:} \quad \frac{\text{Life at 50 mph}}{\text{Life at 70 mph}} = \left(\frac{70}{50}\right)^{2.06}$$

Desired result: 90% probability of no failure before 10,000 miles, i.e., no more than 10% of a population fails before 10,000 miles.

The shape of the failure distribution is assumed to be identical at 50 and 70 mph, thus the left side of the inverse power law equation shown above can be used to represent life at 10% failures, or:

$$\frac{\text{Life at 10\%, failures at 50 mph (10,000 miles desired)}}{\text{Life at 10\% failures at 70 mph}} = \left(\frac{70}{50}\right)^{2.06}$$

Thus: Life at 70 mph = 10,000/2 = 5,000

Therefore, if 10% or less of the tires tested at 70 mph fail by 5,000 test miles, we can conclude that 10% or less of tires driven at 50 mph will fail in 10,000 miles. Thus we have a 90% probability that a tire will last 10,000 miles at 50 mph.

8.7.5.2 The Arrhenius Acceleration Model

The Arrhenius acceleration model is widely used to predict life as a function of temperature. It applies specifically to those failure mechanisms that are temperature related and which are within the range of validity for the model.

It states that : $\text{Life} = A(e)^{\frac{E}{kT}}$ (8.43)

where:

Life = a measure of life e.g., median life of a population of parts
A = a constant determined by experiment for the parts involved
e = the base of the natural logarithms
E = activation energy (electron volts - a measure of energy) this is a unique value for each failure mechanism (Examples of the activation energies for some silicon semiconductor failure mechanisms are shown in Table 8.7-1.)
k = Boltzman's constant = 8.62 x 10^{-5} eV/K
T = Temperature (Degrees Kelvin)

SECTION 8: RELIABILITY DATA COLLECTION AND ANALYSIS, DEMONSTRATION, AND GROWTH

TABLE 8.7-1: ACTIVATION ENERGIES ASSOCIATED WITH VARIOUS SILICON SEMICONDUCTOR FAILURE MECHANISMS

DEVICE ASSOCIATION	FAILURE MECHANISM	RELEVANT FACTORS*	ACCELERATING FACTORS*	ACCELERATION (E_A APPARENT ACTIVATION ENERGY)
Silicon Oxide And	Surface Charge Accumulation	Mobile Ions V, T	T	Bipolar: E_A = 1.0-1.05eV MOS: E_A = 1.2-1.35eV
	Dielectric Breakdown	E, T	E	E_A = 0.3-2.0 eV
Silicon-Silicon Oxide Interface	Charge Injection	E, T	E, T	E_A =1.3eV(Slow Trapping) E_A =1.2eV "P" Channel E_A =1.05eV "N" Channel
Metallization	Electromigration	T, J, A Gradients of T and J Grain Size	T, J	E_A =0.5-1.2eV J to J^4 E_A =0.3eV Small Grain Size 0.5eV Typical Al 0.9eV Contact Windows
	Corrosion Chemical Galvanic Electrolytic	Contamination Humidity (H) V, T	H, V, T	Strong H Effect E_A =0.3-0.6eV (for V may have thresholds
	Contact Degradation	T, Metals Impurities	Varied	E_A = 0.9eV
Bonds and Other Mechanical Interfaces	Intermetallic Growth	T, Impurities Bond Strength	T	Al • Au: E_A = 1.0-1.05eV
	Fatigue	Temperature Cycling, Bond Strength	T Extremes in Cycling	E_A = 0.3-1.0eV
Hermeticity	Seal Leaks	Pressure Differential Atmosphere	Pressure Temperature Cycling	

* V - Voltage E - Electric Field A - Area
T - Temperature J - Current Density H - Humidity

“A” and “E” are typically calculated from test data using graphical methods. Special Arrhenius graph paper with a logarithmic life vertical scale and an inverse absolute temperature horizontal scale (in degrees Centigrade) is used. A straight line plot on this paper supports the assumption that an Arrhenius relationship holds (see Figure 8.7-1).

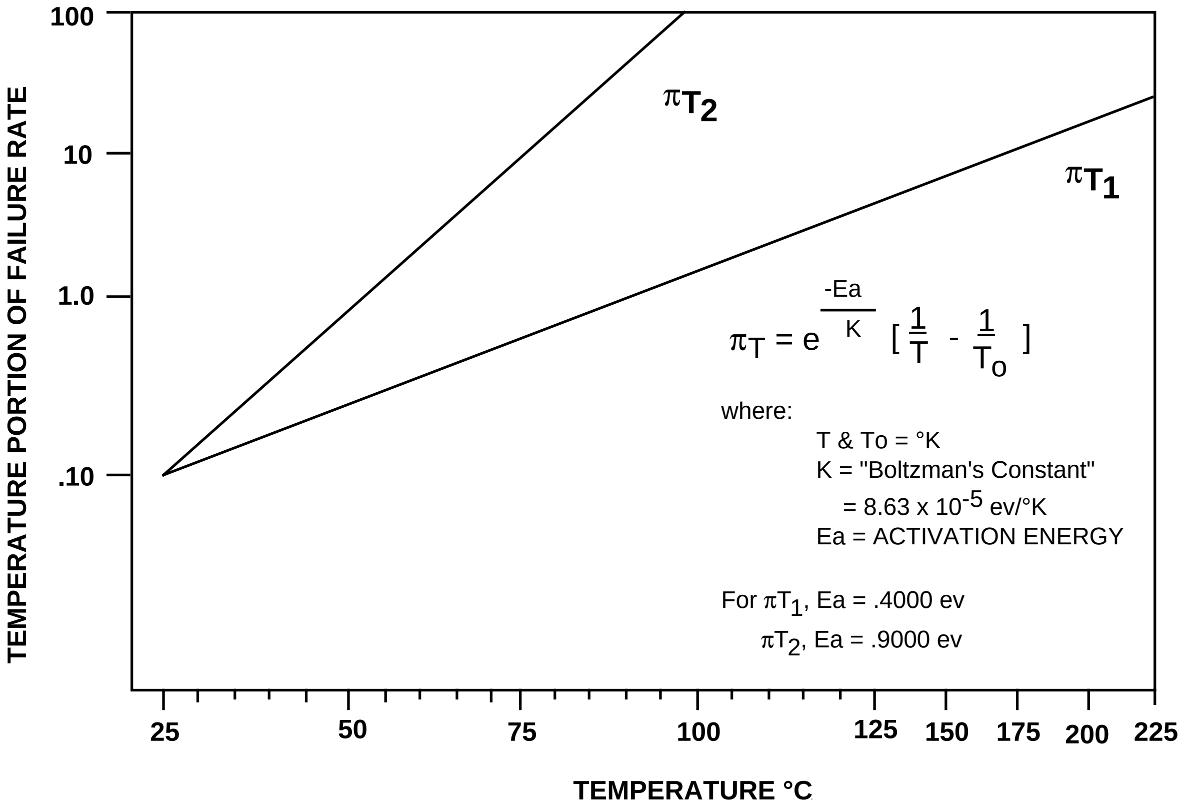


FIGURE 8.7-1: ARRHENIUS ACCELERATION MODEL

8.7.5.3 Miner's Rule - Fatigue Damage

Quantification of metal fatigue under the varying load conditions which an item sees in service is frequently a major design concern. Fatigue researchers have proposed various cumulative damage models. The simplest of these models is Miner's rule. Miner's rule states that cumulative damage (CD) is:

$$CD = \sum_{i=1}^{k} \frac{C_{s_i}}{N_i} \le 1 \qquad (8.44)$$

where:

C_{S_i} = number of cycles applied at a given mean stress S_i

N_i = the number of cycles to failure under stress S_i, (as determined from an S-N diagram for that specific material)

k = the number of loads applied

Thus it is assumed that at the end of life (point of failure) CD = 1.

Miner assumes that every part has a finite useful fatigue life and every cycle uses up a small portion of that life. Failure is likely to occur when the summation of incremental damage from each load equals unity. (Miner's rule does not extend to infinity, however. It is valid only up to the yield strength of the material, beyond that point it is no longer valid.)

We can then construct an accelerated fatigue test by combine Miner's Rule with the previously discussed inverse power law. The inverse power law (equation 8.42) stated that the damage-accumulation rate is proportional to a power of the current stress. Thus:

$$\left[\frac{\text{Life at normal stress}}{\text{Life at accelerated stress}}\right] = \left[\frac{\text{accelerated stress}}{\text{normal stress}}\right]^N$$

where:

N = the acceleration factor derived from the slope of the S-N curve

Accelerated cumulative fatigue damage could therefore be calculated by combining Miner's rule (equation 8.44) and the power law (equation 8.42). Thus from equation 8.45:

$$CD = \sum \frac{C_{si}}{N_i}$$

and from equation 8.42, for accelerated stress causing failure in one cycle:

$$\frac{N_i}{1} = \left(\frac{S_1}{S_i}\right)^{\alpha}$$

where:

- α = N from the inverse power law = material dependent parameter (slope of the S-N curve)
- N_i = the number of cycles to failure under stress S_i
- S_i = stress level associated with N_i cycles
- S_1 = stress level required for failure in 1 stress reversal

Thus:

$$CD = \sum_{i=1}^{k} \frac{C_{S_i}}{\left(\frac{S_1}{S_i}\right)^{\alpha}} = \sum_{i=1}^{k} C_{S_i}\left(\frac{S_i}{S_1}\right)^{\alpha} = n_i\left(\frac{s_i}{S_1}\right)^{\alpha} \tag{8.45}$$

where:

n_i = the number of applied stress reversals at a single stress level i
s_i = stress level associated with n_i

An S-N diagram is commonly used to present the data from equation 8.45. The S-N diagram plots the number of stress cycles required to break an item at a given stress level. The power of accelerated fatigue testing can then be demonstrated by simplifying equation 8.45 and assuming a material parameter. Since S is a constant:

$$CD \propto n_i (s_i)^{\alpha} \tag{8.46}$$

The cumulative fatigue damage then becomes proportional to the number of stress cycles and their associated stress level. To illustrate, calculate the increase in cumulative fatigue damage during accelerated testing when the stress level (s_i) is doubled, assuming (for the sake of illustration only that) the material parameter α = 10, then:

$$\Delta CD \propto n_i (2)^{10} = n_i (1024)$$

Thus the fatigue damage accumulates 1024 times (2^{10}) faster than what it would at the baseline stress. Hence, a 20-second test with the applied stress doubled becomes the equivalent of a 300-minute vibration test at normal stress level! Properly applied, this technique can be a powerful tool. In this example *(assuming that the yield strength of the material was not exceeded during the test)*, identifying design problems quickly could be readily achieved using an accelerated stress test.

8.7.6 Advanced Concepts In Accelerated Testing

The intent here is not to get deeply involved in the mechanics of accelerated testing, especially not the advanced concepts, but rather to make the user aware of some of the more common practices in the discipline, such as non-constant stress profiles, combined stress profiles and more recent developments in the field.

Historically, most accelerated testing is done using a single stress and a constant stress profile. This includes cycled stress (e.g. temperature cycling between specified limits) where the cycle (upper and lower temperature limits and rate of change of temperature), rather than the temperature is fixed. In accelerated testing, however, the stress profile need not be constant and a combination of stresses may also be used. Some common non-constant stress profiles and combined stress profiles variations include:

- Step Stress Profile Test
- Progressive Stress Profile Test
- Highly Accelerated Life Test (HALT) (Equipment-level)

- Highly Accelerated Stress Screens (HASS) (Equipment-level)
- Highly Accelerated Temperature and Humidity Stress Test (HAST) (Part-level)

Highly accelerated testing is the systematic application of environmental stimuli at levels well beyond those anticipated during product use. Thus, the results need to be carefully interpreted. It is used to identify relevant faults and to assure that the resulting products have a sufficient margin of strength above that required to survive the normal operating environments. Highly accelerated testing attempts to greatly reduce the time needed to precipitate these defects. The approach may be used either for development testing or for screening.

HALT is a development tool and HASS is a screening tool. They are frequently employed in conjunction with one another. They are new, and are in conflict with the classical approach to accelerated testing; thus, they are controversial. Their specific goal, however, is to improve the product design to a point where manufacturing variations and environment effects have minimal impact on performance and reliability. There is usually no quantitative life or reliability prediction associated with highly accelerated testing.

8.7.6.1 Step Stress Profile Testing

Using a step stress profile, test specimens are subjected to a given level of stress for a preset period of time, then they are subjected to a higher level of stress for a subsequent period of time. The process continues at ever increasing levels of stress, until either; all specimens fail, or the time period at the maximum level stress ends, as shown in Figure 8.7-2. This approach provides more rapid failures for analysis, but with this technique it is very difficult to properly model the acceleration and hence to quantitatively predict the item life under normal usage.

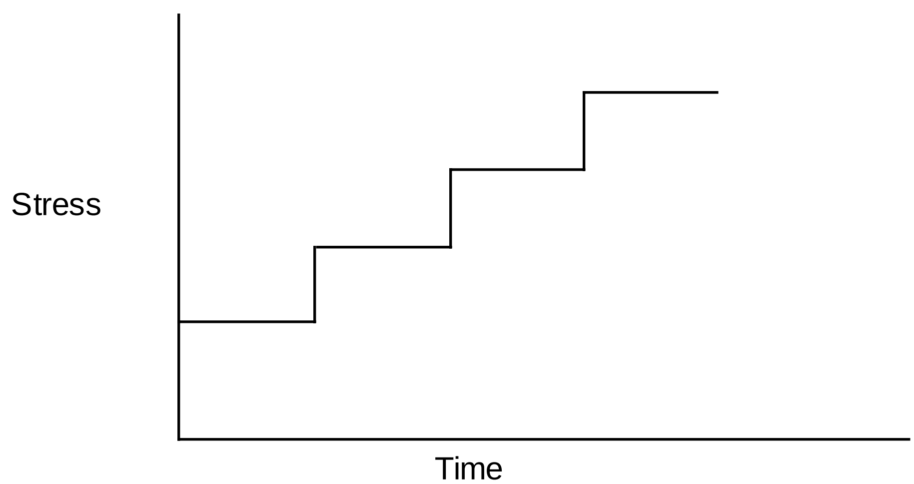


FIGURE 8.7-2: STEP STRESS PROFILE

How much to increase the stress in any single step is a function of many variables and is beyond the scope of this discussion. However, the general rule to follow in the design of such a test is to eventually exceed the expected environments by a comfortable margin so that all members of the population can be expected to survive both the field environment and the screen environments, assuming of course that they are defect free.

8.7.6.2 Progressive Stress Profile Testing

A progressive stress profile or “ramp test” is another frequently used approach (see Figure 8.7-3). With this approach the level of stress is continuously increased with time. The advantages and disadvantages are the same as those for step stress testing, but with the additional difficulty of accurately controlling the rate of increase, of the stress.

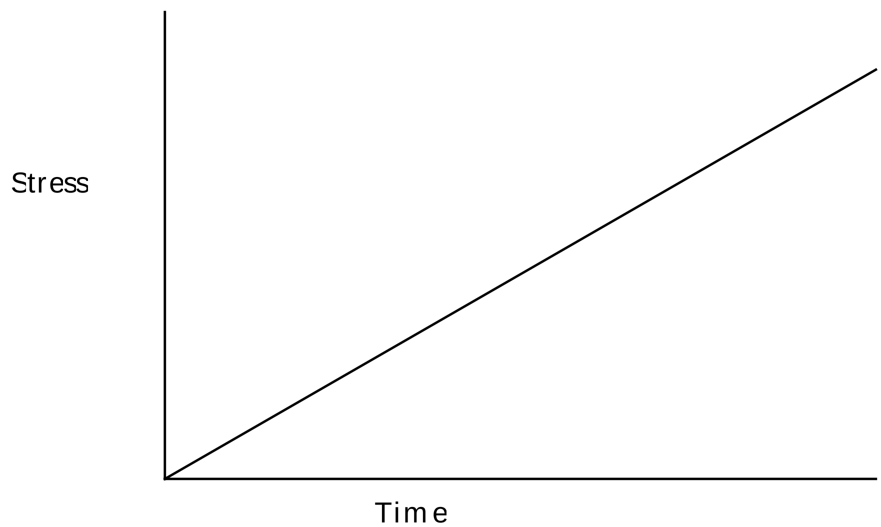


FIGURE 8.7-3: PROGRESSIVE STRESS PROFILE

8.7.6.3 HALT Testing

The term HALT was coined in 1988 by Gregg K. Hobbs (Ref. [8]). HALT (also, sometimes referred to as STRIFE (Stress plus Life) testing) is a development test, an enhanced form of step stress testing. It is typically used to identify design weaknesses and manufacturing process problems and to increase the margin of strength of the design rather than to predict quantitative life or reliability of the product.

HALT testing begins with step stress testing in generic stresses such as temperature, rate of change of temperature, vibration, voltage, power cycling and humidity. In addition, product unique stresses such as clock frequency, DC voltage variation and even component value variation may be the accelerated stimuli.

The type of the vibration stimuli used for HALT (and HASS) testing is unique. It is not based upon the universally accepted accelerated (power) spectral density concept. Thus it does not utilize classical, single-axis, sinusoidal vibration or a random vibration spectrum, generated by acceleration-controlled electro-dynamic shakers. Instead an unconventional multi-axial pneumatic (six degree of freedom) impact exciter is typically used. This type of equipment generates a highly unique broadband accelerated shock response spectrum (SRS). This is effectively a repeated shock environment rather than a vibration environment and is, in its self, much more severe than a classical vibration spectrum. Because of the choice of this shock stimuli spectrum, the resulting data cannot be easily correlated with either: (a) the normal environment or with (b) classical vibration testing using classical vibration modeling approaches. Thus quantitative prediction of life or reliability is not usually possible with HALT and HASS.

Using HALT the step stress process continues until stress levels well above those expected in the normal operational environments are exceeded. Throughout the process continuous evaluation is performed to determine how to make the unit able to withstand the increasing stress. Generally temporary fixes are implemented just so that the test can continue. When a group of fixes is identified, a permanent block change is then implemented.

After one stimuli has been elevated to a level felt to be sufficient, another stimuli is selected for step stress testing. This progression continues until all stimuli have been applied separately. Then combined stresses are run to exploit the synergism between the stresses, that is, the combined effect may generate larger stresses than either stress alone would create. After design fixes for the identified problems have been implemented, a second series of step stresses are run to verify the fixes, assure that the fixes themselves have not introduced new problems and to look for additional problems which may have been missed due to the limited sample size. This aspect of HALT must be taken into account in selecting the appropriate stress levels since a slight increase in stress can greatly reduce the number of cycles to failure.

For all of these stimuli, the upper and lower operating limits and the destruct limits should be found or at least understood. Understood means that although the limits are not actually found, they are verified to be well beyond the limits which may be used in any future HASS test and even farther beyond the normal field environments. For example, a product may be able to withstand an hour of random vibration at 20 G_{rms} without failure. Although the destruct limit may not have been found, it is certainly high enough for most commercial equipment intended for non-military environments where the screen environment may be 10 G_{rms} random vibration for 5 minutes and the worst field environment is a truck ride while in an isolation container. This example of the capability far exceeding the field environment is quite common when HALT is properly applied.

There are several reasons for ascertaining both the operating limits and the destruct limits. Knowledge of the operating limits is necessary in order to assess if suitable design margins exist and how large the margins are likely to be as a function of population. It is also necessary to formulate failure detection tests. These can be run during any future HASS test since the detection tests run during stimulation are necessary for high detectability of precipitated defects. Knowledge of the destruct limits is required in order to determine the design margins in non-operating environments and to assure that any future HASS environments are well below destruct levels.

8.7.6.4 HASS Testing

HASS is a form of accelerated environmental stress screening. It presents the most intense environment of any seen by the product, but it is typically of a very limited duration. HASS is designed to go to "the fundamental limits of the technology." This is defined as the stress level at which a small increase in stress causes a large increase in the number of failures. An example of such a fundamental limit might be the softening point of plastics.

HASS requires that the product have a sufficient margin of strength above that required to survive the normal use environments. Temperature, vibration levels, voltage and other stimuli exceeding the normal levels are used in HASS to force rapid defect precipitation in order to make the screens more effective and economical. The use of HASS requires a thorough knowledge of the product's ability to function at the extended ranges of simulation and also detailed knowledge about the failure mechanisms which limit these stimuli levels. Design and process changes are usually made to extend the functional and destruct levels of the equipment in order to assure large design and process margins as well as to allow HASS, with its attendant cost savings, to be performed. These saving can potentially produce orders of magnitude reduction in screening cost as well as significant quality improvements. One risk is that the item may be overdesigned.

Development of screening levels to be used in HASS begins during HALT testing. Operational levels and destruct levels are used as guidelines to select environmental limits during HASS. Two levels of environmental stimuli are chosen for each accelerated screening environment: the precipitation level and the detection level. Precipitation is the manifestation of a latent, or dormant, product flaw (i.e., it changes from a latent state to a patent or evident, detectable state). Detection is the observation that an abnormality exists. The observation may be made visually, electronically, audibly, etc.

The precipitation levels are chosen to be well below the destruct level, but beyond the operational limits. During the precipitation screen, the test item may not operate within the required limits but functional operation must be maintained and it must be monitored. These levels serve as the acceleration factor to minimize the time necessary to precipitate faults. The detection stress level is chosen outside of or just below the operational level determined during HALT testing. During the detection portion of the screen, operational parameters are monitored for compliance with the requirements. Once the screening parameters have been set, a proof-of-screen test must be performed to ensure that the accelerated screening levels are not damaging the product. The proof-of-screen is performed by simply running multiply accelerated screening profiles until either the product wears out or assurance is gained that the screen is not devouring appreciable useful life. Typically, repeating the screening environment 10 times is acceptable proof, provided there is no evidence of product wear out.

Its is critical that the product be powered up and monitored during HASS. A large portion, typically greater than 50%, of the faults identified during screening are soft or intermittent faults. Not having complete diagnostics and detection these faults can be disastrous. An intermittent fault in the factory is very likely to be an early failure in the field.

HASS is a time compressed environmental stress screen applied at the earliest functional level of assembly. Complete functional monitoring of the test item is extremely important. Non-detected faults correlate with early life failures and dissatisfied customers. A poorly designed screen can be worse than no screen at all! Thus it is important to perform proof-of-screen evaluations prior to screening in production, to ensure that the screen does not appreciably reduce the useful life of the product. One must be receptive to changing the screen if field data indicates that a specific failure mechanism is escaping the screen. Thus an effective screening process is a dynamic process.

8.7.6.5 HAST (Highly Accelerated Temperature and Humidity Stress Test)

With the vast recent improvements in electronics technology and the speed with which these technology improvements are occurring, accelerated tests which were designed just a few years ago may no longer be adequate and efficient for today’s technology. This is especially true for those accelerated tests intended specifically for microelectronics. For example, due to the improvements in plastic IC packages, the previous virtually universally accepted 85°C/85%RH Temperature/Humidity test now typically takes thousand of hours to detect any failures in new integrated circuits. In most cases the test samples finish the entire test without any failures. A test without any failures tells us very little. Yet we know that products still fail occasionally in the field; thus, we need to further improved our accelerated tests.

Without test sample failures we lack the knowledge necessary to make product improvements. Therefore the accelerated test conditions must be redesigned accordingly (e.g., utilize higher temperatures) to shorten the length of time required for the test, to make it more efficient and hence more cost effective. This is the background for today’s focus (at the component level) upon Highly Accelerated Temperature and Humidity Stress Testing.

8.7.7 Accelerated Testing Data Analysis and Corrective Action Caveats

An accelerated test model is derived by testing the item of interest at a normal stress level and also at one or more accelerated stress levels. Extreme care must be taken when using accelerated environments to recognize and properly identify those failures which will occur in normal field use and conversely those that are not typical of normal use. Since an accelerated environment typically means applying a stress level well above the anticipated field stress, accelerated stress can induce false failure mechanisms that are not possible in actual field use. For example, raising the temperature of the test item to a point where the material properties change or where a dormant activation threshold is exceeded could identify failures which cannot occur during normal field use. In this situation, fixing the failure may only add to the product cost without an associated increase in reliability. Understanding the true failure mechanism is paramount to elimination of the root cause of the failure.

The key to a successful accelerated testing program is to properly identify the failure mechanism and then eliminate the fault. Accelerating an environment such as temperature or vibration will uncover a multitude of faults. Each of these faults must be analyzed until the failure mechanism is fully understood. Chasing the wrong failure mechanism and implementing corrective action which does not eliminate the true cause of failure adds to the product's cost but does not improve product reliability.

A systematic method of tracking faults identified during accelerated testing ensures that problems are not forgotten or conveniently ignored. Each fault must then be tracked from the moment it is identified until either: a) corrective action is verified and documented or, b) a decision is made not to implement correction action. The failure tracking system must be designed to track the short term progress of failures over time.

When quantitative estimate of life or reliability is needed, the failure distribution must be determined for each stress condition. Next a model is derived to correlate the failure distributions. This is done to quantitatively predict performance under normal use, based upon the observed accelerated test data.

Constant stress prediction models frequently employ a least-square fit to the data using graphical methods such as those previously described in Section 8.3.1 or statistical methods such as those described in Section 8.3.2. However, when non-constant stresses are used, correctly plotting the data is much more complicated. Also, in many cases it may be necessary to use more elaborate techniques, such as those described in Section 8.3.2.4, to account for censored data.

Censored data is defined as data for test specimens which do not have a recorded time to failure. Some of the reasons for censoring data include:

(1) A unit may still be running without failure when the test ends

(2) The failure may be for some reason other than the applied test stress (e.g. mishandling)

(3) The item may have been removed from the test before failure for various reasons.

Complex censored data cases usually require powerful analysis tools, e.g., maximum likelihood methods, and cumulative damage models. Such tools can be cumbersome to use, but fortunately there are a number of statistically based computer programs to assist in these analyses.

Identifying which corrective action will solve the problem frequently involves multiple engineering and production disciplines. Multiple discipline involvement is necessary to prevent finding a "fix" which cannot be economically built in production. Corrective action frequently involves a pilot build process which confirms that the "fix" does not introduce unanticipated new problems.

Corrective action verification should be performed in quick steps whenever possible. The accelerated testing environment is reapplied to verify that the proposed corrective action does eliminate the problem. Documenting the action taken is necessary to prevent reoccurrence and to ensure that production is modified to conform to the design change. Documentation should be shared throughout the organization to ensure that reoccurrence is indeed prevented. Conversely, a decision might be made not to implement corrective action based upon a monetary risk assessment.

Corrective action is expensive, if the problem affects only a small portion of the product population, the anticipated warranty repair cost will probably also be low. Thus the program management may elect to live with the identified risk. The decision, however, must always be based upon the root cause of the failure not applying to the intended use of the product, e.g., the failure mechanism cannot occur in normal field usage. This decision should always be made with due caution. Historically, some "non-relevant" or "beyond normal use" failures do recur in the field and become very relevant.

8.8 References for Section 8

1. Engineering Design Handbook: Reliability Measurement. January 1976, AMCP-706-198, AD#A027371.

2. Horn, R., and G. Shoup, "Determination and Use of Failure Patterns," Proceedings of the Eighth National Symposium on Reliability and Quality Control, January 1962.

3. VanAlvin, W. H., ed., Reliability Engineering. Englewood Cliffs, NJ: Prentice-Hall Inc., 1966.

4. Lloyd, R.K. and M. Lipow, Reliability: Management, Methods, and Mathematics, TRW, Redondo Beach, CA, second edition, 1977.

5. Mann, N., R. Schafer and N. Singpurwalla, Methods of Statistical Analysis of Reliability and Life Data. New York, NY: John Wiley and Sons, 1974.

6. Quality Assurance Reliability Handbook. AMCP 702-3, U.S. Army Materiel Command, Washington DC 20315, October, 1968, AD#702936.

7. Turkowsky, W., Nonelectronic Reliability Notebook. RADC-TR-69-458, March 1970.

8. Hobbs, G. K., Highly Accelerated Life Tests - HALT, unpublished, contained in seminar notes, “Screening Technology” © April 1990.

9. Harris, C.M., Crede, C.E, “Shock and Vibration Handbook,” McGraw-Hill, 1961.

10. Nelson, Dr. Wayne, “Accelerated Testing,” John Wiley & Sons, 1990.

11. “Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Mean Life Criterion),” Quality Control and Reliability Technical Report, TR3, Office of the Assistant Secretary of Defense (Installations and Logistics), September 30, 1961.

12. “Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Hazard Rate Criterion),” Quality Control and Reliability Technical Report TR4, Office of the Assistant Secretary of Defense (Installations and Logistics), February 28, 1962.

13. “Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Reliable Life Criterion),” Quality Control and Reliability Technical Report, TR6, Office of the Assistant Secretary of Defense (Installations and Logistics), February 15, 1963.

14. Crow, L. H., “On Tracking Reliability Growth,” Proceedings 1975 Annual Reliability & Maintainability Symposium, pp 438-443.

15. Discrete Address Beacon System (DABS) Software System Reliability Modeling and Prediction, Report No. FAA-CT-81-60, prepared for U.S. Department of Transportation, FAA Technical Center, Atlantic City, New Jersey 08405, June 1981.

16. Reliability Growth Study, RADC-TR-75-253, October 1975, ADA023926.

17. Green, J. E., “Reliability Growth Modeling for Avionics,” Proceedings AGARD Lecture Series No 81, Avionics Design for Reliability, April 1976.

18. MIL-HDBK-781A, “Reliability Test Methods, Plans and Environments for Engineering, Development, Qualification and Production,” April 1996.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

9.0 SOFTWARE RELIABILITY

9.1 Introduction

Hardware reliability engineering was first introduced as a discipline during World War II to evaluate the probability of success of ballistic rockets. The 1950's brought more advanced methods to estimate life expectancies of mechanical, electrical and electronic components used in the defense and aerospace industry. By the 1960's, reliability engineering had established itself as an integral part of end user product development in commercial products as well as military applications. (Ref. [1]).

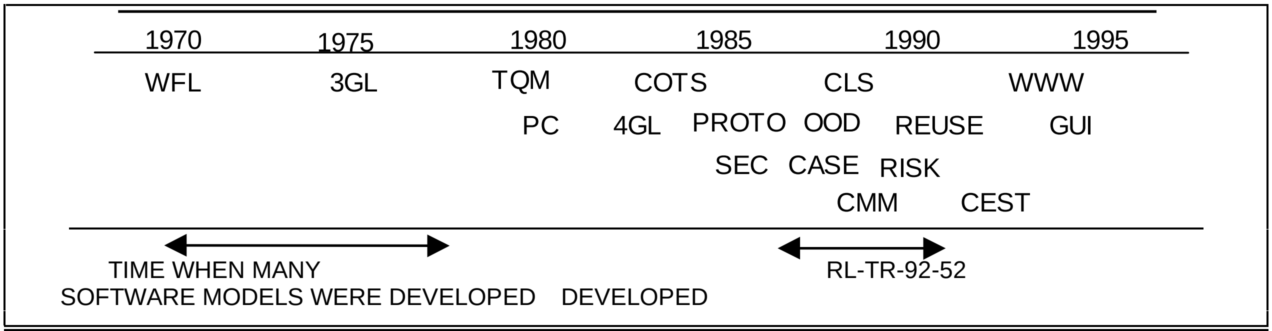
The *software reliability* discipline is much younger, beginning in the mid 1970's when the software development environment was reasonably stable. Most of software reliability models were developed during this time of software stability. However, a surge of new technology, new paradigms, new structured analysis concepts, and new ways of developing software emerged in the late 1980's and continues to this date. Figure 9.1-1 provides a chronological reference for some of the elements which comprise the current software development environment and add to its complexity.

As more and more systems that are a part of everyday life become more and more dependent upon software, perceptions about software reliability have changed. Increasing control by software of items such as dishwashers, ovens and automobiles, along with liability issues associated with these products, has led to an increased awareness of the criticality of reducing "hidden" software errors. Additionally, the influx of computers into financial and security-related operations requires a guarantee of data integrity.

Software engineers uniformly do not have an analogous view of reliability. Webster defines reliable as "giving the same result on successive trials." This definition, when extrapolated to include "forever," more closely resembles the view of reliability imposed on software engineers. In general, the reliability metric for software is used to describe the probability of the software operating in a given environment within the designed range of input without failure. Therefore, *software reliability* is defined as the probability that software will not cause a system failure over a specified time under specified conditions. This probability is a function of the inputs to and use of the system, as well as the presence of latent software faults. The system inputs determine whether any latent faults will be encountered during system operation.

SECTION 9: SOFTWARE RELIABILITY

WHEN THESE SOFTWARE ENGINEERING CONCEPTS WERE INTRODUCED



LABEL	TERM	LABEL	TERM
3GL	3rd Generation Languages	PC	Personal Computers
4GL	4th Generation Languages	PROTO	Rapid Prototyping
CASE	Computer Assisted Software Engineering Tools	REUSE	Reuse
CEST	Software Cost Estimation	RISK	Software Risk Management
CLS	Client Server Technology	RL-TR-	Software Reliability, Measurements and Testing
CMM	Capability Maturity Model	SEC	Software Engineering Curricula
COTS	Commercial Off-the-Shelf Software	TQM	Total Quality Management
GUI	Graphical User Interfaces	WFL	Waterfall Development Model
OOD	Object Oriented Design	WWW	World Wide Web (Internet)

FIGURE 9.1-1: SOFTWARE ENVIRONMENT TIMELINE

Additional differences between hardware and software reliability include:

(1) The age of the software has nothing to do with its failure rate. If the software has worked in the past, it will work in the future, everything else remaining the same (i.e., no hardware, software or interface changes). Software does not rust or exhibit other hardware wearout mechanisms.

(2) The frequency of software use does not influence software reliability. The same software can be used over and over and, if it did not fail the first time, it will not fail any other time in identical usage (same range of inputs with no hardware, software or interface changes). In contrast, physical parts wear from usage, resulting in failure.

(3) Software does become obsolete as user interface standards evolve and hardware become antiquated.

(4) With the exception of documentation and storage/transfer media, software, unlike hardware, cannot be held or touched. Typical methods of judging a hardware item include observing size and material composition, quality of assembly (form, fit and finish), and compliance with specification. For example, one can observe how well two gears mesh or if a transistor has sufficient current capacity for a circuit application. These physical concepts do not apply to software.

(5) Software cannot be judged prior to use by the same methods as hardware, i.e., there is no equivalent to incoming inspection.

(6) Software must be matched with hardware before it can ever be tested. If a failure occurs, the problem could be hardware, software, or some unintended interaction at the hardware/software interface.

(7) In general, hardware will either work or not in a given application. Software, aside from total failure, has varying degrees of success according to its complexity and functionality.

(8) Although not executable, documentation usually is considered an integral part of the software. Documentation which does not fully or accurately describe the operation can be considered to be just as much a failure as a software crash. When a user expects on-line help and does not get it (either because it is not activated or because what was provided was incorrect or incomplete), the software does not meet the user's expectation and, therefore, is not perfectly reliable. In contrast, documentation is usually not assessed when evaluating hardware reliability.

Admittedly there are differences between hardware and software. Rather than dwelling on the differences, we should look at the similarities. Some of these are:

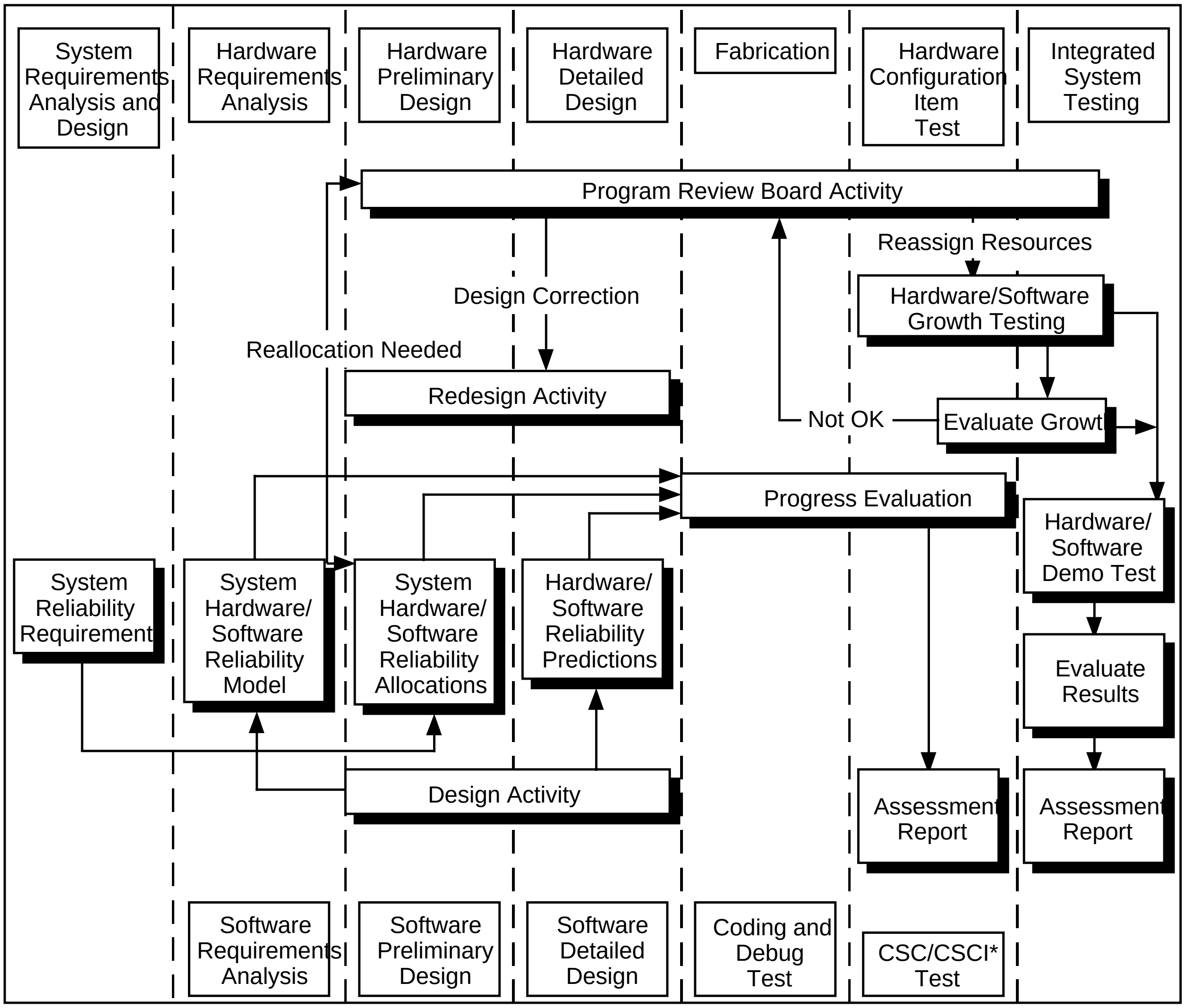
(1) Hardware reliability is a function of equipment complexity; intuitively one would expect the same to be true of software.

(2) Solid state electron devices (e.g., transistors, microcircuits) if fabricated properly, do not have any wearout mechanisms that one can see over a long time period. The defects which cause failure (other than obvious misapplication of the device) are built-in during the initial fabrication of the device; the same is true of software.

(3) Hardware reliability can be improved by reliability growth testing, e.g., a test-analyze-and-fix program to discover, identify, and correct failure modes and mechanisms which would cause early equipment failure. This is similar to finding and eliminating "bugs" in a software program, thus increasing its reliability.

Thus, we should be concentrating on the duality that exists between the successful hardware approaches and the emerging software approaches. Once this is accepted, the whole problem is simplified because the hardware and software problems can be approaches together in a total system context.

The duality between hardware and software is graphically portrayed in Figure 9.1-2 which illustrates the key elements of hardware and software programs during the life cycle phases of system development. The basic difference occurs during full scale engineering development, when hardware is fabricated and tested while software is coded (programmed) and debugged.



* Computer Software Component/Computer Software Configuration Item

FIGURE 9.1-2: HARDWARE/SOFTWARE SYSTEM LIFE CYCLE RELATIONSHIP (REF. [2])

9.2 Software Issues

Quality Focus One essential concept for both hardware and software is that the customer's perception of quality is extremely important. Quality is delivering what the customer wants or expects. Customers must be considered during the specification and design stages of development. Since various customer groups have conflicting interests and view quality and reliability differently, it is important to analyze the customer base.

For example, the organization funding a project is one customer, the user another. If they are different organizations, their expectations may be in conflict. Quality for the funding organization may be interpreted as “delivering on time and within budget” with “conformance to requirements” viewed as having less priority. In contrast, the customer who depends on the system’s functionality to meet organizational needs is probably not as concerned with development schedule or cost. The pilot of a jet fighter expects the hardware and software to work perfectly regardless of whether the various sub-systems were delivered on time or within budget. Any failure, for any reason, may be catastrophic. On the other hand, those accountable for verifying that the jet will not fail are very much interested in ensuring that both the hardware and software have been thoroughly tested and that the reliability assessment process is consistent with what has been used in other systems that have proved to be as reliable as predicted. The expectation is that quality consists of evidence that everything possible has been done to ensure failure-free operation, providing very high reliability.

The Software Engineering Institute (SEI) Capability Maturity Model (CMM) provides a framework for organizing small evolutionary steps into five maturity levels. These levels provide successive foundations for continuous improvement. Details of each level are found in “Capability Maturity Model for Software (Version 1.1),” CMU/SEI-93-TR-024, Software Engineering Institute, and are summarized in the following paragraphs.

Level 1. At the initial level, Level 1, the organization typically lacks a stable environment for developing and maintaining software. In this case, the benefits of good software engineering practices are undermined by ineffective planning and reactive systems. Since the software process is not stable, the software process capability is unpredictable. Schedules, budgets, functionality, and product quality also are generally unpredictable.

Level 2. An organization at the repeatable level, Level 2, has developed policies for managing software projects and has procedures for implementing those policies. Experience gained on one software development project is used to plan and manage new, similar projects. One criteria for Level 2 is the institutionalization of effective management processes for software development. This institutionalization allows successful practices developed on earlier projects to be repeated, although specific processes may differ from project to project. An effective process has the following characteristics: practiced, documented, enforced, measured and improvable.

A Level 2 organization has basic software management controls in place. Managers of software projects track costs, schedule, and functionality. They monitor the project to identify problems in meeting commitments. Software requirements and associated work products are baselined and the integrity of the configuration is controlled. Defined project standards are available and faithfully followed. A strong customer-supplier relationship is established with any subcontractors.

Level 3. Level 2 is called the defined level. At this level, the standard process for developing and maintaining software throughout the organization is documented. Software engineering and management processes are integrated into a coherent whole. Effective software processes are exploited in the development of the organization's standard software process. Training is conducted across the organization to ensure managers and staff have the knowledge and skills needed to carry out their role in the process. One group is responsible for the organization's software process activities.

The characteristics of a well-defined software process include readiness criteria, inputs, work performance standards and procedures, verification mechanisms, outputs, and completion criteria. A well-defined software process gives management good insight into technical progress.

Level 4. At the managed level, Level 4, quantitative defect goals for software and the software process are established. Productivity and defect rates for important software process activities are measured across all projects as part of an organization-wide measurement program. All measurement data is entered into a common data base and used to analyze process performance. Project managers control assigned projects and processes by reducing variations in performance to fall within acceptable limits. Risks associated with moving up the learning curve of a new application domain are known, tracked, and managed.

Level 5. The highest level of maturity is aptly called the optimizing level. Here the organization has the means and will to continuously improve the process. Weaknesses are identified and processes are strengthened proactively, with the prevention of defects being the objective. Data on the effectiveness of the software process are collected and used to conduct cost-benefit analyses of new technologies and proposed process changes. Innovative ideas that capitalize on the best software engineering practices are identified and implemented throughout the organization.

At Level 5, the software process capability is characterized as continuously improving. This continuous improvement results from constantly striving to improve the range of process capability, thereby improving process performance of projects. Improvement comes in the form of incremental advancement of existing processes and innovative application of new technologies and methods.

Organizational Structure. The typical sequential organizational structure does not support significant cross communication between hardware and software specialists. An organization's internal communication gap can be assessed by considering the questions in Table 9.2-1. The answers help determine if the organizational structure creates two "separate worlds." If reliability is important and a communication gap exists, then the organization needs to break down the communication barriers and get all parts of the technical community to focus on a common purpose. Activities may involve awareness training, cross training, organizational restructuring, implementing/improving a metrics program, reengineering the overall system development processes as well as the sub-system (i.e., hardware and software) processes, or instituting a risk assessment/risk management program.

Reliability Terminology. While hardware-focused reliability engineers have adopted a common set of concepts and terms with explicit meaning, the software community has not yet reached consensus and, hence, no universally adopted terminology set is in place. Many concepts, fundamental to the discussion and development of software reliability and quality, have several meanings. Worse, they are often used interchangeably!

TABLE 9.2-1: ASSESSING THE ORGANIZATIONAL COMMUNICATIONS GAP

- Is the software group a separate entity?
- Does the organization consider software as an engineering discipline?
- What is the career path for hardware/software, or system engineers?
- What forums exist for interaction engineers, and project managers?
- Who heads up system development? Hardware engineers? Software engineers? Others?
- Is there an expressed need for quantifying system reliability?
- Who has defined the system reliability metric?
- Who is responsible for assessing system reliability?
- What metrics are in place for assessing system reliability?
- What program is in place for testing system reliability?

For instance, software engineers often use "*defect*", "*error*", "*bug*", "*fault*", and "*failure*" interchangeably. Capers Jones (Ref. [3]) defined these terms as follows:

(1) Error: A mistake made by a programmer or software team member that caused some problem to occur.

(2) Bug: An error or defect that finds its way into programs or systems.

(3) Defect: A bug or problem which could cause a program to either fail or to produce incorrect results.

(4) Fault: One of the many nearly synonymous words for a bug or software defect. It is often defined as the manifestation of an error.

Some software specialists define a "*failure*" as any inappropriate operation of the software program while others separate "*faults*" and "*failures*" on a time dimension relative to when a defect is detected: "*faults*" are detected before software delivery while "*failures*" are detected after delivery. To the hardware community this appears to be an artificial distinction; yet it is important to be aware of the differentiation since both terms are used in actual practice. Software people talk about "*fault rate*" and "*failure rate*", with the latter term having a different meaning than that used with regard to hardware.

RobertDunn (Ref. [4]) defines a software defect as "Either a fault or discrepancy between code and documentation that compromises testing or produces adverse effects in installation, modification, maintenance, or testing". In contrast, Putnam and Myers (Ref. [5]) define a defect as "A software fault that causes a deviation from the required output by more than a specified tolerance. Moreover, the software need produce correct outputs only for inputs within the limits that have been specified. It needs to produce correct outputs only within a specified exposure period." Since these definitions differ, a count of the number of defects will yield different results, and, hence, a different defect rate, depending on the counter's definition.

Dunn separates defects into three classes (he feels that it is fairly easy for experienced programmers to relate to each of these):

(1)	Requirements Defects:	Failure of software requirements to specify the environment in which the software will be used, or requirements documentation that does not reflect the design of the system in which the software will be employed.
(2)	Design Defects:	Failure of designs to satisfy requirements, or failure of design documentation to correctly describe the design.
(3)	Code Defects:	Failure of code to conform to software designs.

Typical requirements defects include indifference to the initial system state, incomplete system error analysis and allocation, missing functions, and unquantified throughput rates or necessary response times. The many kinds of design defects include misinterpretation of requirements specifications, inadequate memory and execution time reserves, incorrect analysis of computational error, and infinite loops. Possible code defects include unreachable statements, undefined variables, inconsistency with design, and mismatched procedure parameters.

Other software experts have different classifications. For example, Putnam and Myers define six classes of defects:

(1) Requirements Defects
(2) Design Defects
(3) Algorithmic Defects
(4) Interface Defects
(5) Performance Defects
(6) Documentation Defects

<u>Life Cycle Considerations</u>. Hardware reliability often assumes that the *hazard rate* (i.e., failure rate per unit time, often shortened to the failure rate) follows the "bathtub" curve, illustrated in Figure 9.2-1. Failures occur throughout the item's life cycle; the hazard rate initially is decreasing, then is uniform, and finally is increasing.

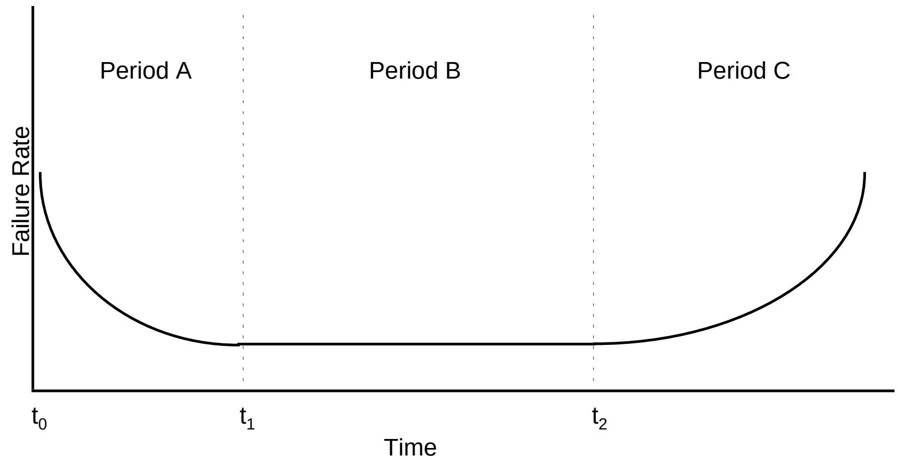


FIGURE 9.2-1: BATHTUB CURVE FOR HARDWARE RELIABILITY

The time points on the plot are defined as follows:

(1) Time t_0 is the time the population of components is activated or put into service ("fielded" or "distributed"); usually this is after the completion of development and production (whose times are not shown on the figure; i.e., design, build and test times are not included). Failures occurring during Period A, from t_0 to t_1, are said to be due to *infant mortality*.

(2) Time t_1 is the time when nearly all items with manufacturing defects have failed and have been removed from the population. Failures occurring during Period B, from t_1 to t_2, are assumed to be *random*, i.e., not due to any specific factor. The user is confident that the component will remain in service during this period. The probability that the component will function until time t_2 is expressed as the probability of success or the *reliability*.

(3) Time t_2 is the end of the *useful life* when components begin to exhibit end-of-life failures. Those failures occurring during Period C, after t_2, are considered to be due to *wearout*.

In hardware, the number of infant mortality failures observed in the field can be reduced by testing (*screening*) the components or assemblies prior to distribution (i.e., in the bathtub curve, the height of the curve in Period A can be reduced; alternatively the length of time attributable to infant mortality (Period A) can be reduced, causing t_1 to be moved closer to t_0). In the case of electronic components, this screen consists of operating, or *burning in*, the component for a time usually less than or equal to t_1. In the case of mechanical components, the screen may also include visual inspection. In addition, a random sample of the items may be tested to demonstrate adherence to specification. These procedures may be performed by the item manufacturer prior to distribution to ensure that shipped components have few or no latent failures. Otherwise, the purchasing organization takes the responsibility for these activities.

When modeling the failure characteristics of a hardware item, the factors which contribute to the random failures must be investigated. The majority are due to two main sources:

(1) *Operating stress* is the level of stress applied to the item. The operating stress ratio is the level of stress applied relative to its rated specification. For example, a resistor rated to dissipate 0.5 watts when actually dissipating 0.4 watts is stressed at 80% of rated. Operating stresses are well defined and measurable.

(2) *Environmental stresses* are considered to be those due to the specific environment (temperature, humidity, vibration, etc.) that physically affect the operation of the item being observed. For example, an integrated circuit having a rated temperature range of 0° to 70°C that is being operated at 50°C is within operational environment specification. Environmental stresses also can be well defined and measurable.

When transient stresses occur in hardware, either in the operating stresses or the environmental stresses, failures may be induced which are observed to be random failures. For this reason, when observing failures and formulating modeling parameters, care must be taken to ensure accurate monitoring of all of the known stresses.

The same "*bathtub*" curve for hardware reliability strictly does not apply to software since software does not typically wearout. However, if the hardware life cycle is likened to the software development through deployment cycle, the curve can be analogous for times up to t_2. For software, the time points are defined as follows:

(1) Time t_0 is the time when testing begins. Period A, from t_0 to t_1, is considered to be the *debug* phase. Coding errors (more specifically, errors found and corrected) or operation not in compliance with the requirements specification are identified and resolved. This is one key difference between hardware and software reliability. The "clock" is different. Development/test time is NOT included in the hardware reliability calculation but is included for software.

(2) Time t_1 is the initial *deployment* (distribution) time. Failures occurring during Period B, from t_1 to t_2, are found either by users or through post deployment testing. For these errors, work-arounds or subsequent releases typically are issued (but not necessarily in direct correspondence to each error reported).

(3) Time t_2 is the time when the software reaches the end of its useful life. Most errors reported during Period C, after t_2, reflect the inability of the software to meet the changing needs of the customer. In this frame of reference, although the software is still functioning to its original specification and is not considered to have failed, that

specification is no longer adequate to meet current needs. The software has reached the end of its useful life, much like the wearout of a hardware item. Failures reported during Period C may be the basis for generating the requirements for a new system.

Usually hardware upgrades occur during Period A, when initial failures often identify required changes. Software upgrades, on the other hand, occur in both Periods A and B. Thus, the Period B line is not really "flat" for software but contains many mini-cycles of Periods A and B: an upgrade occurs, most of the errors introduced during the upgrade are detected and removed, another upgrade occurs, etc. Hence, Figure 9.2-2 might be a better representation of the software life cycle. Although the failure rate drops after each upgrade in Period B, it may not reach the initial level achieved at initial deployment, t_1. Since each upgrade represents a mini development cycle, modifications may introduce new defects in other parts of the software unrelated to the modification itself. Often an upgrade focuses on new requirements; its testing may not typically encompass the entire system. Additionally, the implementation of new requirements may inversely impact (or be in conflict with) the original design. The more upgrades that occur, the greater the likelihood that the overall system design will be compromised, increasing the potential for increased failure rate, and hence lower reliability. This scenario is now occurring in many legacy systems which have recently entered Period C, triggering current reengineering efforts.

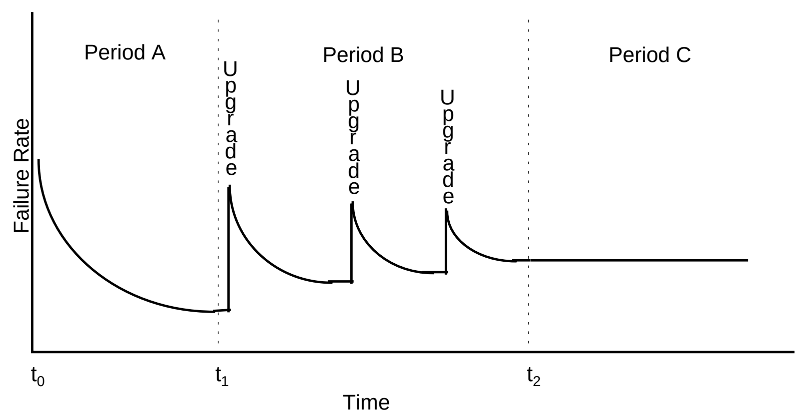


FIGURE 9.2-2: REVISED BATHTUB CURVE FOR SOFTWARE RELIABILITY

In software, the screening concept is not applicable since all copies of the software are identical. Additionally, typically neither operating stresses nor operational environment stresses affect software reliability. The software program steps through the code without regard for these factors. Other quality characteristics, such as speed of execution, may be effected, however. The end user might consider a "slow" program as not meeting requirements.

Table 9.2-2 summarizes the fundamental differences between hardware and software life cycles.

TABLE 9.2-2: SUMMARY: LIFE CYCLE DIFFERENCES

Life Cycle	**Pre t_0**	**Period A (t_0 to t_1)**	**Period B (t_1 to t_2)**	**Period C (Post t_2)**
HARDWARE	Concept Definition Development Build Test	Deployment Infant Mortality Upgrade	Useful Life	Wearout
SOFTWARE	Concept Definition Development Build	Test Debug/Upgrade	Deployment Useful Life Debug/Upgrade	Obsolescence

9.3 Software Design

Once the requirements have been detailed and accepted, the design will be established through a process of allocating and arranging the functions of the system so that the aggregate meets all customer needs. Since several different designs may meet the requirements, alternatives must be assessed based on technical risks, costs, schedule, and other considerations. A design developed before there is a clear and concise analysis of the system's objectives can result in a product that does not satisfy the requirements of its customers and users. In addition, an inferior design can make it very difficult for those who must later code, test, or maintain the software. During the course of a software development effort, analysts may offer and explore many possible design alternatives before choosing the best design.

Frequently, the design of a software system is developed as a gradual progression from a high-level or logical system design to a very specific modular or physical design. Many development teams, however, choose to distinguish separate design stages with specific deliverables and reviews upon completion of each stage. Two common review stages are the preliminary design and the detailed design.

9.3.1 Preliminary Design

Preliminary or high-level design is the phase of a software project in which the major software system alternatives, functions, and requirements are analyzed. From the alternatives, the software system architecture is chosen and all primary functions of the system are allocated to the computer hardware, to the software, or to the portions of the system that will continue to be accomplished manually.

During the preliminary design of a system, the following should be considered:

(1) Develop the architecture
- system architecture -- an overall view of system components
- hardware architecture -- the system's hardware components and their interrelations
- software architecture -- the system's software components and their interrelations

(2) Investigate and analyze the physical alternatives for the system and choose solutions

(3) Define the external characteristics of the system

(4) Refine the internal structure of the system by decomposing the high-level software architecture

(5) Develop a logical view or model of the system's data

9.3.1.1 Develop the Architecture

The architecture of a system describes its parts and the ways they interrelate. Like blueprints for a building, there may be various software architectural descriptions, each detailing a different aspect. Each architecture document usually includes a graphic and narrative about the aspect it is describing.

The software architecture for a system describes the internal structure of the software system. It breaks high-level functions into subfunctions and processes and establishes relationships and interconnections among them. It also identifies controlling modules, the scope of control, hierarchies, and the precedence of some processes over others. Areas of concern that are often highlighted during the establishment of the software architecture include: system security, system administration, maintenance, and future extensions for the system.

Another aspect of the software architecture may be the allocation of resource budgets for CPU cycles, memory, I/O, and file size. This activity often leads to the identification of constraints on the design solution such as the number of customer transactions that can be handled within a given period, the amount of inter-machine communication that can occur, or the amount of data that must be stored.

The first software architecture model for a system is usually presented at a very high level with only primary system functions represented. An example of a high-level software architecture is presented in Figure 9.3-1. As design progresses through detailed design, the architecture is continually refined.

9.3.1.2 Physical Solutions

Unless a software system has been given a pre-defined physical solution, an activity called environmental selection occurs during the preliminary design of a system. This is the process of investigating and analyzing various technological alternatives to the system and choosing a solution based upon the system's requirements, the users' needs, and the results of the feasibility studies. Aspects of a system that are generally selected at this time are: the hardware processing unit; computer storage devices; the operating system; user terminals, scanners, printers and other input and output devices; and the computer programming language.

In some cases, hardware and software items such as communications hardware and software, report writers, screen management systems, or database management systems are available "off-the-shelf." In other cases, unique requirements of the system may dictate the development of specific hardware and software items, specially designed for the system. The additional resources required to customize the system must be estimated and reviewed.

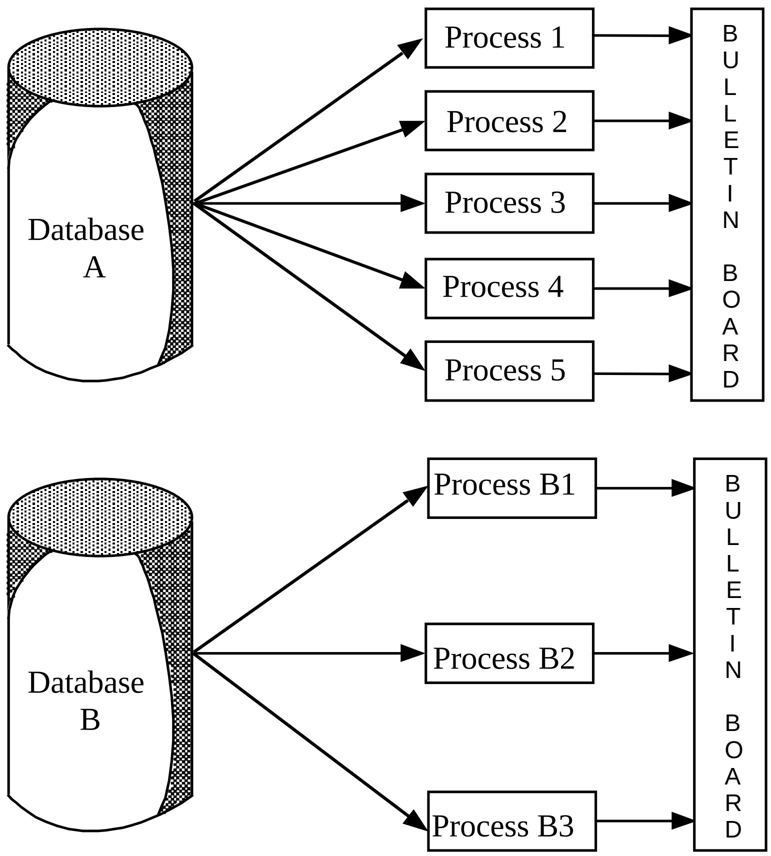


FIGURE 9.3-1: HIGH-LEVEL SOFTWARE ARCHITECTURE EXAMPLE

9.3.1.3 External Characteristics

Following the software system's functional allocation and physical environment selection, the details of the external or observable characteristics of a system can be developed. Included here would be terminal screen displays, report formats, error message formats, and interfaces to other systems.

A human factors engineer may be part of the design team concerned with the observable characteristics of a software system. This person specializes in the analysis of the human-machine interface. When a system's targeted users are novice computer users or when a system requires extensive manual data entry, human factors engineering can be a very important aspect of the design.

9.3.1.4 System Functional Decomposition

The activity of breaking a high-level system architecture into distinct functional modules or entities is called functional decomposition. When preparing to decompose a software system, the design team must decide what strategy they will use. Many decomposition strategies have been written about and are advocated; most the variations of the widely used top-down or bottom-up approaches. (Ref. [13]).

Top-down design is the process of moving from a global functional view of a system to a more specific view. Stepwise refinement is one technique used in top-down design. With this method, design begins with the statement of a few specific functions that together solve the entire problem. Successive steps for refining the problem are used, each adding more detail to the functions until the system has been completely decomposed.

A bottom-up design strategy for a software system is often used when system performance is critical. In this method, the design team starts by identifying and optimizing the most fundamental or primitive parts of the system, and then combining those portions into the more global functions. (Ref. [14] and [15]).

9.3.2 Detailed Design

Detailed design or low-level design determines the specific steps required for each component or process of a software system. Responsibility for detailed design may belong to either the system designers (as a continuation of preliminary design activities) or to the system programmers.

Information needed to begin detailed design includes: the software system requirements, the system models, the data models, and previously determined functional decompositions. The specific design details developed during the detailed design period are divided into three categories: for the system as a whole (system specifics), for individual processes within the system (process specifics), and for the data within the system (data specifics). Examples of the type of detailed design specifics that are developed for each of these categories are given below.

9.3.2.1 Design Examples

System specifics:

(1) Physical file system structure
(2) Interconnection records or protocols between software and hardware components
(3) Packaging of units as functions, modules or subroutines
(4) Interconnections among software functions and processes
(5) Control processing
(6) Memory addressing and allocation
(7) Structure of compilation units and load modules

Process specifics:

(1) Required algorithmic details
(2) Procedural process logic
(3) Function and subroutine calls
(4) Error and exception handling logic

Data specifics:

(1) Global data handling and access
(2) Physical database structure
(3) Internal record layouts
(4) Data translation tables
(5) Data edit rules
(6) Data storage needs

9.3.2.2 Detailed Design Tools

Various tools such as flowcharts, decision tables, and decision trees are common in detailed software design. Frequently, a structured English notation for the logic flow of the system's components is also used. Both formal and informal notations are often lumped under the term pseudocode. This is a tool generally used for the detailed design of individual software components. The terminology used in pseudocode is a mix of English and a formal programming language. Pseudocode usually has constructs such as "IF ..., THEN ...," or "DO ... UNTIL ...," which can often be directly translated into the actual code for that component. When using pseudocode, more attention is paid to the logic of the procedures than to the syntax of the notation. When pseudocode is later translated into a programming language, the syntactical representation becomes critical.

9.3.2.3 Software Design and Coding Techniques

Specific design and code techniques are related to error confinement, error detection, error recovery and design diversity. A summary of the each technique is included in Table 9.3-1 and Table 9.3-2.

TABLE 9.3-1: SOFTWARE DESIGN TECHNIQUES

Design Techniques
• Recovery designed for hardware failures • Recovery designed for I/O failures • Recovery designed for communication failures • Design for alternate routing of messages • Design for data integrity after an anomaly • Design for replication of critical data • Design for recovery from computational failures • Design to ensure that all required data is available • Design all error recovery to be consistent • Design calling unit to resolve error conditions • Design check on inputs for illegal combinations of data • Design reporting mechanism for detected errors • Design critical subscripts to be range tested before use • Design inputs and outputs within required accuracy

TABLE 9.3-2: SOFTWARE CODING TECHNIQUES

Coding Techniques	
• All data references documented • Allocate all system functions to a CSCI • Algorithms and paths described for all functions • Calling sequences between units are standardized • External I/O protocol formats standardized • Each unit has a unique name • Data and variable names are standardized • Use of global variables is standardized • All processes within a unit are complete and self contained • All inputs and outputs to each unit are clearly defined • All arguments in a parameter list are used • Size of unit in SLOC is within standard • McCabe's complexity of units is within standard • Data is passed through calling parameters • Control returned to calling unit when execution is complete	• Temporary storage restricted to only one unit - not global • Unit has single processing objective • Unit is independent of source of input or destination of output • Unit is independent of prior processing • Unit has only one entrance and exit • Flow of control in a unit is from top to bottom • Loops have natural exits • Compounded booleans avoided • Unit is within standard on maximum depth of nesting • Unconditional branches avoided • Global data avoided • Unit outputs range tested • Unit inputs range tested • Unit paths tested

9.4 Software Design and Development Process Model

Software development can occur with no formal process or structure (called "ad hoc" development) or it can follow one of several approaches (i.e., methods or models). Ad hoc development usually is the default used by relatively inexperienced developers or by those who only develop software as an aside or on rare occasions. As developers become more experienced, they tend to migrate from operating in an ad hoc fashion to using more formal structured methodologies. These major software development process models have evolved based upon actual practice. The selection is based upon several basic concepts, as summarized in

Table 9.4-1 and described throughout this section.

However, it is important to realize that what is actually being practiced may not fully correspond to the theory of any one model. In reality, developers often customize a model by implementing one or a combination of several elements of the models described. What is important is to understand enough about what constitutes the organization's software development process to be able to identify what characterizes the process used and to determine whether it is stable. The process that is in place will determine not only what data are available but also when they are available and whether they are adequate for determining the software reliability and quality performance levels as defined by the customer's contract requirements.

TABLE 9.4-1: SOFTWARE DEVELOPMENT PROCESS SELECTION

Approach	**When to Use**
Waterfall Model or Classic Development Model	When the detailed requirements are known, and are very stable When the type of application has been developed before When the type of software class (e.g., compilers or operating systems) has been demonstrated to be appropriate When the project has a low risk in such areas as getting the wrong interface or not meeting stringent performance requirements When the project has a high risk in budget and schedule predictability and control
Prototyping Approach	When the input, processing, or output requirements have not been identified To test concept of design or operation To test design alternatives and strategies To define the form of the man-machine interface
Spiral Model	To identify areas of uncertainty that are sources of project risk To resolve risk factors To combine the best features of the classic model and prototyping
Incremental Model	When a nucleus of functionality forms the basis for the entire system When it is important to stabilize staffing over the life of the project
Cleanroom Model	When a project can be developed in increments When staff size is sufficient to perform independent testing (staff > 6) When the approach has management support

9.4.1 Ad Hoc Software Development

The reality in many organizations where software development is not the main focus is that the development process is *ad hoc*. This is a polite way of saying that a defined structured process does not exist. The development effort is subject to the habits and operating styles of the individuals who comprise the project team. Responsibility for the project, and for interaction with the customer, is often in the hands of a non-software engineer. The software is viewed as having a supporting role to the project as a whole. Communication regarding requirements is primarily verbal and seldom documented. It is assumed that requirements are understood by all parties. Additionally, requirements change throughout the development effort. There is seldom a focus on design; design and code become merged into one task. Testing is the responsibility of the development team, and is often reduced to a random selection of functionality because there is no time to do a thorough job. Documentation, including design documents, is often written after the code is completed, and then reflects what was developed rather than serving as a guide for development. The project schedule is often determined by who is available to work rather than who is best qualified, the amount of dollars available, and an arbitrary completion date that typically is derived from something other than the functionality to be developed. The driving force is "having something to show by a specified date."

9.4.2 Waterfall Model

The *Waterfall Model* is presented in Figure 9.4-1. In its most simplistic interpretation it suggests that the process is strictly sequential, that there is a flow of ideas through the phases, with each phase having a distinct beginning and end and each phase enhancing the development to result in a software product that is operational when the bottom of the waterfall is reached.

The original intention of this model was that the development process is stable if all rework requires going back only one step in the process in order to be rectified. For example, if analysis revealed that initial requirements were incomplete then further requirements gathering would be implemented. If a particular design could not be coded correctly in the given environment then the design would be revisited. Testing would uncover coding errors which would be fixed before final delivery. The model suggests that the phases follow a time line, but this does not allow for revisiting previous phases when a problem is discovered.

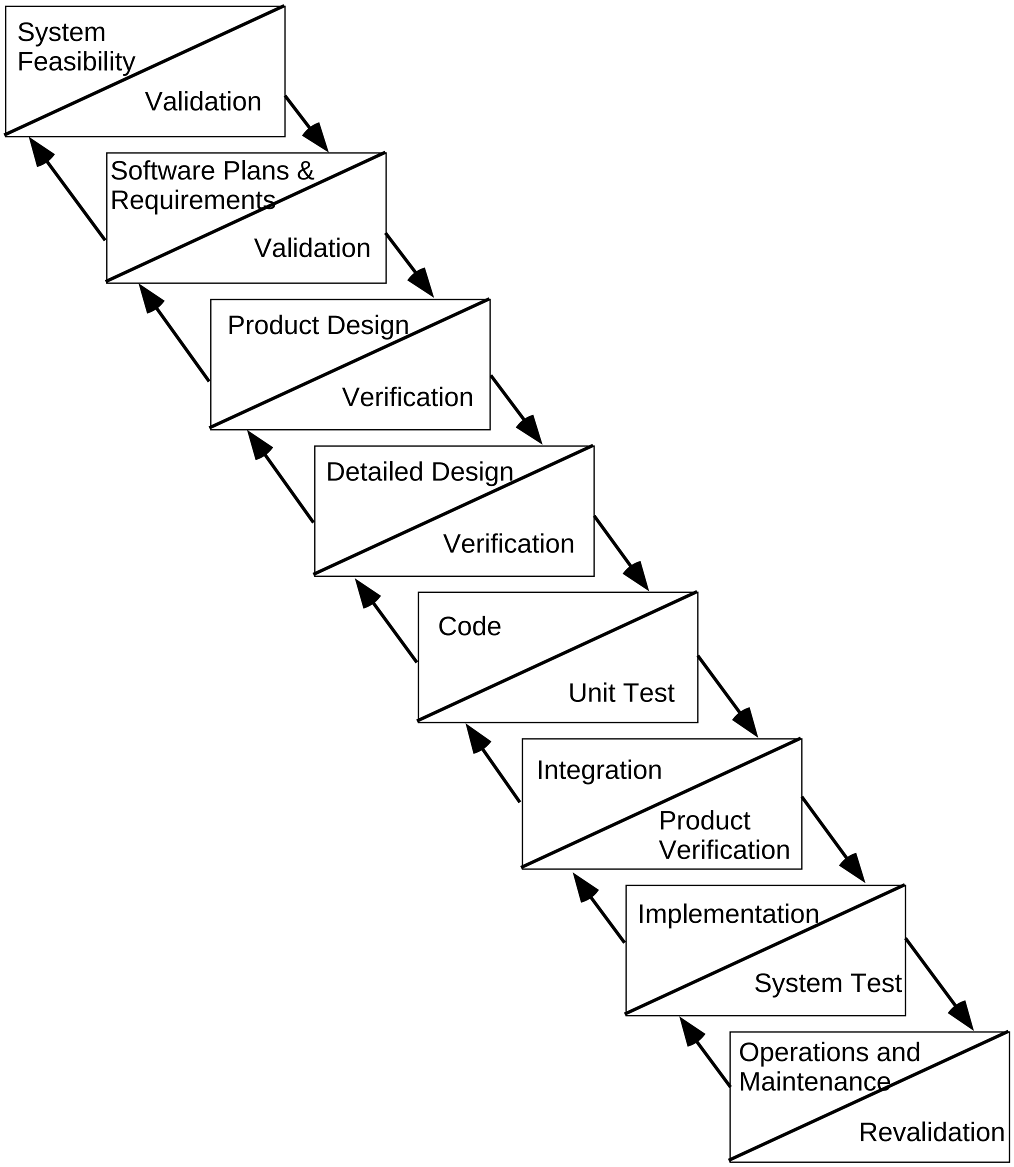
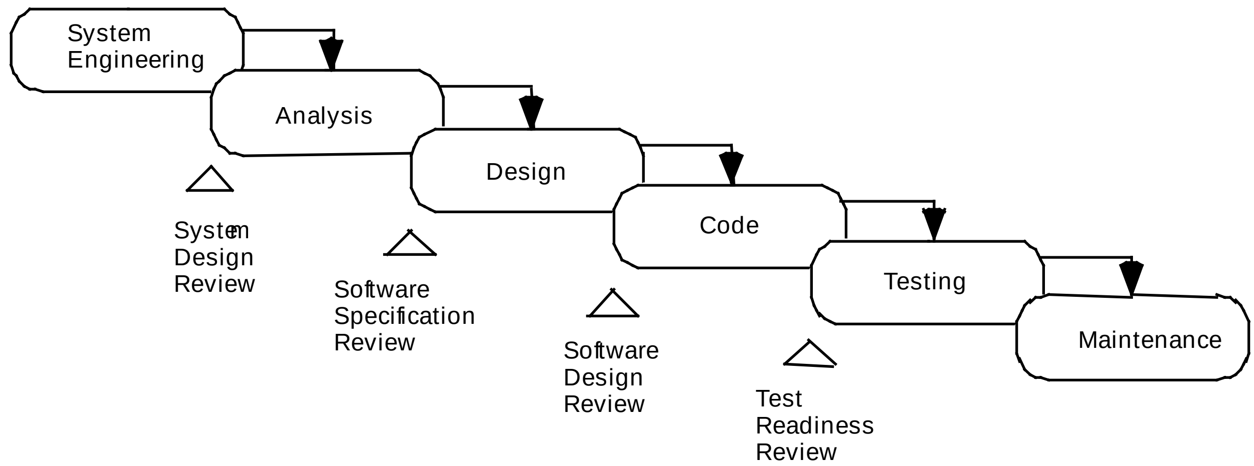


FIGURE 9.4-1: WATERFALL MODEL (REF. [6])

9.4.3 Classic Development Model

The *Waterfall Model* was later augmented to include precise phase ends and continuing activities, and has come to be known as the *Classic Development Model*; see Figure 9.4-2. This model provides a systemic approach to software development consisting of consecutive phases that begin with system engineering (sometimes called system requirements definition) and progress through requirements analysis, design, coding, testing, and maintenance. Each phase is defined in Figure 9.4-2.



PHASE	DESCRIPTION
System Engineering (sometimes called Requirements Definition)	When software is part of a larger system, work begins by establishing requirements for all system elements and then allocating some subset of these requirements to software. This is essential since software must interface with other elements such as hardware, people, and databases. Requirements are defined at the system level with a small amount of top-level design and analysis. It is during this phase that developers identify previously developed subsystems that can be reused on the current system.
Requirements Analysis	The requirements definition process is now intensified and focused specifically on the software. The development team performs functional or object-oriented analysis and resolves ambiguities, discrepancies, and to-be-determined (TBD) specifications. To understand the nature of the software to be built, the developers must understand the information domains for the software, as well as the required functions, performance, and interfaces. Requirements for both the system and the software are documented and reviewed with the sponsor/user.
Design	Software design is actually a multi-step process that focuses on four distinct attributes of the software: data structure, software architecture, procedural detail, and interface characterization. The design process translates requirements into a representation of the software that can be assessed for quality before coding begins. During this step, the developers perform structured, data driven, or object-oriented analysis. Like requirements, the design is documented and becomes part of the software configuration.
Code	The design is translated (coded) into a machine-readable form. If design has been performed in a detailed manner, coding can be accomplished mechanistically. The developers also reuse existing modules (or objects), with or without modification, and integrate it into the evolving system.
Test	Once new code has been generated or reused code has been modified, software testing begins. The unit test process focuses on the logical internals of the software, ensuring that all statements have been tested. The integration and system testing process focuses on the functional externals, testing to uncover errors and to ensure that the defined input will produce actual results that agree with required results. During acceptance testing, a test team that is independent of the software development team examines the completed system to determine if the original requirements are met. After testing the software is delivered to the customer.
Maintenance	Software may undergo change (one possible exception is embedded software) after it is delivered for several reasons (i.e., errors have been encountered, it must be adapted to accommodate changes in its external environment (e.g., new operating system), and/or customer requires functional or performance enhancements). Software maintenance reapplies each of the preceding phases, but does so in the context of the existing software.

FIGURE 9.4-2: THE CLASSIC DEVELOPMENT MODEL (REF. [7])

The *Classic Development Model* includes the notion of *validation* and *verification* at each of the phases. Validation is defined as testing and evaluating the integrated system to ensure compliance with the functional performance and interface requirements. Verification is defined as determining whether or not the product of each phase of the software development process fulfills all the requirements resulting from the previous phase. The purpose of the *validation* associated with the analysis and design model phases is to determine if the right product is being built. In revalidation activity that occurs after the software functionality has been defined, the purpose is to determine if the right product is still being built. *Verification* activity, associated with product design is to determine if the product is being built right, including the right components and their inter-combinations. This Classic Model has a definite and important role in software engineering history. It provides a template into which methods for analysis, design, coding, testing, and maintenance can be placed. It remains *the most widely used procedural model* for software engineering.

The classic model does have weaknesses. Among the problems that are sometimes encountered when the classic development process model is applied are:

(1) It emphasizes fully elaborated documents as completion criteria for early requirements and design phases. This does not always work well for many classes of software, particularly interactive end-user applications. Also, in areas supported by fourth-generation languages (such as spreadsheet or small business applications), it is unnecessary to write elaborate specifications for one's application before implementing it.

(2) Often the customer cannot state all requirements explicitly. The classic model requires this and has difficulty accommodating the natural uncertainty that exists at the beginning of many projects.

(3) The customer must have patience. A working version of the program is not available until late in the project schedule. Errors in requirements, if undetected until the working program is reviewed, can be costly.

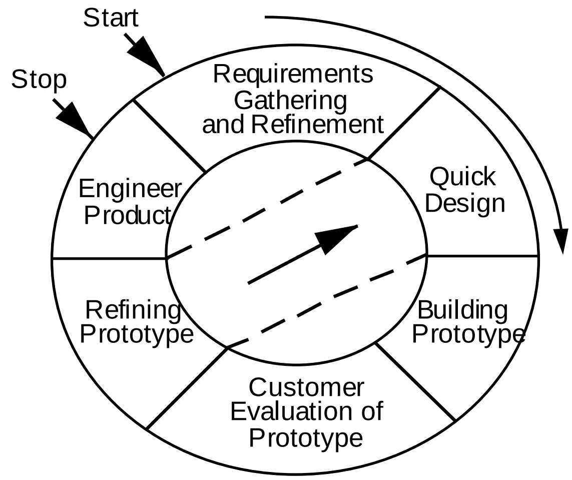
9.4.4 Prototyping Approach

Prototyping is a process that enables the developer to create a model of the software to be built. The steps for prototyping are identified and illustrated in Figure 9.4-3. The model can take one of three forms:

(1) A model that depicts the human-machine interaction in a form that enables the user to understand how such interaction will occur

(2) A working prototype that implements some subset of the functions required of the desired software

(3) An existing program that performs part or all of the functions desired, but has other features that will be improved upon in the new development effort



Step	Description
Requirements Gathering and Refinement	The developer and customer meet and define the overall objectives for the software, identify whatever requirements are known, and outline areas where further definition is mandatory.
Quick Design	The quick design focuses on a representation of those aspects of the software that will be visible to the user (e.g., user interface and output formats).
Prototype Construction	A prototype is constructed to contain enough capability for it to be used to establish or refine requirements, or to validate critical design concepts. If a working prototype is built, the developer should attempt to make use of existing software or apply tools (e.g., report generators, window manager) that enable working programs to be generated quickly.
Customer Evaluation	The prototype is evaluated by the customer and is used to refine requirements or validate concepts.
Prototype Refinement	The process of iteration occurs as the prototype is "tuned" to satisfy the needs of the customer, while at the same time enabling the developer to better understand what needs to be done.

FIGURE 9.4-3: STEPS IN THE PROTOTYPING APPROACH

Using an iterative rapid prototyping approach, the concept of the software system gradually unfolds; each iteration continues to explore the functionality that is desired. This process is comparable to performing "what if" analyses. The developer uses the prototype to generate suggestions from users, including ideas for innovations and plans for revision of the prototype itself, or the process it supports.

Rapid prototyping can significantly improve the quality and reliability of software if the methodology is properly used. However there are severe adverse impacts to quality and reliability when the developer or the customer perceives the prototype to be the completed project. Adversaries of prototyping claim that prototyping should not replace the traditional development cycle for these reasons:

(1) If a system is needed badly the prototype may be accepted in its unfinished state and pressed into service without necessary refinement. Eventually, as deficiencies are realized, a backlash is likely to develop, requiring maintenance efforts which are extremely costly compared to the cost of doing it right the first time.

(2) It tends to shape the approach to a capability before it is thoroughly understood.

(3) The real costs of supporting prototypes after delivery are not well documented. Therefore there is little evidence to support statements claiming that the cost of software is less for systems developed under rapid prototyping methods.

(4) Prototyping can be difficult to manage as a project within a larger project.

The key to successful prototyping is to define the rules of the game at the beginning; that is, the customer and developer must both agree that the prototype is built to serve as a mechanism for defining requirements or validating critical design concepts. It is then discarded (at least in part) and the actual software is engineered with an eye toward quality and maintainability.

9.4.5 Spiral Model

The *Spiral Model* for software development is presented in Figure 9.4-4. This model has been developed to encompass the best features of both the Classic Model and prototyping, while at the same time adding the element of risk analysis that is missing in both these process models. In the simplest sense it represents the normalization of a “trial and error” methodology. It is used to explore the possibilities in situations where a need exists but the exact requirements are not yet known.

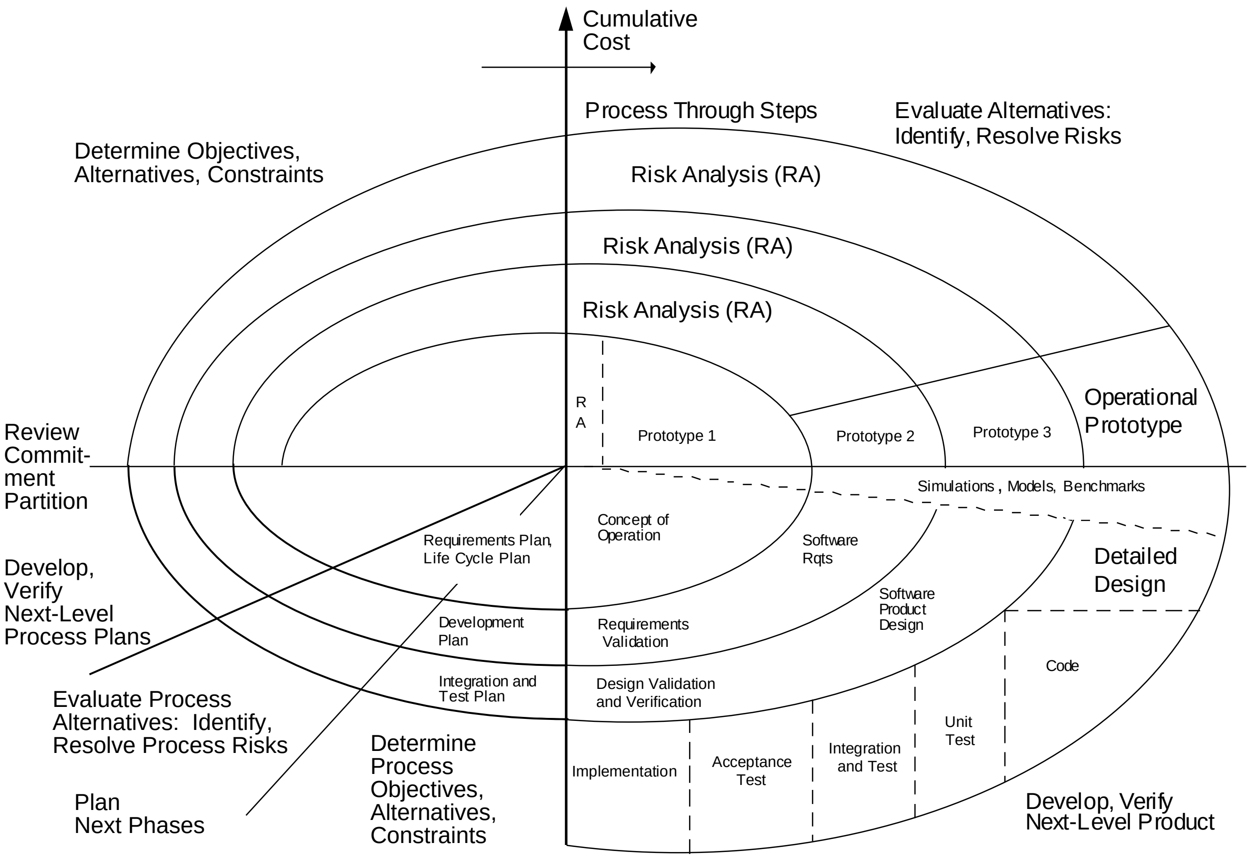


FIGURE 9.4-4: SPIRAL MODEL (REF. [7])

The model defines four major activities represented by the four quadrants (starting at upper left and progressing clockwise) of the figure:

(1) Planning: Determination of objectives, alternatives and constraints

(2) Risk analysis: Analysis of alternatives and identification/resolution of risks

(3) Engineering: Development of the “next-level” product

(4) Customer evaluation: Assessment of the results of engineering

With each iteration around the spiral (beginning at the center and working outward), progressively more complete versions of the software are built. During the first cycle around the spiral, objectives, alternatives, and constraints are defined and risks are identified and analyzed. If risk analysis indicates that there is uncertainty in requirements, prototyping may be used in the engineering quadrant to assist both the developer and the customer. Simulations and other models may be used to further define the problem and refine requirements. The customer evaluates the engineering work and makes suggestions for modification. Based on customer input, the next phase of planning and risk analysis occur. At each cycle around the spiral, the culmination of risk analysis results in a “go, no-go” decision. If risk is too great, the project can be terminated. However, if the flow around the spiral path continues, each path moves the developer outward toward a more complete model of the system, and, ultimately, to the operational system itself. Every cycle around the spiral requires engineering that can be accomplished using either the classic or prototyping approaches.

Like the other development process models, the spiral model is not a panacea. The following are some of the reasons why it is not right for all developments:

(1) It may be difficult to convince the sponsor that the evolutionary approach is controllable.

(2) It demands risk assessment expertise, and relies on this expertise for success.

(3) If major risk areas are not uncovered during risk analysis, problems will undoubtedly occur.

(4) The model itself is relatively new and has not been used as widely as the Classic or prototyping approaches. It will take a number of years before its effectiveness and efficiency can be determined with certainty.

9.4.6 Incremental Development Model

The *Incremental Development Model* can be followed using a sequential approach or an iterative approach. In a sequential approach, once a step has been completed, a developer never returns to that step or to any step previous to that step. In an iterative approach, if there is sufficient reason to do so, the developer may return to a previously completed step, introduce a change, and then propagate the effects of that change forward in the development. Projects actually can rarely follow the sequential forward flow. Iteration is generally necessary.

The *Incremental Development Model* is based on developing the software in increments of functional capability with a series of overlapping developments and a series of staggered deliveries. As indicated in Figure 9.4-5, each increment is developed under the phased approach described for the Classic Development Model. Each increment undergoes structural, or top-level design, detailed design, code and unit test, integration and test, and delivery. The nucleus of the software, the “cornerstone” functionality that is the foundation for use, must be addressed in the structural design of the first increment. Additional capability is then added with successive increments. Note that all software efforts do not lend themselves to incremental development because it is often not possible to distinguish a nucleus of functional capability.

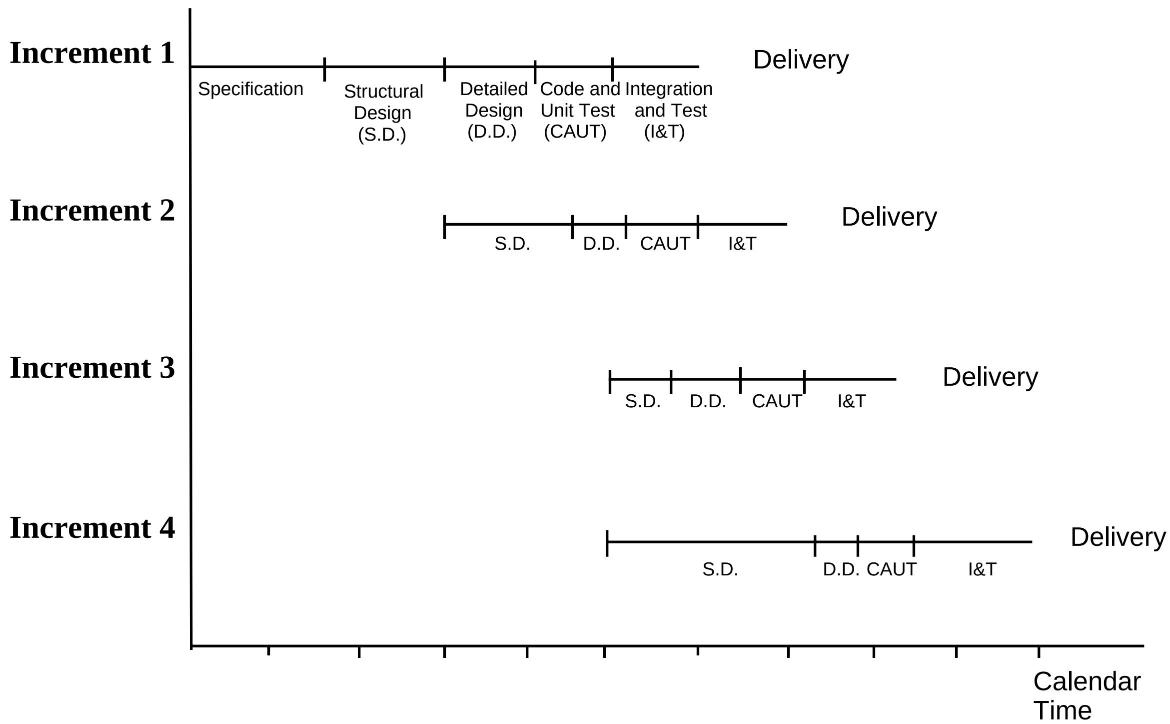


FIGURE 9.4-5: INCREMENTAL DEVELOPMENT MODEL (REF. [7])

Incremental development has been used successfully on many large projects. It is frequently used when the technical risks make it difficult to predict time scales for development, or when there is uncertainty about some aspects of the project. This approach also tends to level out or flatten the project's labor distribution curve. The design, program, and test teams can remain at relatively constant strength dealing with each increment in turn. Additionally, increments are easier to test and the cost of refinements is less expensive than with the single-shot Classic Development Model.

Incremental development is a useful approach when some functions within the software system have more stringent reliability requirements than others. Design efforts for a given increment can focus on attaining the desired reliability. Another feature of incremental development is that while the timeframe from project start to end may be identical to that of a project developed with the classic model, this model places operational software in the customer's hands long before project end.

9.4.7 Cleanroom Model

Cleanroom Software Engineering (Ref. [8] and [9]) (CSE) or just "*Cleanroom*" is a metaphor that comes from the integrated circuit manufacturing process where the environment must be free from all contaminants. If one were to rank all software development methodologies according to the amount of structure inherent in the methodology, the ad hoc development would be the lower bound (lack of structure) and cleanroom methodology would be the upper bound (very structured). Figure 9.4-6 illustrates the essential steps of the cleanroom development process.

The uniqueness of this approach is that it has embedded principles of total quality such as the use of teams, use of statistical process control techniques, and the commitment to "Do the right things right the first time" into the development process. The approach focuses on the aspects of the development that have the greatest impact on quality. Software reliability is specifically defined and measured as part of the certification process. Cleanroom Certification Test Teams provide scientific certification of software reliability -- they do not test it in.

Cleanroom methodology is premised on the notion that the best way to produce software approaching zero defects is to focus on defect prevention by clarifying requirements, developing precise functional and usage specifications, and then using them as the guide for planning and design, and for test development. It further presumes that correctness verification of the design will detect and eliminate most remaining significant defects before the software is actually built. The design effort entails writing pseudo code which is then subjected to correctness verification. The resulting pseudo code is so thorough and precise that it can be easily translated into the specified language. The actual coding is considered to be trivial relative to the development of pseudo code because the complex logic is addressed during pseudo code development.

In this methodology, the focus of testing reflects usage, not the structure of the software. Usage is inherent in the execution behavior of the software. Statistical Usage Testing is a process of testing software the way users intend to use it. The entire focus is on external system behavior, not on the internals of design and implementation. Test cases are randomly generated based on probability distributions that model anticipated software use in all possible circumstances including unusual and stressed situations. By definition, testing is designed to detect the more serious and/or high frequency defects first. Thus this testing method is more effective at improving software reliability in less time than traditional testing techniques. Data recorded includes execution time up to the point of each failure in appropriate units (measured in Central Processing Unit (CPU) time, clock time, or number of transactions, etc.). After execution of the test runs, the results are assessed and quality and performance measures computed.

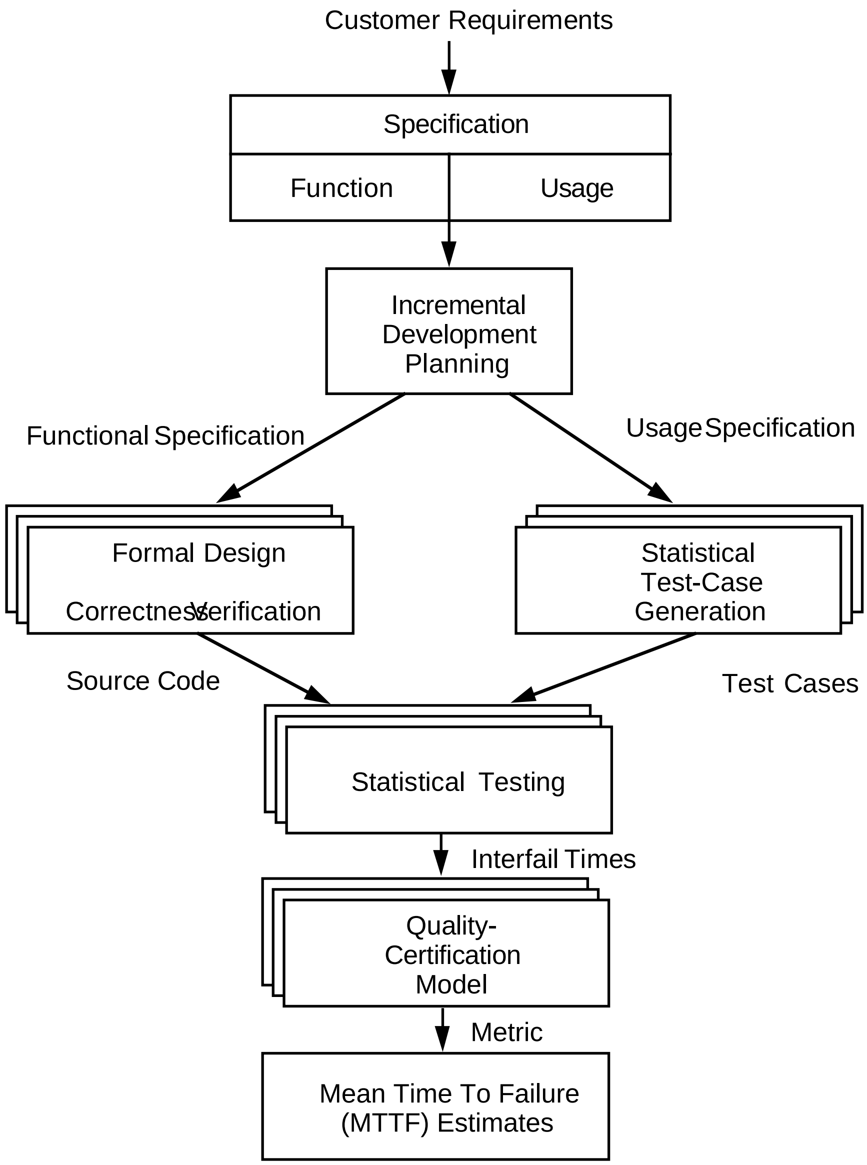


FIGURE 9.4-6: THE CLEANROOM DEVELOPMENT PROCESS (REF. [10])

Advocates for this methodology claim good results, namely that it is possible to produce software that approaches zero defects and deliver it on time and within budget. There is no way of knowing if a software system has zero defects, but as failure-free executions occur during testing, given the completeness of test coverage, one can conclude that there is a high probability that the software is at or near zero defects and will not fail during actual usage. The cleanroom approach has been adopted by more than 30 software development organizations as noted in the "Software Technology Support Center Guide (1995)." The cleanroom methodology has been applied to new systems, maintenance and evolution of existing systems, and re-engineering of problem systems. As of the end of 1993, cleanroom methodology used to develop a variety of projects totaling more than one million lines of code has shown extraordinary quality compared to traditional results. A summary of cleanroom performance measures is given in Table 9.4-2. It should be noted, however, that these results are achieved with cleanroom teams composed of adequately trained journeyman programmers.

TABLE 9.4-2: CLEANROOM PERFORMANCE MEASURES (REF. [11])

Software Development Practices	**Defects During Development** (defects per KLOC*)	**Operational Failures** (failures per KLOC)	**Resultant Productivity** (LOC*/Staff Month)
Traditional Software-as-art	50 - 60	15 - 18	Unknown
Software Engineering	20 - 40	2 - 4	75 - 475
Cleanroom Engineering	0 - 5	< 1	> 750

*** KLOC - Thousand Lines of Code * LOC - Lines of Code**

Adversaries claim that it is an unrealistic methodology for the following reasons:

(1) The required statistical knowledge is beyond the realm of most software engineers.

(2) The testing strategies are too complicated to expect the average developer to use.

(3) It is too complicated for use on small projects.

(4) The paradigm shift required is so radical that software people will never accept it.

Software Reliability Prediction and Estimation Models

Software reliability models have been in existence since the early 1970's; over 200 have been developed. Certainly some of the more recent ones build upon the theory and principles of the older ones. Some of the older models have been discredited based upon more recent information about the assumptions and newer ones have replaced them. This review of software reliability is not meant to be an exhaustive review of every model ever developed but, rather, a discussion of some of the major models in use today, highlighting issues important to the reliability engineer.

Prediction vs. Estimation Models

Software reliability modeling is generally used for one of two purposes: to make *predictions* and for *estimation*. Software reliability prediction models use historical data for similar systems while estimation models use data collected during test. Prediction, therefore, is usually less accurate than estimation. The objective of software prediction is to predict the potential reliability (fault rate) early in the development process. Insight into potential reliability allows improvements in software management to be considered before coding and testing start. The objective of the estimation process is to determine the number of faults remaining in the software just prior to testing so that the length of the test can be determined. Table 9.5-1 provides a comparison of prediction and estimation models.

TABLE 9.5-1: COMPARING PREDICTION AND ESTIMATION MODELS

Issues	Prediction Models	Estimation Models
Data Reference	Uses historical data	Uses data from the current software development effort
When Used In Development Cycle	Usually made prior to development or test phases; can be used as early as concept phase	Usually made later in life cycle (after some data have been collected); not typically used in concept or development phases
Time Frame	Predict reliability at some future time	Estimate reliability at either present or some future time

9.5.1 Prediction Models

The most basic prediction model involves the use of an organization's internal data, based on extensive experience and tracking, to develop predictions. Four other prediction models have been developed: *Musa's Execution Time Model,* (Ref. [12]), *Putnam's Model*, (Ref. [5]), and two models developed at Rome Laboratory and denoted by their technical report numbers: the *TR-92-52 Model* (Ref. [16]) and the *TR-92-15 Model* (Ref. [17]). Each prediction model, its capabilities and description of outputs is summarized in Table 9.5-2.

9.5.1.1 In-house Historical Data Collection Model

A few organizations predict software reliability by collecting and using the database of information accumulated on each of their own software projects. Metrics employed include Product, Project Management, and Fault indicators. Statistical regression analysis typically is used to develop a prediction equation for each of the important project characteristics. Management uses this information to predict the reliability of the proposed software product as well as to plan resource allocation.

TABLE 9.5-2: SOFTWARE RELIABILITY PREDICTION TECHNIQUES

Prediction Model	Capabilities	Description of Outputs
Historical Data Collection Model	Can be most accurate, if there is organization wide commitment.	Produces a prediction of the failure rate of delivered software based on company wide historical data.
Musa's Model	Predicts failure rate at start of system test that can be used later in reliability growth models.	Produces a prediction of the failure rate at the start of system test.
Putnam's Model	The profile of predicted faults over time and not just the total number is needed. Can be used with the other prediction models.	Produces a prediction in the form of a predicted fault profile over the life of the project.
TR-92-52 Model	Allows for tradeoffs.	Produces a prediction in terms of fault density or estimated number of inherent faults.
TR-92-15 Model	Has default factors for estimating number of faults	Estimates faults during each development phase.

9.5.1.2 Musa's Execution Time Model

Developed by John Musa (Ref. [12]) of Bell Laboratories in the mid 1970s, this was one of the earliest reliability prediction models. It predicts the initial failure rate (intensity) of a software system at the point when software system testing begins (i.e., when time, t = 0). The *initial failure intensity*, λ_0, (faults per unit time) is a function of the unknown, but estimated, total number of failures expected in infinite time, N. The prediction equation is shown below; terms are explained in Table 9.5-3.

$$\lambda_0 = k \times p \times w_o$$

For example, a 100 line (SLOC) FORTRAN program with an average execution rate of 150 lines per second has a predicted failure rate, when system test begins, of λ_0 = k x p x w_0 = (4.2E-7) x (150/100/3) x (6/1000) = .0126E-7 = 1.26E-9 faults per second (or 1 fault per 7.9365E8 seconds which is equivalent to 1 fault per 25.17 years).

It is important to note that this time measure is *execution time*, not calendar time. Since hardware reliability models typically are in terms of calendar time, it is not feasible to use Musa's prediction in developing an overall system reliability estimate unless one is willing to assume that calendar time and execution time are the same (usually not a valid assumption).

TABLE 9.5-3: TERMS IN MUSA'S EXECUTION TIME MODEL

Symbol	Represents	Value
k	Constant that accounts for the dynamic structure of the program and the varying machines	k = 4.2E-7
p	Estimate of the number of executions per time unit	p = r/SLOC/ER
r	Average instruction execution rate, determined from the manufacturer or benchmarking	Constant
SLOC	Source lines of code (not including reused code)	
ER	Expansion ratio, a constant dependent upon programming language	Assembler, 1.0; Macro Assembler, 1.5; C, 2.5; COBAL, FORTRAN, 3; Ada, 4.5
w_0	Estimate of the initial number of faults in the program	Can be calculated using: w_0 = N x B or a default of 6 faults/1000 SLOC can be assumed
N	Total number of inherent faults	Estimated based upon judgment or past experience
B	Fault to failure conversion rate; proportion of faults that become failures. Proportion of faults not corrected before the product is delivered.	Assume B = .95; i.e., 95% of the faults undetected at delivery become failures after delivery

9.5.1.3 Putnam's Model

Trachtenberg (formerly of General Electric) and Gaffney (of then IBM Federal Systems, now Loral) examined defect histories, by phases of the development process, for many projects of varying size and application type. Based on their work, Putnam (Ref. [5]) assigned the general normalized Rayleigh distribution to describe the observed reliability, where k and a are constants fit from the data and t is time, in months:

$$R(t) = k \exp(-at^2)$$

The corresponding probability density function, f(t), the derivative of R(t) with respect to t, is of the general form:

$$f(t) = 2ak\, t \exp(-at^2)$$

Putnam further developed an ordinal (i.e., not equally spaced in real time) scale to represent the development process milestones; see Table 9.5-4. Of special interest is Milestone 7, denoted by t_d, corresponding to the end of the development phases and the beginning of full operational capability; this point was defined as occurring at the 95th percentile (i.e., 95% of all defects have been detected at this point in the software development). Using t_d as the reference basis, he then developed the expressions for the model constants, a and k, in terms of N and t_d. The final equation to predict the expected number of defects per month as a function of the schedule month and the total number of inherent defects, N, is given by:

$$f(t) = (6N/t_d^2)\ t \exp(-3t^2/t_d^2)$$

TABLE 9.5-4: PUTNAM'S TIME AXIS MILESTONES

Milestone #	Milestone
0	Feasibility study
1	Preliminary design review, function design complete
2	Critical design review, detailed design complete
3	First code complete
4	Start of system integration test
5	Start of user systems test
6	Initial operational capability; installation
7	Full operational capability; reliability about 95% in routine usage
8	99% reliability achieved by stress testing
9	99.9% reliability, assumed debugged

For example, suppose a FORTRAN program is being developed; the plan is that it will be fully operational (Milestone 7) in 10 calendar months resulting in t_d^2 to be 10^2 or 100. The defects expected per month during development are calculated using the expression:

$$f(t) = .06\ N\ t \exp(-.03t^2)$$

Calculation results are shown in Figure 9.5-1, where t is the month number, f(t) is the expected proportion of the total number of defects to be observed in month t, and F(t) represents the cumulative proportion. The Milestone number, based on the planned development schedule is also shown for comparison; Milestone 7, corresponding to the 95th percentile, is, indeed, in Month 10, Milestone 8, at the 99th percentile, is expected to occur in scheduled Month 13, and Milestone 9, at .999, is not expected to be reached by the end of scheduled Month 15.

t	f(t)	F(t)	Mile #
1	0.058	0.058	
2	0.106	0.165	1
3	0.137	0.302	
4	0.149	0.451	2
5	0.142	0.592	
6	0.122	0.715	3
7	0.097	0.811	4
8	0.070	0.881	5
9	0.048	0.929	6
10	0.030	0.959	7
11	0.017	0.976	
12	0.010	0.986	
13	0.005	0.991	8
14	0.002	0.993	
15	0.001	0.994	

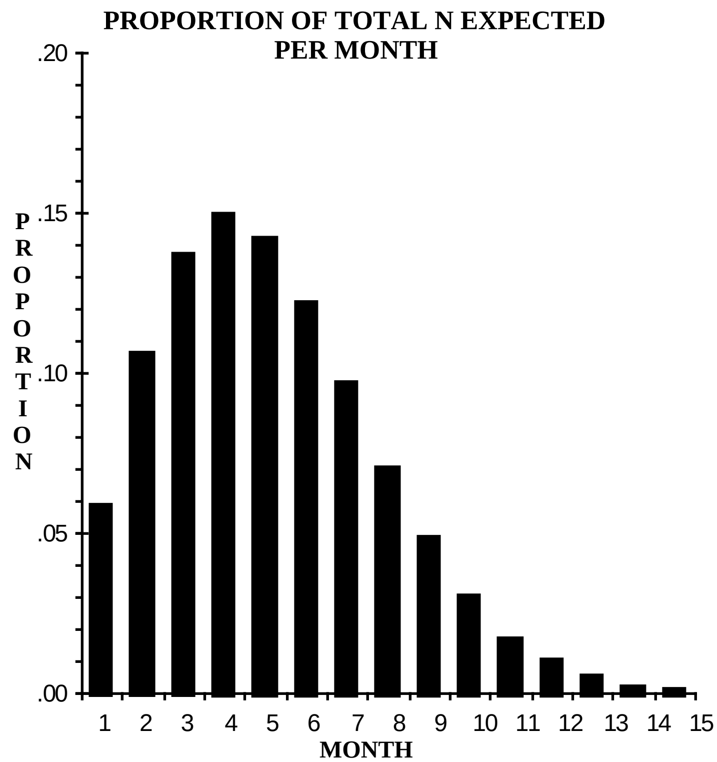


FIGURE 9.5-1: EXPECTED PROPORTION OF THE TOTAL NUMBER OF DEFECTS

One major benefit of this model is that the expected number of faults can be predicted for various points in development process as compared to Musa's model that provides the prediction when system testing begins (i.e., at Milestone 4) only.

Another corollary to this model is that the mean time to the next defect (MTTD) is given by 1/f(t). This is only meaningful after Milestone 4 (since prior to that point the system would not have been developed so defects could not be detected). As the development progresses, (i.e., t increases), the MTTD increases since defects are being eliminated.

9.5.1.4 Rome Laboratory Prediction Model: RL-TR-92-52 (Ref. [16])

This is a method for predicting *fault density at delivery time* (i.e., at Putnam's Milestone 6) and subsequently using this fault density to predict the *total number of inherent faults*, N, and the *failure rate*. It also provides a mechanism for allocating software reliability as a function of the software characteristics as well as assessing trade-off options. The basic terminology of this model is presented in Table 9.5-5. The underlying assumption is that Source Lines of Code (SLOC) is a valid size metric.

TABLE 9.5-5: RL-TR-92-52 TERMINOLOGY

Terms	Description
A	Factor selected based on Application type; represents the baseline fault density
D	Factor selected to reflect the Development environment
S	Factor calculated from various "sub-factors" to reflect the Software characteristics
SLOC	The number of executable Source Lines Of Code; lines that are blank or contain comments to enhance the readability of the code are excluded
FD	Fault density; for the purposes of this model it is defined as the ratio of faults to lines of code (faults/SLOC)
N	Estimate of total number of inherent faults in the system; prediction is derived from the fault density and the system size
C	Factor representing a Conversion ratio associated with each application type; values are determined by dividing the average operational failure rate by the average fault density in the baseline sample set.

It is recognized as one of a few publicly available prediction models based upon extensive historical information. Predictions are based on data collected on various types of software systems developed for the Air Force; see Table 9.5-6.

TABLE 9.5-6: AMOUNT OF HISTORICAL DATA INCLUDED

Application Type	# of Systems	Total SLOC
Airborne	7	540,617
Strategic	21	1,793,831
Tactical	5	88,252
Process Control	2	140,090
Production Center	12	2,575,427
Developmental	6	193,435
TOTAL	53	5,331,652

The basic equations are:

Fault Density = FD = A x D x S (faults/line)
Estimated Number of Inherent Faults = N = FD x SLOC
Failure Rate = FD x C (faults/time)

The model consists of factors that are used to predict the fault density of the software application. These factors are illustrated in Table 9.5-7.

TABLE 9.5-7: SUMMARY OF THE RL-TR-92-52 MODEL

Factor	Measure	Range of Values	Phase Used In*	Trade-off Range
A - Application	Difficulty in developing various application types	2 to 14 (defects/KSLOC)	A-T	None - Fixed
D - Development organization	Development organization, methods, tools, techniques, documentation	.5 to 2.0	If known at A, D-T	The largest range
SA - Software anomaly management	Indication of fault tolerant design	.9 to 1.1	Normally, C-T	Small
ST - Software traceability	Traceability of design and code to requirements	.9 to 1.0	Normally, C-T	Large
SQ - Software quality	Adherence to coding standards	1.0 to 1.1	Normally, C-T	Small
SL - Software language	Normalizes fault density by language type	Not applicable	C-T	N/A
SX - Software complexity	Unit complexity	.8 to 1.5	C-T	Large
SM - Software modularity	Unit size	.9 to 2.0	C-T	Large
SR - Software standards review	Compliance with design rules	.75 to 1.5	C-T	Large

Key **A - Concept or Analysis Phase**

D - Detailed and Top Level Design
C - Coding
T - Testing

The following are benefits of using this model:

(1) It can be used as soon as the concept of the software is known

(2) During the concept phase, it allows "what-if" analysis to be performed to determine the impact of the development environment on fault density

(3) During the design phase, it allows "what-if" analysis to be performed to determine the impact of software characteristics on fault density

(4) It allows for system software reliability allocation because it can be applied uniquely to each application type comprising a software system

(5) The prediction can be customized using unique values for the A, S, and D factors based upon historical software data from the specific organization's environment while the following are drawbacks:

(a) Factors and values used were generated based on software developed for the Air Force; if the software in question does not match one of the Air Force-related application types, then the average value must be selected. The Air Force application types do not map well to software developed outside the military environment

(b) Use of SLOC as the size metric is becoming more and more irrelevant with recent changes in software development technology, such as Graphical User Interface (GUI) system development, and the use of Commercial Off-the-Shelf (COTS) software

9.5.1.5 Rome Laboratory Prediction Model: RL-TR-92-15 (Ref. [17])

This technical report, produced by Hughes Aircraft for Rome Laboratory, examined many software systems. It resulted in an *average fault rate prediction* value of 6 faults/1000 SLOC. (This was the default value for fault rate, w_0, used in Musa's Execution Time Model).
In addition, a set of 24 predictor factors, listed in Table 9.5-8, was used to estimate the three main variables of interest:

(1) Number of faults detected during each development phase (DP)

(2) Man-hours utilized during each phase (UT)

(3) Size of product (S)

The resultant equations were:

(1) $f(DP) = 18.04 + .05 \text{ x } (.009 X_1 + .99 X_2 + .10 X_3 - .0001 X_4 + .0005 X_5)$

(2) $f(UT) = 17.90 + .04 \text{ x } (.007 X_1 + .796 X_2 + .08 X_3 - .0003 X_4 + .0003 X_5 + .00009 X_6 + .0043 X_7 + .013 X_8 + .6 X_9 + .003 X_{10})$

(3) $f(S) = 17.88 + .04 \text{ x } (.0007 X_1 + .8 X_3 + .01 X_8 + .6 X_9 + .008 X_{23} + .03 X_{25})$

where the coefficients and descriptions of the variables of the regression model are listed in the table.

TABLE 9.5-8: REGRESSION EQUATION COEFFICIENTS

X	Description of Variable	Coefficients EQ 1	EQ 2	EQ 3
1	Number of faults in software requirements specification	.009	.007	.007
2	Requirements statement in specification	.99	.796	NA
3	Pages in specification	.10	.08	.80
4	Man-months spent in requirements analysis	.0001	-.0003	NA
5	Requirements change after baseline	.0005	.0003	NA
6	Number of faults in preliminary design document	NA	.00009	NA
7	Number of CSCS	NA	.0043	NA
8	Number of units in design	NA	.013	.01
9	Pages in design document	NA	.6	.6
10	Man-months spent in preliminary design	NA	.003	NA
11	Number of failures in design document	NA	NA	NA
12	Man-months spent in detailed design	NA	NA	NA
13	Design faults identified after baseline	NA	NA	NA
14	Design faults identified after internal review	NA	NA	NA
15	Number of executable SLOC	NA	NA	NA
16	Faults found in code reviews	NA	NA	NA
17	Average years of programmer experience	NA	NA	NA
18	Number of units under review	NA	NA	NA
19	Average number of SLOC per unit	NA	NA	NA
20	Average number branches in unit	NA	NA	NA
21	Percentage branches covered	NA	NA	NA
22	Nesting depth coverage	NA	NA	NA
23	Number of times an unit is unit tested	NA	NA	.008
24	Man-months for coding and unit test	NA	NA	NA
25	Equals (X13 + X14 + X16)	NA	NA	.03

The results indicate that thirteen of the 24 hypothesized factors had no effect on the three variables of interest. Further, the most important estimators involved the software requirements specification, including the number of requirement statements, number of faults in these statements, and the total number of pages in the specification.

The benefits of this model are:

(1) It can be used prior to system testing to estimate reliability

(2) It includes cost and product parameters as well as fault and time

The disadvantages of this model are:

(1) It was based on data collected by one organization in one industry/application type

(2) It does not disclose the unit of measure for specification size

9.5.2 Estimation Models

The fault count and fault rate models are the most common type of estimation techniques. Each makes assumption about how faults arrive (detected) and how they are corrected. The fault count models include: Exponential, Weibull and Bayesian techniques. Also, included in the estimation model scenario are the test coverage and fault tagging methods.

9.5.2.1 Exponential Distribution Models

In general, *exponential models* assume that the software is in an operational state and that all faults are independent of each other. The time to failure, t, of an individual fault follows the exponential distribution:

$$f(t) = \lambda \exp(-\lambda t)$$

with the general form for the reliability given by:

$$R(t) = \exp(-\lambda t)$$

and the mean time to the next failure (MTTF) expressed as:

$$MTTF = 1/\lambda$$

The notations used in the general case for the exponential distribution model are shown in Table 9.5-9 and illustrated in Figure 9.5-2.

TABLE 9.5-9: NOTATIONS FOR THE EXPONENTIAL DISTRIBUTION MODEL

Notation	Explanation
N	Total number of defects
n	Number of defects to date
c	Number of defects corrected to date
N-n	Defects not yet manifested
N-c	Defects yet to be corrected
n_f	Fault count
λ_f	Fault rate
t_f	Future time
n_p	Fault count at present time
λ_p	Fault rate at present time
t_p	Present time

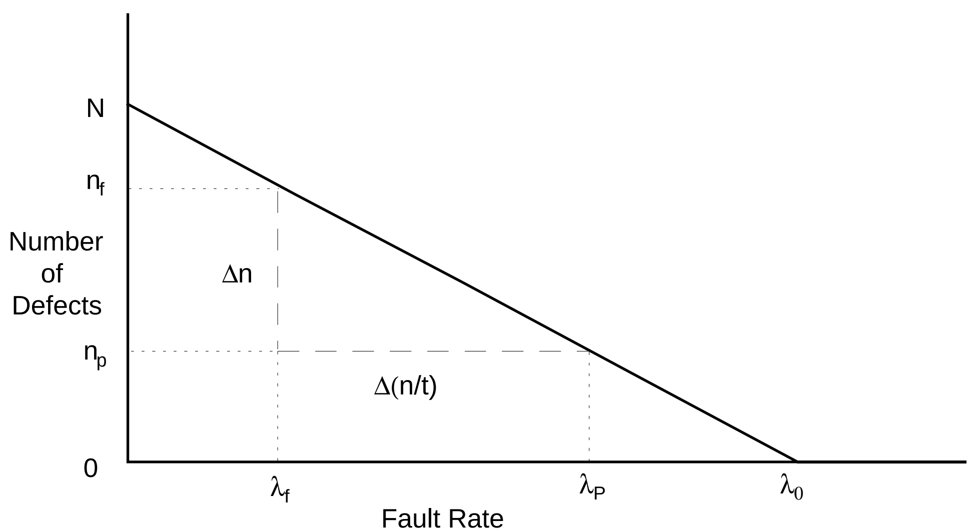


FIGURE 9.5-2: EXPONENTIAL MODEL BASIS

Advocates of the exponential model note its simplicity and its parallelism to the hardware reliability framework. The major disadvantage is that it cannot be used early in the software development since the product must be operational before the model can be used. Hence, it cannot be used for early reliability assessment.

Table 9.5-10 summarizes the various exponential models including assumptions and comments.

TABLE 9.5-10: VARIOUS EXPONENTIAL MODELS*

Model	MTTF	Dn	Dt	Assumptions/Comments
General Exponential	1/[k(N - c)]	$k^{-1} \lambda_p / \lambda_f$	$k^{-1} \mathit{ln} (\lambda_p / \lambda_f)$	Faults are equal in severity and probability of detection Fault rate directly related to number of faults remaining to be corrected
Lloyd-Lipow	1/[k(N - n)]	$k^{-1} \lambda_p / \lambda_f$	$k^{-1} \mathit{ln} (\lambda_p / \lambda_f)$	Fault rate directly related to number of faults remaining to be detected
Musa’s Basic		$N/\lambda_0 (\lambda_p - \lambda_f)$	$N/\lambda_0 \mathit{ln} (\lambda_p - \lambda_f)$	References an initial fault rate at time 0 (beginning of system test)
Musa’s Logarithmic		$f^{-1} \mathit{ln} (\lambda_p / \lambda_f)$	$f^{-1} (1/\lambda_f - 1/\lambda_p)$	Some faults are more likely to be found before others Rate of fault detection decreases exponentially
Shooman’s	1/[kSLOC - ((N/SLOC) - (C/SLOC))]	$k^{-1} \lambda_p / \lambda_f$	$k^{-1} \mathit{ln} (\lambda_p / \lambda_f)$	Adjusts for changing product size; each parameter is normalized for lines of code
Goel-Okumoto				Faults can cause other faults Faults may not be removed immediately

* where k is a constant of proportionality, N is the number of inherent faults, c is the number of corrected faults and n is the number of detected faults

<u>General Exponential Model</u>

In the general case, the model assumes that all faults are equal in severity and probability of detection and that each is immediately corrected upon detection. The fault rate, λ, is assumed to be directly related to the number of faults remaining in the software. That is, λ is a function of the number of corrected faults, c:

$$\lambda = k\,(N - c)$$

where k is a constant of proportionality. In actual application, k is typically estimated from the slope of the plot of the observed fault rate vs. number of faults corrected.

The projection of the number of faults needed to be detected to reach a final failure rate λ_f is given by:

$$\Delta n = (1/k)\, \lambda_p / \lambda_f$$

where k is the same proportionality constant used above.

The projection of the time necessary to reach a projected fault rate is given by:

$$\Delta t = (1/k) \ln \left[\lambda_p / \lambda_f \right]$$

The major disadvantage of this specific approach is that not only must the defects be detected but they also must be corrected.

Lloyd-Lipow Model (Ref. [18] and [19])

The *Lloyd-Lipow Model* exponential model also assumes that all faults are equal in severity and probability of detection. The difference from the previous model is that in this Lloyd-Lipow approach, the fault rate, λ, is assumed to be directly related to the number of faults remaining to be **detected** (not corrected) in the software. That is, λ is a function of the number of detected faults, n:

$$\lambda = k\,(N - n)$$

The expressions for the mean-time-to-failure (MTTF), Δn, and Δt are the same as in the general exponential model.

This form of the exponential model does not require defect correction, just detection. However, the validity of the use of the exponential model in this situation has been questioned.

Musa's Basic Model (Ref. [12])

Musa's Basic Model is another form of the general exponential model. It utilizes the initial (i.e., at the start of software testing) fault rate, λ_0, where either λ_0 is estimated from the data or computed (λ_0 = N/k) based on a guess for N and the estimate for k, the previously referenced slope value.

In this model, the fault rate after n faults have been detected is a fraction of the original fault rate:

$$\lambda_n = \lambda_0 (1 - n/v)$$

where:

n is usually expressed as μ and v is usually expressed as υ

while the expression for the fault rate at time t is given by:

$$\lambda_t = \lambda_0 \exp\left[-(\lambda_0/\upsilon)\tau\right]$$

where:

υ = N/B, where N is the number of inherent faults and B is the fault reduction ratio, usually assumed to be 95% (i.e., 95% of the faults undetected at delivery become failures after delivery

τ = System test time

The projection of the number of faults needed to be detected to reach a final failure rate λ_f is given by:

$$\Delta n = N/\lambda_0 \, (\lambda_p - \lambda_f)$$

The projection of the time necessary to reach a projected failure rate is given by:

$$\Delta t = N/\lambda_0 \ln (\lambda_p - \lambda_f)$$

The disadvantage is that this model is very sensitive to deviations from the assumptions. In addition, as noted with Musa's previous work, the units are execution time, not calendar time.

Musa's Logarithmic Model (Ref. [12])

Musa's Logarithmic Model has different assumptions than the other exponential models:

(1) Some faults are likely to be found before others

(2) The rate of fault detection is not constant, but decreases exponentially

In this model, the fault rate after n faults have been detected is a function of the original fault rate:

$$\lambda_n = \lambda_0 \exp(-ft)$$

while the expression for the fault rate at time t is given by:

$$\lambda_t = \lambda_0/(\lambda_0 f \; t + 1)$$

where:

f = failure intensity decay parameter, the relative change of n/t over n.

The projection of the number of faults needed to be detected to reach a final failure rate λ_f is given by:

$$\Delta n = 1/f \ \ln(\lambda_p/\lambda_f)$$

The projection of the time necessary to reach a projected failure rate is given by:

$$\Delta t = 1/f \ (1/\lambda_f - 1/\lambda_p)$$

The major benefit of this model is that it does not require an estimate for N. Since the value for f can be estimated prior to actual data occurrence, the model can be used earlier in the development cycle to estimate reliability.

The disadvantage of this model, typical for most exponential models, is that the model assumptions must be valid for the results to be valid. In particular, the assumption that the rate of fault detection decreases exponentially has not been confirmed with many real data sets. In addition, as noted with Musa's previous work, the units are execution time, not calendar time, making direct comparison with hardware reliability difficult.

<u>Shooman's Model (Ref. [20])</u>

The *Shooman's Model* is similar to the general exponential model except that each fault count is normalized for the lines of code at that point in time. Earlier, $\lambda = k$ (N - c); here, it is given by:

$$\lambda = k \ \text{SLOC} \ (\text{N/SLOC} - \text{c/SLOC})$$

The equation of MTTF = $1/\lambda$ uses Shooman's expression for λ. The equations for Δn and Δt are the same as the general exponential case.

The advantage of Shooman's model is that it adjusts for the changing size of the software product. The disadvantages are that it must be used later in development after the LOC have been determined and that the general exponential assumptions may not apply.

<u>Goel-Okumoto Model (Ref. [19])</u>

This model is different from other exponential models because it assumes that faults can cause other faults and that they may not be removed immediately. An iterative solution is required. This model is expressed as:

$$\lambda_t = ab\exp(-bt)$$

where a and b are resolved iteratively from the following:

$$n/a = 1 - \exp(-bt) \quad \text{and} \quad n/b = a\,t\,\exp(-bt) + \sum t_i\ ,$$

where the summation is over i = 1,..., n,

Use N and k as starting points for solving for these two equations simultaneously.

The major benefit of this model is that it can be used earlier than other exponential models while its major disadvantage is that it is very sensitive to deviations from the assumptions.

9.5.2.2 Weibull Distribution Model (Ref. [19])

The *Weibull Model* is one of the earliest models applied to software. It has the same form as that used for hardware reliability. There are two parameters: a, the scale parameter (a > 0), and b, the shape parameter that reflects the increasing (b > 1), decreasing (b < 1) or constant (b = 1) failure rate.

The mean time to next failure is given by:

$$\text{MTTF} = (b/a)\,\Gamma(1/a)$$

where Γ (c) is the complete Gamma Function $= \int_0^{\infty} y^{c-1} e^{-y}\,dy$

The reliability at time t is given by:

$$R(t) = \exp[-(t/b)^{a}]$$

The benefits of the Weibull model is its flexibility to take into account increasing and decreasing failure rates. The disadvantage of this model is more work is required in estimating the parameters over the exponential model.

9.5.2.3 Bayesian Fault Rate Estimation Model

The *Bayesian* approach does not focus on the estimated inherent fault count, N, but rather concentrates on the fault/failure rate. The classical approach assumes that reliability and failure rate are a function of fault detection while the Bayesian approach, on the other hand, assumes that a software program which has had fault-free operation is more likely to be reliable. The Bayesian approach also differs because it is possible to include an assessment of "prior knowledge" (therefore, it is sometimes called a "subjective" approach).

The Thompson and Chelson's model (Ref. [19]) assumes that:

(1) Software is operational

(2) Software faults occur at some unknown rate λ that is assumed to follow a Gamma Distribution with parameters X_i and $f_i + 1$

(3) Faults are corrected in between test periods but not during test periods

(4) Total number of faults observed in a single testing period of length t_i follows a Poisson distribution with parameter λt_i

The model assumes that there are i test periods, each with length, t_i (not assumed equal); where the number of faults detected during that period is represented by f_i. The subjective information is inserted as occurring in period 0, i.e., t_0 and f_0 represent the prior information. If there is no prior information, these values are set to zero. On the other hand, if there is a great deal of experience, t_0 might be very large, especially relative to the expected evaluation time; the value for the prior number of faults also depends on past experience and is independent from the prior of time.

Let T_i represent the cumulative total of the test period lengths over the entire range, i.e., from period 0 to i and let F_i represent the cumulative total of the faults, f_i, over the entire range, i.e., from period 0 to i.

Then, the reliability at time t (in interval i) is expressed as a function of the values from the previous (i - 1) interval as well as the current i^{th} interval data:

$$R(t) = [T_{i-1}/(T_{i-1} + t)]^{F_{i-1}}$$

The failure rate estimate at time t (in interval i) is given by:

$$\lambda(t) = (F_{i-1} + 1)/T_{i-1}$$

The benefits of this model are related to its assumptions: N is not assumed fixed, reliability is not assumed to be directly a function of N, and faults are not assumed to be corrected immediately. The disadvantage is that Bayesian Models, in general, are not universally accepted since they allow for the inclusion of prior information reflecting the analyst's degree of belief about the failure rate.

9.5.2.4 Test Coverage Reliability Metrics

Test coverage advocates have defined software reliability as a function of the amount of the software product that has been successfully verified or tested. Three such metrics are discussed below. The first is a simple ratio based upon the rate of successful testing during the final acceptance test. The second and third provide a metric based on ways of combining the results from both white-box and black-box testing.

Advocates of this approach to reliability explain that since the data are (or should be) collected and tracked during testing, these measures are readily available and require no additional verification effort. However, to the reliability engineer, these metrics are foreign to anything used in the hardware environment to describe reliability. Further, none of these metrics can be converted to failure rate or used to predict or estimate mean time between failures.

Test Success Reliability Metric

In this approach, (Ref. [19]) reliability is simply defined as the ratio of the number of test cases executed successfully during acceptance (black-box) testing, defined as s, to the total number of test cases executed during acceptance testing, defined as r:

$$R = s/r$$

The validity of the result is dependent on the size of r as well as the ability for r to represent the total operational profile of the software. During the very late stages of testing, immediately prior to delivery, this model may be used for accepting or rejecting the software.

IEEE Test Coverage Reliability Metric

This method (Refs. [21] and [22]) assumes that *reliability is dependent upon both the functions that are tested (black-box) and the product that is tested (white-box).* It assumes that both types of testing have to be completed for the test coverage to be complete. The reliability value is defined as the product of two proportions, converted to a percent:

$$R = p(\text{functions tested}) * p(\text{program tested}) * 100\%$$

where:

p(functions tested) = Number of capabilities tested/total number of capabilities

p(program tested) = Total paths and inputs tested/total number of paths and inputs

Leone's Test Coverage Reliability Metric

This approach (Ref. [23]) is similar to the IEEE Model except that it *assumes that it is possible to have either white or black box testing and still have some level of reliability*. Two white-box variables, a and b, and two black-box variables, c and d, are assessed. The reliability is the weighted sum of the four proportions:

$$R = ((a * w1) + (b * w2) + (c * w3) + (d * w4)) / (w1 + w2 + w3 + w4)$$

where:

a = Number of independent paths tested/total number of paths
b = Number of inputs tested/total number of inputs
c = Number of functions verified/total number of functions
d = Number of failure modes addressed/total number of failure modes

The values for w1, w2, w3, w4 represent weights. If all parameters are equally important, these weights all are set to 1; however, if there are data to support that some parameters are more important than others, then these more important parameters would receive higher weights.
This model has two underlying assumptions. First, independent paths are identified using information from testing procedures. Second, failure models (Ref. [24]) are identified using Fault Tree Analysis or Failure Modes, Effects, and Criticality Analysis.

9.5.3 Estimating Total Number of Faults Using Tagging

Tagging (Ref. [23]) is used to estimate the total number of faults in the software, N, based on the number observed during testing. It is based on *seeding*, a method of introducing faults into the software and then determining how many of these faults are found during testing in order to estimate the total number of faults.

To illustrate, suppose the number of fish in a pond, N, is of interest. One way to develop an estimate is to capture and tag some number of the fish, T, and return them to the pond. As fish are caught, the number of tagged fish, t, is recorded as well as the number untagged, u. The total number of untagged fish U, is estimated using the proportion: u/U = t/T. Then, the total number of fish is estimated as the sum: N = U + T.

The steps used in the basic seeding approach for a software fault estimation are:

(1) A set of faults which represents the faults which would typically be found in operational usage is identified.

(2) Faults are injected into software without the testers or developers being aware of them. The total number of injected faults is T.

(3) Test software and identify all faults found. Let t = the number of faults detected that were injected and let u = the number of faults detected which were not injected.

(4) The total number of faults which are not injected is estimated by U, where:

$$u/U = t/T$$

(5) The total number of faults, N, is estimated by:

$$N = U + T$$

(6) The injected faults are removed.

In general, this approach is not recommended based upon the issues identified below:

(1) How can faults which are typical of operational usage be identified and then injected in a completely random manner? (without bias)?
(2) Seeding assumes faults are due to coding mistakes. What about faults due to requirements, design and maintenance errors?
(3) During the course of a typical testing cycle, faults are typically corrected. Do the injected faults get corrected during this process, or at the very end.
(4) Will seeded faults prevent real faults from being detected?
(5) How can injected faults be kept a secret when the maintainer goes to fix them? How can it be justified to spend resources fixing injected faults?
(6) What is the action at the end of testing, go back to the original version with no injected faults (and no corrected real faults) or remove (hopefully all) the injected faults?

An alternative *Dual Test Group Approach* is similar to basic seeding except that two groups are used. It assumes that:

(1) Two independent test groups are testing the same software at same time.
(2) Groups do not share information on faults detected in testing.
(3) Groups create their own test plans, but test the same functionality of the software.
(4) Groups are equal in experience and capabilities.

This model predicts N, the total number of faults, based upon three numbers, n_1, n_2, and n_{12} where n_1 and n_2 represent the number of faults found by Group 1 and Group 2, respectively, while n_{12} is the number of faults found by both groups.

The total number of faults, N, is estimated by:

$$N = R + n_1 + n_2 + n_{12}$$

where:
R is the estimated total number of remaining faults

This model assumes that as the number of faults found by both groups increases, the number remaining decreases. As testing continues, it is assumed that d_{12} will increase. This means that when there are few faults left in the software (i.e., as R approaches 0), both test groups will begin finding the same faults. This may not be the case, however, since both test groups may be inefficient. The basic assumptions also may be unrealistic. It is not always possible or economical to have two completely independent test groups with equal experience level and capabilities. It also may not be easy to keep the groups independent and equal in experience.

9.6 Software Reliability Allocation

Software reliability allocation involves the establishment of reliability goals for individual computer software configuration items (CSCI) based on top-level reliability requirements for all the software. It is very important that this activity, allocations, be established early in the program so that criteria for evaluating the achieved reliability of each element can be established. Table 9.6-1 describes five allocation techniques. These techniques are based on the type of execution expected or the operational profile or the software complexity.

The allocation of a system requirement to software elements makes sense only at the software system or CSCI level. Once software CSCIs have been allocated reliability requirements, a different approach is needed to allocate the software CSCI requirements to lower levels. The reliability model for software differs significantly from hardware due to its inherent operating characteristics. For each mode in a software system's (CSCI) operation, different software modules (CSCs) will be executing. Each mode will have a unique time of operation associated with it. A model should be developed for the software portion of a system to illustrate the modules which will be operating during each system mode, and indicate the duration of each system mode. An example of this type of model is shown in Table 9.6-2 for a missile system.

TABLE 9.6-1: SOFTWARE RELIABILITY ALLOCATION TECHNIQUES (REF. [2])

Technique	Procedure Name	Use Description
Sequential Execution (see 9.6.1)	Equal apportionment applied to sequential software CSCIs	Use early in the SW development process when the software components are executed sequentially
Concurrent Execution (see 9.6.2)	Equal apportionment applied to concurrent software CSCIs	Use early in the SW development process and the software components are executed concurrently
Operational Profile (see Ref. [2])	Mission or Operational Profile Allocation	Use when the operational profile of the CSCIs are known
Operational Criticality (see 9.6.3)	Allocation based on operational criticality factors	Use when the operational criticality characteristics of the software is known
Complexity (see 9.6.4)	Allocation based on complexity factors	Use when the complexity factors of the software components are known

The software reliability model will include the number of source lines of code (SLOC) expected for each module. These data, along with other information pertaining to software development resources (personnel, computing facilities, test facilities, etc.) are used to establish initial failure intensity predictions for the software modules.

To assist in the proper selection of an allocation technique, a flow diagram is provided in Figure 9.6-1.

TABLE 9.6-2: SOFTWARE FUNCTIONS BY SYSTEM MODE - EXAMPLE

System Mode	Modules	SLOC
Standby - 2 Hours	Built-in Test (BIT)	4000
	1760 Interface	750
	Flight Sequencing	2000
	Prelaunch Initialization	900
	TOTAL	**7650**
Prelaunch - 20 Minutes	BIT	4000
	Navigation	1000
	Flight Sequencing	2000
	Prelaunch Initialization	900
	Navigation Kalman Filter	2000
	TOTAL	**9900**
Post-Launch - 10 Minutes	BIT	4000
	Interface	7000
	Navigation	1000
	Infrared Seeker Control	500
	Flight Sequencing	2000
	Terminal Maneuver	1000
	Other Post-Launch	24500
	Navigation Kalman Filter	2000
	TOTAL	**42000**

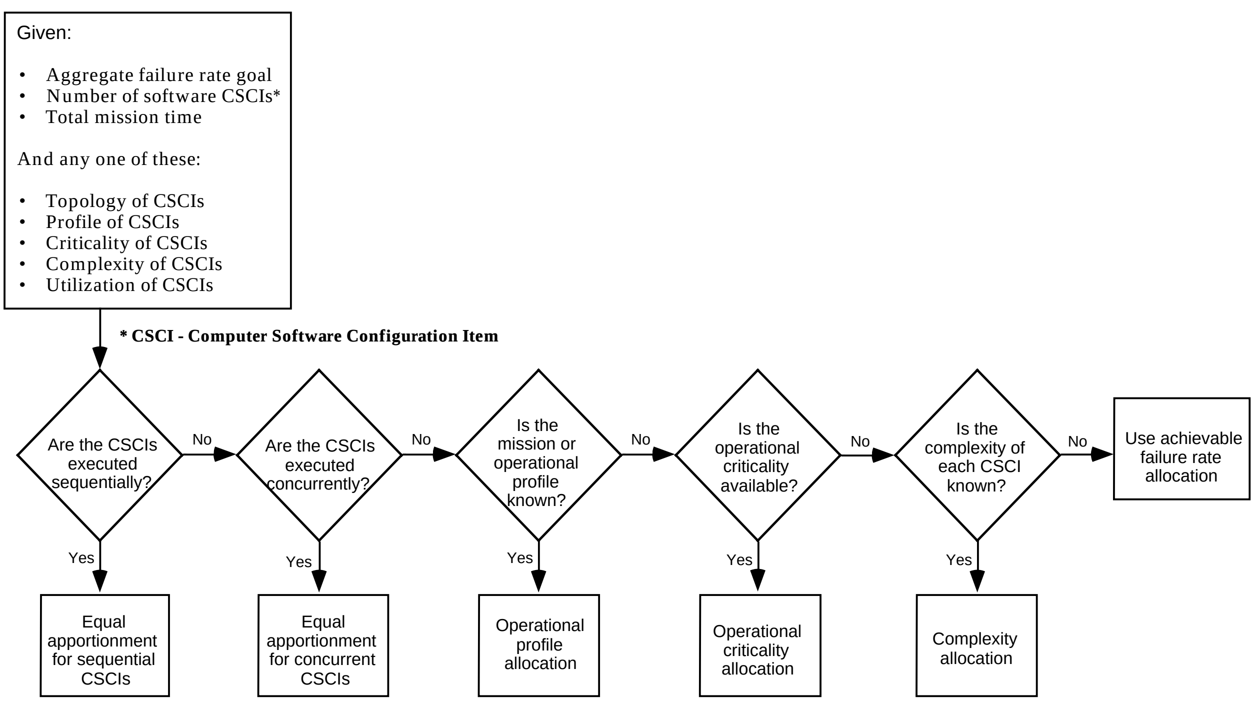


FIGURE 9.6-1: RELIABILITY ALLOCATION PROCESS (REF. [2])

9.6.1 Equal Apportionment Applied to Sequential Software CSCIs

This technique is used to allocate a failure rate goal to each individual software CSCI when the CSCIs are executed sequentially. This procedure should be used only when the failure rate goal of the software aggregate (λ_s), and the number of software CSCIs in the aggregate (N), are known. The aggregate's failure rate goal is either specified in the requirements or is the result of an allocation performed at a higher level in the system hierarchy.

Steps:

(1) Determine the failure rate goal for the software aggregate; λ_s

(2) Determine the number of software CSCIs in the aggregate; N

(3) For each software CSCI, assign the failure rate goal as follows:

$$\lambda_{i\,(CSCI)} = \lambda_s \text{ (failures per hour)}$$

where:
i = 1, 2, ..., N

Example:

A software aggregate is required to have a maximum of 0.05 failures per hour. The aggregate consists of five software CSCIs that are executed one after another, that is, the five CSCIs run sequentially. All CSCIs must succeed for the system to succeed (this is a series system).

Then, using the equal apportionment technique, the failure rate goal for the i^{th} software CSCI is assigned to be:

$$\lambda_i = \lambda_s = 0.05 \text{ failures per hours}$$

where:
i = 1, 2, ..., 5

9.6.2 Equal Apportionment Applied to Concurrent Software CSCIs

This technique is used to allocate the appropriate failure rate goal to each individual software CSCI, when the CSCIs are executed concurrently. λ_s, the failure rate of the software aggregate, and N, the number of software CSCIs in the aggregate, are needed for this procedure.

Steps:

(1) Determine the failure rate goal for the software aggregate; λ_s

(2) Determine the number of software CSCIs in the aggregate; N

(3) For each software CSCI, assign the failure rate goal as follows:

$$\lambda_{i\ (CSCI)} = \lambda_s / N \quad \text{(failures per hour)}$$

where:
i = 1, 2, ..., N

Example:

A software aggregate has a failure rate goal of 0.05 failures per hour. The aggregate consists of five software CSCIs, which are in series and executed concurrently. Then, the allocated failure rate goal of each of the five software CSCI is:

$$\lambda_i = \lambda_s/N = \frac{0.05}{5} = 0.01 \text{ failures per hour}$$

9.6.3 Allocation Based on Operational Criticality Factors

The operational criticality factors method allocates failure rates based on the system impact of a software failure. Criticality is a measure of the system's ability to continue to operate and the system's ability to be fail-safe. For certain modes of operation, the criticality of that mode may call for a lower failure rate to be allocated. In order to meet very low failure rates, fault-tolerance or other methods may be needed.

The following procedure is used to allocate the appropriate value to the failure rate of each software CSCI in an aggregate, provided that the criticality factor of each CSCI is known. A CSCI's criticality refers to the degree to which the reliability and/or safety of the system as a whole is dependent on the proper functioning of the CSCI. Furthermore, gradations of safety hazards translate into gradations of criticality. The greater the criticality, the lower the failure rate that should be allocated.

Steps:

(1) Determine the failure rate goal of the software aggregate; λ_s

(2) Determine the number of software CSCIs in the aggregate; N

(3) For each i^{th} CSCI, i = 1, 2, ..., N, determine its criticality factor c_i. The lower the c_i the more critical the CSCI.

(4) Determine τ_i' the total active time of the i^{th} CSCI, i = 1, 2, ..., N. Determine T, the mission time of the aggregate.

(5) Compute the failure rate adjustment factor K:

$$K = \frac{\sum_{i=1}^{N} c_i \tau_i'}{T}$$

(6) Compute the allocated failure rate goal of each CSCI

$$\lambda_i = \lambda_s (c_i/K)$$

(Dividing by K makes the allocated CSCI failure rates build up to the aggregate failure rate goal).

<u>Example</u>:

Suppose a software aggregate consisting of three software CSCIs is to be developed. Assume the failure rate goal of the aggregate is 0.002 failures per hour. Suppose that the mission time is 4 hours. Furthermore, the criticality factors and the total active time of the software CSCIs are:

$c_1 = 4$ $\quad \tau'_1 = 2$ hours
$c_2 = 2$ $\quad \tau'_2 = 1$ hour
$c_3 = 1$ $\quad \tau'_3 = 2$ hours

(Note: In this example, since c_3 has the smallest value, this indicates that the third CSCI of this software aggregate is the most critical.)

Compute the adjustment factor K:

$$K = \frac{c_1 \tau_1 + c_2 \tau_2 + c_3 \tau_3}{T} = \frac{(4)(2) + (2)(1) + (1)(2)}{4} = 3$$

Then, the allocated failure rate goals of the software CSCIs are:

$\lambda_1 = \lambda_s (c_1/K)$
= 0.002 (4/3) = 0.0027 failures per hour

$\lambda_2 = \lambda_s (c_2/K)$
= 0.002 (2/3) = 0.0013 failures per hour

$\lambda_3 = \lambda_s (c_3/K)$
= 0.002 (1/3) = 0.00067 failures per hour

9.6.4 Allocation Based on Complexity Factors

The following technique is used to allocate a failure rate goal to each software CSCI in an aggregate, based on the complexity of the CSCIs. There are several types of complexity as applied to software that are listed in Table 9.6-3.

TABLE 9.6-3: COMPLEXITY PROCEDURES

Complexity Type	Description	When it Can Be Used
McCabe's Complexity	A measure of the branches in logic in a unit of code.	From the start of detailed design on.
Functional Complexity	A measure of the number of cohesive functions performed by the unit.	From the start of detailed design on.
Software Product Research Function Points	A measure of problem, code, and data complexity, inputs, outputs, inquiries, data files and interfaces.	From detailed design on.
Software Product Research Feature Points	A measure of algorithms, inputs, outputs, inquiries, data files and interfaces.	From detailed design on.

During the design phase, an estimated complexity factor using any one of these techniques is available. The greater the complexity, the more effort required to achieve a particular failure rate goal. Thus, CSCIs with higher complexity should be assigned higher failure rate goals.

The complexity measure chosen must be transformed into a measure that is linearly proportional to failure rate. If the complexity factor doubles, for example, the failure rate goal should be twice as high.

Steps:

(1) Determine the failure rate goal of the software aggregate; λ_s
(2) Determine the number of software CSCIs in the aggregate; N
(3) For each CSCI, i = 1, 2, ..., N, determine its complexity factor; w_i
(4) Determine the total active time of each CSCI, i = 1, 2, ..., N; τ_i
(5) Determine the mission time of the aggregates; T
(6) Compute the failure rate adjustment factor K:

$$K = \frac{\sum_{i=1}^{N} w_i \, \tau_i}{T}$$

(7) Compute the allocated failure rate of the i[th] CSCI:

$$\lambda_i = \lambda_s \, (w_i / K)$$

<u>Example</u>:

A software aggregate consisting of 4 software CSCI is to be developed. The failure rate goal of the aggregate is 0.006 failures per hour. The mission time is three hours. Furthermore, the complexity factors and the total active time of the software CSCIs are given as:

$$w_1 = 4, \quad \tau_1 = 2 \text{ hours}$$
$$w_2 = 2, \quad \tau_2 = 1 \text{ hour}$$
$$w_3 = 3, \quad \tau_3 = 3 \text{ hours}$$
$$w_4 = 1, \quad \tau_4 = 2 \text{ hours}$$

Compute the failure rate adjustment factor K:

$$K = \frac{\sum_{i=1}^{N} w_i \, \tau_i}{T} = \frac{(4)(2) + (2)(1) + (3)(3) + (1)(2)}{3} = 7$$

Then, the failure rate goal of each software CSCIs is:

$\lambda_1 = \lambda_s (w_1/K)$
$= 0.006 (4/7) = 0.0034$ failures per hour

$\lambda_2 = \lambda_s (w_2/K)$
$= 0.006 (2/7) = 0.0017$ failures per hour

$\lambda_3 = \lambda_s (w_3/K)$
$= 0.006 (3/7) = 0.00026$ failures per hour

$\lambda_4 = \lambda_s (w_4/K)$
$= 0.006 (1/7) = 0.0009$ failures per hour

9.7 Software Testing

Most software experts recommend that an independent organization test a software system. One option is to contract with an outside organization for the testing. If this is not possible, the testing organization should be managerially separate from the design and development groups assigned to the project.

This recommendation is based more on observations of human nature than on substantiated fact. Effective testing groups need to have somewhat of a "destructive" view of a system, so that they can flush out errors and "break" the system. The design and development groups who have built the software system have a "constructive" view, and may therefore find it too difficult to develop the frame of mind required for testing.

9.7.1 Module Testing

Module testing (also called unit or component testing) is the testing of one individual component (that is, one program module, one functional unit, or one subroutine). The objective of module testing is to determine if the module functions according to its specifications.

Module testing is usually conducted by the programmer of the module being tested. It is closely tied to the programmer's development of the code and often becomes an iterative process of testing a component, finding a problem, debugging (finding the reason for the problem in the code), fixing the problem, and then testing again. Module testing is therefore often considered part of the implementation rather than part of the testing phase. Module testing should nevertheless be recognized as a separate function, and should be disciplined. The tester must develop a test plan for the component and must document test cases and procedures. Too often, this discipline is overlooked and testing of individual components becomes "ad hoc" testing with no records about the actual cases, the procedures, or the results.

White box testing is frequently used during module testing. White box testing means that the tester is familiar with the internal logic of the component and develops test cases accordingly.

Code coverage (how much of the code is covered by the testing) and logic path coverage (how many of the logical paths in the code are tested) are two primary considerations when developing test cases for module testing.

9.7.2 Integration Testing

After module testing, the next step in the software testing phase is integration testing. This activity involves combining components in an orderly progression until the entire system has been built. The emphasis of integration testing is on the interaction of the different components and the interfaces between them.

Most often, the programming group performs software integration testing. As with module testing, integration testing is very closely linked to the programming activity since the tester needs to know details of the function of each component to develop a good integration test plan.

Integration Test Techniques. An important decision when planning for integration testing is determining the procedure to be used for combining all the individual modules. There are two basic approaches for doing this: non-incremental testing and incremental testing.

In non-incremental integration testing, all the software components (assuming they have each been individually module tested) are combined at once and then testing begins. Since all modules are combined at once, a failure could be in any one of the numerous interfaces that have been introduced.

The recommended approach for the integration of system components is planned incremental testing. With this method, one component is completely module tested and debugged. Another component is then added to the first and the combination is tested and debugged. This pattern of adding one new component at a time is repeated until all components have been added to the test and the system is completely integrated.

Incremental testing requires another decision about the order in which the components will be added to the test. There are no clear-cut rules for doing this. Testers must base a decision on their knowledge of what makes the most sense for their system, considering logic and use of resources. There are two basic strategies: top-down or bottom-up as shown in Figure 9.7-1.

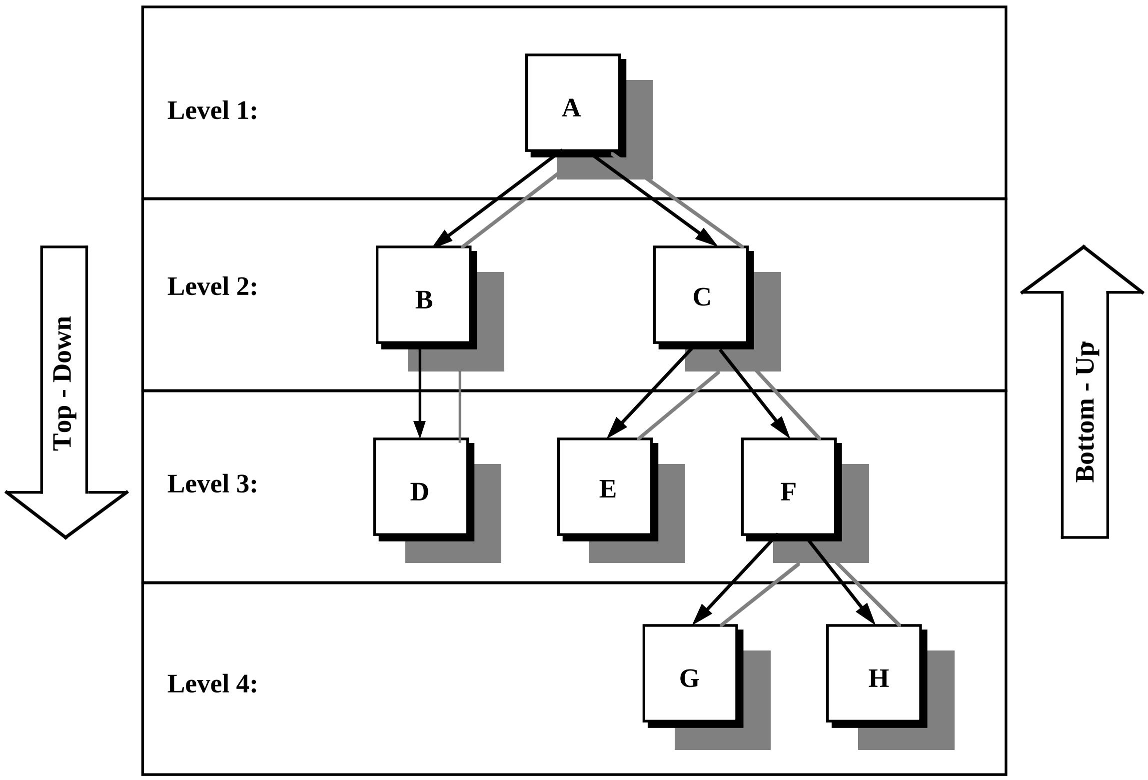


FIGURE 9.7-1: STRUCTURAL REPRESENTATION OF A SOFTWARE SYSTEM

A tester using top-down integration testing on this system begins by module testing and debugging the A component. The next step is to add a new component to the test. In this case, either B or C is added. If B was chosen and tested, either C or D could be the next choice. Some testers prefer to follow one path to completion, while others prefer to complete all the modules on the same level before proceeding to a lower level of the hierarchy.

Bottom-up integration testing reverses top-down testing. With this approach, a tester simply starts at the bottom-most level of the hierarchy and works up. As shown in Figure 9-16, a tester might start by module testing component G. With bottom-up testing, all the components at the bottom of the hierarchy are usually module tested first and then testing proceeds in turn to each of their calling components. The primary rule in bottom-up testing is that a component should not be chosen to be the next one added to the test unless all of the components that it calls have already been tested.

9.7.3 System Testing

System Testing Techniques. System testing is often referred to as “testing the whole system.” Translated literally, that could mean that every input or output condition in the software needs to be tested for every possible logical path through the code. Even in a small system this task could become quite lengthy. In a large, complex system, it would be prohibitively time-consuming and expensive.

The system test organization must develop a strategy for testing a particular system and determine the amount of test coverage required. There is no cookbook for doing so. In a small noncritical system, a very low degree of test coverage may be acceptable. High coverage is needed in a critical software system involving human life. The testers must decide the best plan based on system characteristics, the environment in which the software system will operate, and the testers’ experience.

In general, software system testing is done using black box testing. The tester, viewing the system as a black box, is not concerned with the internals, but rather is interested in finding if and when the system does not behave according to its requirements.

One technique often used for identifying specific test cases is called equivalence partitioning. In this method, an equivalence class is identified so that one test case covers a number of other possible test cases.

Boundary analysis is another technique used in which testing is performed on all the boundary conditions. This method tests the upper and lower boundaries of the program. In addition, it is usually wise to test around the boundaries.

A third technique that should always be applied to the testing of a program is called error guessing. With this method, testers use their intuition and experience to develop specific test cases. A good system tester is usually very effective at doing this.

9.7.4 General Methodology for Software Failure Data Analysis

A step-by-step procedure for software failure data analysis is shown in Figure 9.7-2 and described below:

Step 1: Study the failure data

The models previously described assume that the failure data represent the data collected after the system has been integrated and the number of failures per unit time is statistically decreasing. If, however, this is not the case, these models may not yield satisfactory results. Furthermore, adequate amount of data must be available to get a satisfactory model. A rule of thumb would be to have at least thirty data points.

Step 2: Obtain estimates of parameters of the model

Different methods are generally required depending upon the type of available data. The most commonly used ones are the least squares and maximum likelihood methods.

Step 3: Obtain the fitted model

The fitted model is obtained by first substituting the estimated values of the parameters in the postulated model. At this stage, we have a fitted model based on the available failure data.

Step 4: Perform goodness-of-fit test

Before proceeding further, it is advisable to conduct the Kolmogorov-Smirnov goodness-of-fit test or some other suitable test to check the model fit.

If the model fits, we can move ahead. However, if the model does not fit, we have to collect additional data or seek a better, more appropriate model. There is no easy answer to either how much data to collect or how to look for a better model. Decisions on these issues are very much problem dependent.

Step 5: Computer confidence regions

It is generally desirable to obtain 80%, 90%, 95%, and 99% joint confidence regions for the parameters of the model to assess the uncertainty associated with their estimation.

Step 6: Obtain performance measure

At this stage, we can compute various quantitative measures to assess the performance of the software system. Confidence bounds can also be obtained for these measures to evaluate the degree of uncertainty in the computed values.

9.8 Software Analyses

Two types of analyses will be discussed in this section; the failure modes and effects analysis (FMEA) and the fault tree analysis (FTA). The objective of both analyses is to determine what the system or product software may do or not do that is not desirable. This is opposite of most analyses which attempt to show that the product performs the intended functions. Safety criticality is one area for detailed software analysis.

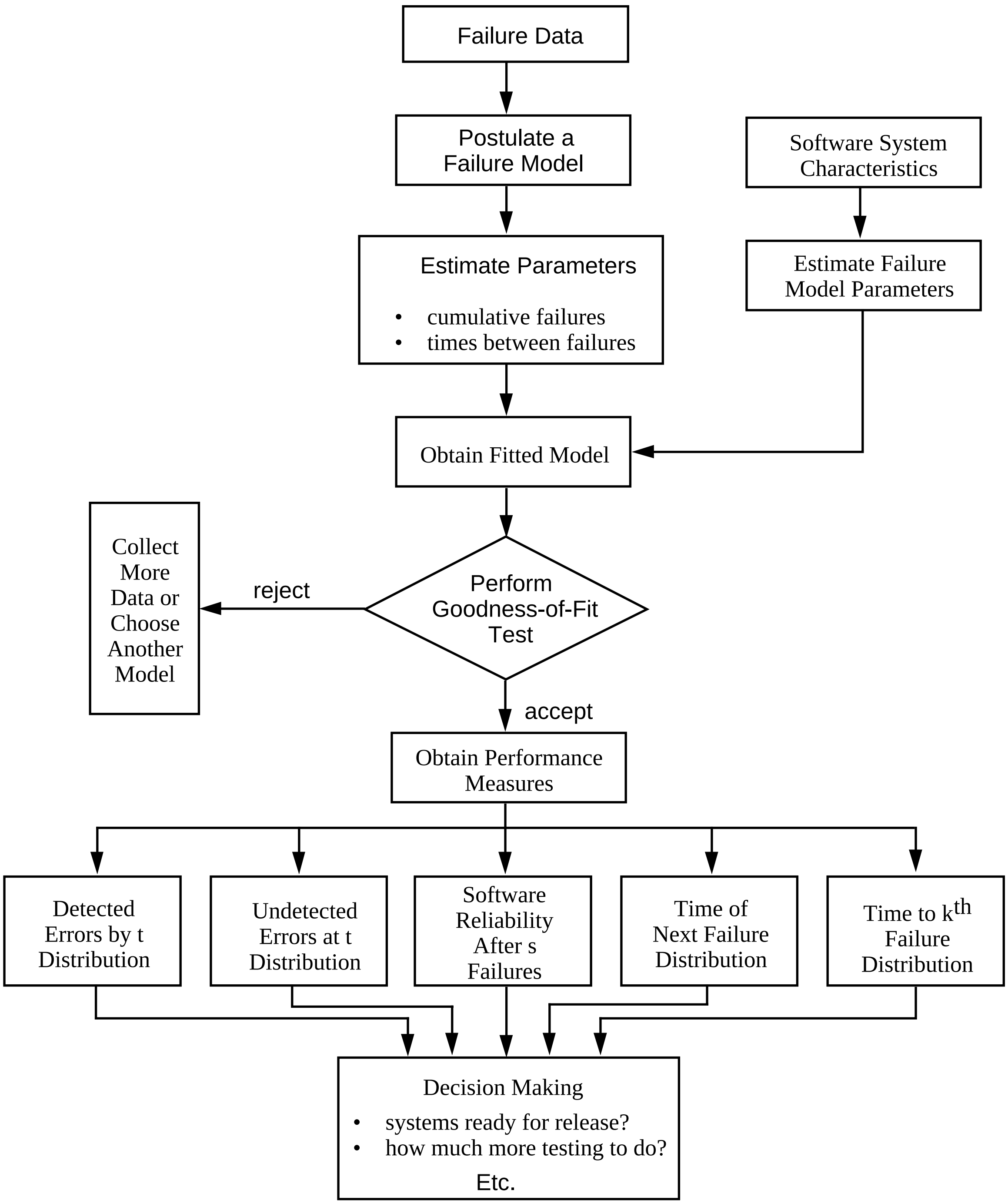


FIGURE 9.7-2: FLOWCHART FOR SOFTWARE FAILURE DATA ANALYSIS AND DECISION-MAKING

9.8.1 Failure Modes

The definition of what constitutes a failure in software is often open to debate. When a program "*crashes*", it has obviously failed due to an error, either in design, coding, testing, or exception handling.

Software sometimes fails to perform as desired. These failures may be due to errors, ambiguities, oversights or misinterpretation of the specification that the software is supposed to satisfy, carelessness or incompetence in writing code, inadequate testing, incorrect or unexpected usage of the software or other unforeseen problems. (Ref. [25]).

However, the software may not crash and still fail. This can be due to a number of criteria which are not always well defined before development. Speed of execution, accuracy of the calculations and other criteria can all be significant factors when identifying lack of successful software operation. In addition, each of these criteria has a specific level of importance and is assessed on a specific scale.

When first using or evaluating a software program, a significant amount of time can be spent determining compliance with specification or applicability of an application. Depending on the person doing the evaluation, or the evaluation process scope and design, the evaluation may or may not fully exercise the program or accurately identify functionality that is unacceptable.

Hardware/software interface problems can also occur, including failures in software due to hardware or communications environment modifications. Software errors can be introduced during software functional upgrades or during either scheduled or unscheduled maintenance.

Software failures are usually considered relative to the application type and the severity of failure as evaluated by the specific end user. Consider the following two examples. One software program contains complex graphical user interfaces that map exactly to the customer's layout requirements; but this program crashes whenever a specific sequence of user inputs and events occurs. Another software program has layout flaws but it does not fail for any sequence of user triggered events. Which program is more reliable?

(1) Is an application reliable if it meets all specified requirements? Then the first is better.

(2) If failure is defined as any crash, then the second is more reliable; in fact, some would say it is perfectly reliable because it does not crash.

9.8.2 Failure Effects

When software fails, a dramatic effect, such as a plane crash, can also be observed. Often, however, the effect of a software failure is not immediately seen or may only cause inconvenience. A supermarket checkout system which incorrectly prices selected items may never be noticed, but a failure has still occurred. An Automatic Teller Machine (ATM) which does not allow user access is a nuisance which results in disgruntled customers. Both of these may be the result of catastrophic software failures, but, in reference to endangering human life, both are minor system failures.

These examples illustrate that is it important to distinguish between the software failure relative to the software's functioning as compared to the software failure relative to the total system's functioning. In the supermarket example, the software may have failed but the checkout continued while in the ATM example, the system did not operate.

9.8.3 Failure Criticality

Both hardware and software fall into two general categories based on the function performed: *mission critical* and *non-mission critical*. *Mission critical* encompasses all failures that are life threatening as well as failures that have catastrophic consequences to society. Table 9.8-1 identifies hardware failure severity levels with respect to both mission and operator. In hardware reliability improvement, usually only catastrophic and critical levels of severity are addressed.

TABLE 9.8-1: HARDWARE FAILURE SEVERITY LEVELS (REF. [26])

Term	Definition
Catastrophic	A failure which may cause death or system loss (i.e., aircraft, tank, missile, ship, etc.).
Critical	A failure which may cause severe injury, major property damage, or major system damage which will result in mission loss.
Marginal	A failure which may cause minor injury, minor property damage, or minor system damage which will result in delay or loss of availability or mission degradation.
Minor (Negligible)	A failure not serious enough to cause injury, property damage, or system damage, but which will result in unscheduled maintenance or repair.

No similar set of criticality classifications has been adopted by the entire software community. Putnam and Myers have defined four classes of software defect severity and identify the corresponding user response as shown in Table 9.8-2. It is interesting to note that this classification is not with respect to operator or mission but views the software as an entity in itself. No application reference is included in the descriptions. Another interesting contrast is that any level of software defect can cause a catastrophic system failure. If the software crashes ("*Critical*"), mis-computes ("*Serious*"), provides a partly correct answer ("*Moderate*") or mis-displays the answer on the screen ("*Cosmetic*"), the resultant failure may be catastrophic, resulting in system and/or operator loss.

TABLE 9.8-2: SOFTWARE FAILURE SEVERITY LEVELS (REF. [5])

Severity	Description	User Response
Critical	Prevents further execution; nonrecoverable.	Must be fixed before program is used again.
Serious	Subsequent answers grossly wrong or performance substantially degraded.	User could continue operating only if allowance is made for the poor results the defect is causing. Should be fixed soon.
Moderate	Execution continues, but behavior only partially correct.	Should be fixed in this release.
Cosmetic	Tolerable or deferrable, such as errors in format of displays or printouts.	Should be fixed for appearance reasons, but fix may be delayed until convenient.

9.8.4 Fault Tree Analysis

Fault tree analysis is performed on software to determine the areas in the product which could cause a potential failure and to determine the risk and severity of any such potential failure. The timing of this analysis is important and should start during the design phase to identify top-level hazards. The analysis can continue through code development and testing to identify paths for testing and verify that safety related hazards will not occur.

The steps for performing a software fault tree are:

(1) Determine failure modes for software starting from top level product and working downward.

(2) Make these failure modes the top nodes of the fault tree. Assume that these failure modes have already occurred and refer to them as events.

(3) When tree completed for top level failure modes, determine risk and severity for each of the bottom nodes on the tree.

Risk (Ref. [27])

1	Remote possibility of happening
2-3	Low probability with similar designs
4-6	Moderate probability with similar designs
7-9	Frequent probability with similar designs
10-	High probability with similar design

Severity (Ref. [27])

1-2	Probably not detected by customer
3-5	Result in slight customer annoyance
6-7	Results in customer dissatisfaction
8-9	Results in high customer dissatisfaction
10-	Results in major customer dissatisfaction, loss of system operation, or non-compliance with government regulations

(4) Using design flows and charts, determine how the failure modes identified in step 1 can occur. Assume that the failure mode has already occurred and identify what causes it. Bottom nodes of tree will be more specific failure modes.

(5) When the tree is completed for the next level of design, identify failure modes associated with this level and identify risk and severity.

(6) Repeat steps 4 and 5 until a satisfactory level of abstraction has been reached (normally determined by customer).

(7) The tree is pruned when risk and severity are insignificant or when the lowest level of abstraction is reached. The tree is expanded when risk and probability are significant.

9.8.5 Failure Modes and Effects Analysis

In contrast to the fault tree top down development, the failure modes and effects analysis is a bottom up approach. That is, a failure mode is selected in a lower level unit and the failure effect through the system is evaluated. The units that will be affected and the probability and criticality of the failure mode is determined from the failure rates, operating time, and criticality ranking.

The steps for applying failure modes and effects analysis to software are:

(1) Determine product level failure modes using step 1 of the Fault Tree Analysis section.
(2) Using a software failure mode chart, work through top level of chart using the top level failure modes and fill in the form. One unit may have several failure modes or no failure modes.
(3) Repeat steps 1 and 2 for the next level of design until lowest level is reached.

Table 9.8-3 lists the categories that must be addressed when performing a complete failure mode and criticality analyses.

An example of a software failure modes and effects analysis is shown in Figure 9.8-1. Each function failure mode and end effect are described.

TABLE 9.8-3: SOFTWARE FAILURE MODES AND CRITICALITY ANALYSIS CATEGORIES

Software FMECA Categories

(1) Unit - Name of software unit at CSCI, CSC or unit level
(2) Function - General function performed by unit
(3) Failure mode - the associated failure mode
(4) The probable cause of the failure in software terms
(5) The effect on the unit, the next level and the top level. Define these effects in terms of processing, output, etc.
(6) Interrupted? If service/mission would be interrupted by this failure mode state so.
(7) Crit - Criticality I - catastrophic, II - critical, III - moderate, IV - negligible
(8) Predictability - If there is some predictability before the failure occurs state so. Normally software failure have no predictability so this will probably always be no
(9) Action - the type of corrective action required. This will either be restart if the problem can be circumvented, or remote corrective action if it can only be fixed in an engineering environment.

No.	Unit	Function	Failure Mode	Probable Cause	Effect On			Interrupt ?	Crit	Action
					Unit	Sub	System			
1	Output	Outputs file into	Output is incorrect	Inputs are invalid and not detected	n/a	none	mission degraded	no	II	lab repair
2	Output	Outputs file into	Output is incorrect	Inputs are correct but not stored properly	n/a	none	mission degraded	no	II	lab repair
3	Output	Outputs file into	Output is incorrect	Values are not computed to spec	n/a	none	mission degraded	no	II	lab repair

FIGURE 9.8-1: EXAMPLE OF SOFTWARE FMECA

9.9 References

1. Coppola, Anthony,*"Reliability Engineering of Electronic Equipment - A Historical Perspective*," IEEE Transactions on Reliability, April 1984, pp. 29-35.

2. Lakey, P.B., and A.M. Neufelder, *"System and Software Reliability Assurance,"* RL-TR-97-XX, Rome Laboratory, 1997.

3. Jones, Capers, "Assessment and Control of Software Risks," Yourdon, 1994.

4. Dunn, Robert, "*Software Defect Removal,"* McGraw-Hill, 1984.

5. Putnam, Lawrence H., and Myers, Ware, "Measures for Excellence: Reliable Software on Time, Within Budget," Prentice-Hill, 1992.

6. Boehm, Barry W., "*Software Engineering Economics*," Prentice-Hall, 1981.

7. Rook, Paul, Editor, "*Software Reliability Handbook*," Elsevier, 1990.

8. Dyer, Michael, "The Cleanroom Approach to Quality Software Development," Wiley, 1992.

9. Mills, Harlan D., "*Cleanroom Software Engineering,"* Center for Software Engineering, 1994.

10. Linger, Richard C., "Cleanroom Process Model," IEEE Software, March 1994, pp. 50-58.

11. Software Technology Support Center, "Guidelines for Successful Acquisition and Management of Software Intensive Systems," Dept. of the Air Force, February 1995.

12. Musa, John, A. Iannino, K. Okumoto, "*Software Reliability: Measurement, Prediction, Application*," McGraw-Hill, 1987.

13. Fairley, .R.E., "*Software Engineering Concepts*," McGraw-Hill, 1985.

14. Booch, Grady, "*Object-Oriented Development,"* IEEE Transactions of Software (February), 1986.

15. Meyer,Bertrand, "*Object-Oriented Construction,"* Prentice-Hall, 1988.

16. McCall, J.A., W. Randell, and J.Dunham, "*Software Reliability, Measurement, and Testing,"* Rome Laboratory, RL-TR-92-52, 1992.

17. Friedman, M.A., P.K. Tran and P.L. Goddard, *"Reliability Techniques for Combined* Hardware and Software Systems," RL-TR-92-15, 1992.

18. Lloyd D.K. and M. Lipow, *"Reliability: Management, Methods, and Mathematics,"* Second Edition, American Society for Quality Control, 1977.

19. Farr, W.H., *"A Survey of Software Reliability Modeling and Estimation,"* Naval Surface Weapons Center, NSWC-TR-82-171, 1983.

20. Shooman, Martin L., *"Software Engineering,"* McGraw Hill, 1983.

21. IEEE STD 982.1, "IEEE Standard Dictionary of Measures to Produce Reliable Software," 1989.

22. IEEE STD 982.2, "Guide for the Use of IEEE Standard Dictionary of Measures to Produce Reliable Software," 1989.

23. Neufelder, A., *"Ensuring Software Reliability,"* Marcel Dekker, 1993.

24. Mahar, David, *"Fault Tree Analysis,"* Reliability Analysis Center (RAC), 1990.

25. Keiller, Peter A. and Douglas R. Miller, *"On the Use and the Performance of Software Reliability Growth Models,"* Software Reliability and Safety, Elsevier, 1991.

26. MIL-STD-1629A, "Procedure for Performing Failure Mode, Effects and Criticality Analysis," Department of Defense, 1980.

27. Texas Instruments, Inc., "Components Sector Training and Organizational Effectiveness, Failure Modes and Effect Analysis," 1993.

10.0 SYSTEMS RELIABILITY ENGINEERING

10.1 Introduction

The material presented in the previous sections of this handbook in a sense set the stage for this section. This section combines the R&M theory and engineering practices previously presented into a cohesive design methodology which can be applied at the system level to optimize system "worth" for minimum life cycle costs.

The "worth" of a particular equipment/system is determined primarily by the effectiveness with which it does its job - its "operational" effectiveness. An acceptable level of effectiveness is required for every operational system.

In the final analysis, the effectiveness of a system can only be really measured when the system is performing its mission in the actual (or accurately simulated) environment for which it was designed. Of critical importance, however, is how system effectiveness can be considered while system design concepts are developed, how it can be ensured during design, and how it can be evaluated during test. Thus, most system effectiveness methodologies address these issues more than measuring system effectiveness after the system is fielded.

Table 10.1-1 represents the system effectiveness concept and the parameters that have been traditionally used (with minor variations) for system effectiveness analysis.

TABLE 10.1-1: CONCEPT OF SYSTEM EFFECTIVENESS

	System Effectiveness is the Net Result of		
	Availability	Dependability	Capability
Measures:	System condition at start of mission	System condition during performance of mission	Results of mission
Determined by:	Reliability Maintainability Human Factors Logistics	Repairability Safety Survivability Vulnerability	Range Accuracy Power Lethality etc.

As can be seen from the table, availability (how often), dependability (how long), and performance capability (how well) are the primary measures of system effectiveness:

(1) **Availability** is a measure of the degree to which an item is in the operable and committable state at the start of the mission, when the mission is called for at an unknown (random) time.

(2) **Dependability** is a measure of the degree to which an item is operable and capable of performing its required function at any (random) time during a specified mission profile, given item availability at the start of the mission. (This definition is different than the definition of dependability as it appears in International Electrotechnical Commission documents.)

(3) **Capability** is a measure of the ability of an item to achieve mission objectives, given the conditions during the mission.

System effectiveness assessment fundamentally answers three basic questions:

(1) Is the system working at the start of the mission?

(2) If the system is working at the start of the mission, will it continue to work during the mission?

(3) If the system worked throughout the mission, will it achieve mission success?

R&M are important contributions to system effectiveness since they are significant factors in consideration of the availability and dependability parameters. However, in the total system design context, as shown in Table 10.1-1, they must be integrated with other system parameters such as performance, safety, human factors, survivability/vulnerability, logistics, etc., to arrive at the optimum system configuration.

Just about all of the system effectiveness methodologies which have been developed and/or proposed in the past 20 years are concerned with this fundamental question of combining the previously mentioned parameters to achieve optimum system design. In Section 10.2, some of the more significant system effectiveness concepts and methodologies are discussed and compared.

10.1.1 Commercial-Off-The-Shelf (COTS) and Nondevelopmental Item (NDI) Considerations

Under the current military acquisition reform initiatives, the Department of Defense is advocating the use of Commercial-Off-The-Shelf (COTS) and Nondevelopmental Items (NDI) in the products it acquires for military applications. Commercial industry has long used NDI in building new products. NDI is any previously developed item used exclusively for government purposes by "federal agency, a state or local government or a foreign government with which the

US has mutual defense cooperation agreement."[1] COTS are items available in a domestic or foreign commercial marketplace. The increased emphasis on commercial products and practices has occurred for a number of reasons. First, the decrease in military spending over the last decade has resulted in an erosion in the industrial base that existed to support development of weapon systems. Second, while technology was driven primarily by the DoD in the past, this is no longer the case. Third, many technologies (e.g., electronics, information, communications) are advancing at such a rapid pace that the government can no longer afford an acquisition process that has historically required at least a 2-3 year cycle to develop, test, and field a system.

The objective of using COTS/NDI is to reduce the development time and risk associated with a new product by reducing or eliminating new design and development, thereby capitalizing on proven designs. Whether it is the government or a private commercial company, using COTS/NDI can potentially reduce costs, risks, and acquisition time. However, some compromises in the required functional performance (including reliability) of the product may be necessary, and other issues, such as logistics support, must also be considered. The decision to use COTS/NDI must be based on a thorough evaluation of its ability to perform the required function in the intended environment and to be operated and supported over the planned life of the product.

A product that is new in every aspect of its design carries with it cost, schedule, and performance risks. These risks are usually high for such a product because of all the unknowns surrounding a totally new design. A product development involving a completely new design is considered revolutionary in nature.

In contrast to a completely new design (revolutionary approach), using a proven product or incorporating proven components and subsystems in a new product is an evolutionary approach. Using COTS/NDI is a way to follow a pattern of new product development in which new design is minimized or eliminated. Some types of NDI include:

- Items available from a domestic or foreign commercial marketplace
- Items already developed and in use by the U.S. government
- Items already developed by foreign governments

COTS/NDI items may constitute the entire product (e.g., a desktop computer) or they may be components or subsystems within the product (e.g., displays, power supplies, etc., used within a control system). The advantages and disadvantages of using COTS/NDI are summarized in Table 10.1-2.

The use of commercial items in military systems is no longer a question of "yes or no" but a question of "to what degree." A pictorial presentation of the commercial/ NDI decision process is shown in Figure 10.1-1 taken from SD-2. The R&M activities needed for COTS/NDI are different than for new development items, as shown in Table 10.1-3. These considerations are

[1] SD-2, Buying Commercial and Nondevelopment Item: A Handbook, Office of the Assistant Secretary of Defense for Production and Logistics, April 1996.

discussed in more detail in the following paragraphs.

For new development programs, the customer imposes reliability requirements in the system specification and development specifications. (In addition, prior to Acquisition Reform, the customer stipulated in the statement of work which tasks the contractor would conduct as part of the reliability program and how (by imposing standards) the tasks were to be conducted).

With commercial items and NDI, the basic product is already designed and its reliability established. Consequently, the reliability assessment should be an operational assessment of the military application in the expected military environments. Since the basic design of a commercial or nondevelopmental item cannot be controlled by the buyer, the objective is to determine whether well-established and sound reliability practices were applied during the item's development.

When considering the use of COTS/NDI equipment, much work needs to be done up front in terms of market research and development of minimum requirements. This means that procurement offices must work closely with the end user to define the minimum acceptable performance specifications for R&M. Market research then needs to be performed to see what COTS/NDI equipment exists that has the potential of meeting defined requirements at an affordable price.

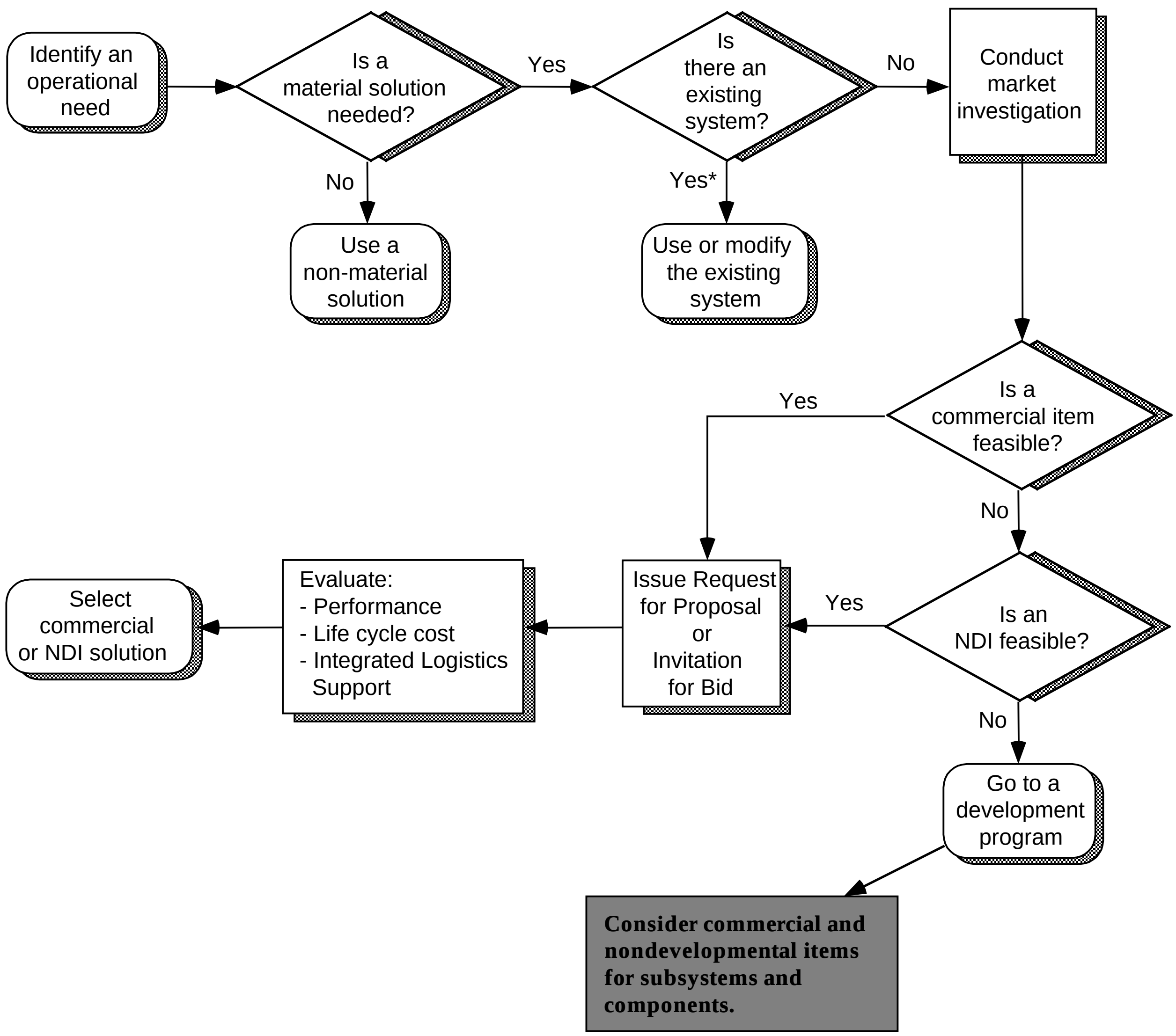
The challenge for market research is obtaining R&M data on COTS/NDI equipment. COTS vendors may not have the kinds of data that exist in military R&M data collection systems. (Text continues after Tables 10.1-1 and 10.1-2 and Figure 10.1-1).

TABLE 10.1-2: ADVANTAGES AND DISADVANTAGES OF COTS/NDI

AREA OF COMPARISON	ADVANTAGES	DISADVANTAGES
Technical, Schedule, and Financial Risk	Decreased technical, financial, and schedule risks due to less new design of components and subsystems. Ideally no research and development costs are incurred.	When NDI items are used as the components and subsystems of a product, integration of those items into the product can be difficult, expensive, and time-consuming.
Performance	There is increased confidence due to established product performance and the use of proven components and subsystems.	Performance trade-offs may be needed to gain the advantages of NDI. Integration may be difficult.
Environmental Suitability	In similar applications, proven ability to operate under environmental conditions.	In new applications, may require modifications external or internal to the equipment to operate.
Leverage	Ability to capitalize on economies of scale, state-of-the-art technology, and products with established quality.	There may not be a perfect match between requirements and available products.
Responsiveness	Quick response to an operational need is possible because new development is eliminated or minimized.	Integration problems may reduce the time saved.
Manufacturing	If already in production, processes are probably established and proven.	Configuration or process may be changed with no advance notice.
Resupply	There is no need for (large) inventory of spares because they can be ordered from supplier.	The long-term availability of the item(s), particularly COTS, may be questionable.
Logistics Support	No organic support may be required (probably not possible). Repair procedures and rates are established.	Supplier support or innovative integrated logistics support strategies may be needed to support the product.

TABLE 10.1-3: R&M ACTIVITIES FOR NEW DEVELOPMENT ITEMS AND FOR COTS

R&M ACTIVITY	TYPE OF ITEM	
	NEW DEVELOPMENT	**COTS/NDI**
Determine Feasibility	Develop requirements based on user needs and technology being used. Estimate achievable level of R&M.	Limited to verifying manufacturer claims.
Understand the Design	Perform FMEA, FTA, and other analyses for entire design. Conduct design reviews. Develop derating criteria. Conduct development testing.	Limited to integration and any modifications.
Parts Selection	Analyze design to determine correct parts application for robust design. Identify needed screening.	None.
Validate the Design	Conduct extensive development testing that addresses all aspects of the design. Identify design deficiencies and take corrective action. Establish achieved levels of R&M.	Limited to what is needed to verify manufacturer claims and to validate integration or required modifications based on the intended environment.
Manufacturing	Design manufacturing processes to retain inherent R&M. Implement statistical process control and develop good supplier relationships.	None if the item is already in production. Otherwise, design the manufacturing process to retain the inherent design characteristics.



* In preparation for the market investigation establish objectives and thresholds for cost, schedule, and performance based on the users' operational and readiness requirements.

FIGURE 10.1-1: THE COMMERCIAL/NDI DECISION PROCESS

SECTION 10: SYSTEMS RELIABILITY ENGINEERING

If design documentation is available, specific R&M tasks, such as prediction and failure modes and effects analysis, may be part of the COTS/NDI evaluation process. Because the prime military contractor is not likely to be the COTS/NDI vendor in this case, both the government and the prime will need to perform the evaluation i.e., a cooperative effort should exist between the two parties.

The amount of testing required to verify that a commercial item or NDI meets the operational requirement is governed by whether the item will be used in the environment for which it was designed and by operators with skills equal to the operators for which it was designed. What may be needed is to require the supplier to furnish operational and environmental characterization data and the results of testing to substantiate reliability and maintainability claims. Also, it may be necessary to require the supplier provide some evidence that the manufacturing processes do not compromise the designed-in reliability and maintainability characteristics. This evidence may include the results of sampling tests, control charts showing that critical processes are in control with a high process capability, and so forth.

10.1.2 COTS/NDI as the End Product

When purchasing COTS/NDI as the total product, the best course of action may be to require only data that substantiates R&M performance claims and to emphasize the role of the manufacturing processes (for NDI not yet in production) in determining the reliability and maintainability of the product. In some cases, even that data may not be needed if either the customer has already determined (through its own testing of samples, for example) that the product has the requisite performance, or if use or independent testing of the product in commercial applications has shown the product's performance to be satisfactory (for example, a personal computer in an office environment). In any case, imposing specific R&M tasks on manufacturers of COTS/NDI, even if they were willing to bid on such a procurement, is usually counterproductive and expensive.

The advantage of using COTS/NDI is that the development is complete (with only minor exceptions); the supplier has already done (or omitted) whatever might have been done to design a reliable and maintainable product. What may be need is to require the supplier to furnish operational and environmental characterization data and the results of testing to substantiate reliability and maintainability claims. Also, it may be necessary to require the supplier provide some evidence that the manufacturing processes do not compromise the designed-in reliability and maintainability characteristics. This evidence may include the results of sampling tests, control charts showing that critical processes are in control with a high process capability, and so forth.

10.1.3 COTS/NDI Integrated with Other Items

When COTS/NDI is being integrated with other items, either new development or other COTS/NDI, the same attention and level of effort that is characteristic of a new development must be given to the integration. R&M and other performance characteristics may be seriously affected by the integration due to feedback, interference and other interactions. The integration may require interface devices, which themselves may present new R&M problems. One would expect a supplier to perform Failure Modes and Effects Analysis (FMEA), Fault Tree Analysis (FTA), and other analyses to ensure that the integration does not compromise the R&M performance of any of the items being integrated and that the resulting product meets the required levels of R&M performance.

10.1.4 Related COTS/NDI Issues

Three of the most important issues associated with using COTS/NDI are the logistics support concept, availability of parts, and performance in the intended military environment. Other issues include configuration management of the COTS/NDI (important if the customer plans to support the product organically), and the availability or development of documentation to support operations and organic maintenance.

10.2 System Effectiveness Concepts

The three generally recognized components of system effectiveness previously defined (availability, dependability, capability) will be used as the basis for description and comparison of the concepts and formulations of system effectiveness. It should be recognized that all of these effectiveness components must be derived from an analysis of the operational needs and mission requirements of the system, since it is only in relation to needs and missions that these basic components can be meaningfully established.

Many semantic difficulties arise when discussing systems effectiveness and its components. These difficulties result from the fact that some people use the same words to mean different things or different words to mean the same things.

Definitions of many of the terms used in the following paragraphs were provided in Section 3 and will not be repeated here.

10.2.1 The ARINC Concept of System Effectiveness (Ref. [1])

One of the early attempts to develop concepts of system effectiveness was delineated by ARINC (Aeronautical Radio Inc.) in its book "Reliability Engineering." It contains some of the earliest published concepts of systems effectiveness and represents one of the clearest presentations of these concepts from which many of the subsequent descriptions have been derived. The definition of systems effectiveness applied in this early work is: "Systems effectiveness is the probability that the system can successfully meet an operational demand within a given time when operated under specified conditions." This definition includes the concepts that system effectiveness

(1) Can be measured as a **probability**

(2) Is related to operational **performance**

(3) Is a function of **time**

(4) Is a function of the **environment** or conditions under which it is used

(5) May vary with the **mission** to be performed

Although it is not essential to describe system effectiveness in terms of probability as opposed to other quantitative measures, it has often been found convenient to do so. The ARINC model may be expressed such that system effectiveness probability, P_{SE} is the product of three probabilities as follows:

$$P_{SE} = P_{OR} \cdot P_{MR} \cdot P_{DA} \tag{10.1}$$

where:

P_{OR} = operational readiness probability
P_{MR} = mission reliability probability
P_{DA} = design adequacy probability

This equation states that the effectiveness of the system is the product of three probabilities: (1) the probability that the system is operating satisfactorily or is ready to be placed in operation when needed; (2) the probability that the system will continue to operate satisfactorily for the period of time required for the mission; and (3) the probability that the system will successfully accomplish its mission, given that it is operating within design limits.

10.2.2 The Air Force (WSEIAC) Concept (Ref. [2])

A later definition of system effectiveness resulted from the work of the Weapon System Effectiveness Industry Advisory Committee (WSEIAC) established in late 1963 by the Air Force System Command. The WSEIAC definition of system effectiveness is: "System effectiveness is a measure of the extent to which a system may be expected to achieve a set of specific mission requirements and is a function of availability, dependability, and capability." The definition may be expressed as:

$$SE = ADC \tag{10.2}$$

where:

A = availability
D = dependability
C = capability

See definitions in Section 10.1.

These are usually expressed as probabilities as follows:

(1) "A" is the vector array of various state probabilities of the system at the beginning of the mission.

(2) "D" is a matrix of conditional probabilities over a time interval, conditional on the effective state of the mission during the previous time interval.

(3) "C" is also a delineal probability matrix representing the performance spectrum of the system, given the mission and system conditions, that is, the expected figures of merit for the system.

Basically, the model is a product of three matrices:

- Availability row vector A
- Dependability matrix D
- Capability matrix C

In the most general case, assume that a system can be in different states and at any given point in time is in either one or the other of the states. The **availability row vector** is then

$$\mathbf{A} = (a_1, a_2, a_3, \ldots, a_i, \ldots a_n) \tag{10.3}$$

where a_i is the probability that the system is in State i at a random mission beginning time. Since the system can be in only one of the n states and n is the number of all possible states it can be in (including the down states in which the system cannot start a mission), the sum of all the probabilities a_i in the row vector must be unity, i.e.,

$$\sum_{i=1}^{n} a_i = 1 \tag{10.4}$$

The **dependability matrix** D is defined as a square n • n matrix

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n} \\ \vdots & & & & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \cdots & d_{nn} \end{bmatrix} \tag{10.5}$$

where the meaning of the element d_{ij} is defined as the expected fraction of mission time during which the system will be in State j if it were in State i at the beginning of the mission. If system output is not continuous during the mission but is required only at a specific point in the mission (such as over the target area), d_{ij} is defined as the probability that the system will be in State j at the time when output is required if it were in State i at mission start.

When no repairs are possible or permissible during a mission, the system upon failure or partial failure cannot be restored to its original state during the mission and can at best remain in the State i in which it started the mission or will degrade into lower states or fail completely. In the case of no repairs during the mission, some of the matrix elements become zero. If we define State 1 as the highest state (i.e., everything works perfectly) and n the lowest state (i.e., complete failure), the dependability matrix becomes triangular with all entries below the diagonal being zeros.

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \cdot\cdot & d_{1n} \\ 0 & d_{22} & d_{23} & \cdot\cdot & d_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdot\cdot & d_{nn} \end{bmatrix} \tag{10.6}$$

If the matrix is properly formulated the sum of the entries in each row must equal unity. For example, for the first row we must have

$$d_{11} + d_{12} + \ldots + d_{1n} = 1 \tag{10.7}$$

and the same must apply to each subsequent row. This provides a good check when formulating a dependability matrix.

The **capability matrix**, C, describes system performance or capability to perform while in any of the n possible system states. If only a single measure of system effectiveness is of importance or of interest, C will be a one column matrix with n elements, such as

$$\mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix} \tag{10.8}$$

where c_j represents system performance when the system is in State j.

System effectiveness, SE, in the WSEIAC model is then defined as

$$SE = \left[a_1, a_2, ..., a_n \right] \bullet \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \bullet \begin{bmatrix} C_1 \\ C_2 \\ \cdot \\ \cdot \\ \cdot \\ C_n \end{bmatrix} \tag{10.9}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \bullet d_{ij} \bullet c_j \tag{10.10}$$

Reference [2] contains several numerical examples of how to perform system effectiveness calculations using the WSEIAC model. Also, Ref. [3], Chapter VII, discusses the model at length and provides numerical examples.

10.2.3 The Navy Concept of System Effectiveness (Ref. [4])

In the early 1960's, the Navy developed a system effectiveness concept which also combines three basic system characteristics: performance, availability and utilization. It can be expressed as "a measure of the extent to which a system can be expected to complete its assigned mission within an established time frame under stated environmental conditions." It may also be defined mathematically as "the probability that a system can successfully meet an operational demand through a given time period when operated under specified conditions."

Mathematically it has been formulated as follows:

$$E_S = PAU \tag{10.11}$$

where:

E_S = index of system effectiveness

P = index of system performance - a numerical index expressing system capability, assuming a hypothetical 100% availability and utilization of performance capability in actual operation

A = index of the system availability - a numerical index of the extent to which the system is ready and capable of fully performing its assigned mission(s)

U = index of system utilization - a numerical index of the extent to which the performance capability of the system is utilized during the mission

The components of the Navy model are not as readily computed as are those of the ARINC and WSEIAC models. The Navy has stated that the terms PU and A are similar to the WSEIAC terms C and AD (Ref. [5]) and that PAU can be translated into the analytical terms P_C and P_T

where:

P_C **performance capability** - a measure of adequacy of design and system degradation

P_T **detailed time dependency** - a measure of availability with a given utilization

Thus the Navy model is compatible with the WSEIAC model in the following way:

$$f(PAU) = f(P_C, P_T) = f(A,D,C) \tag{10.12}$$

The WSEIAC, Navy and ARINC concepts of system effectiveness are depicted in Figure 10.2-1.

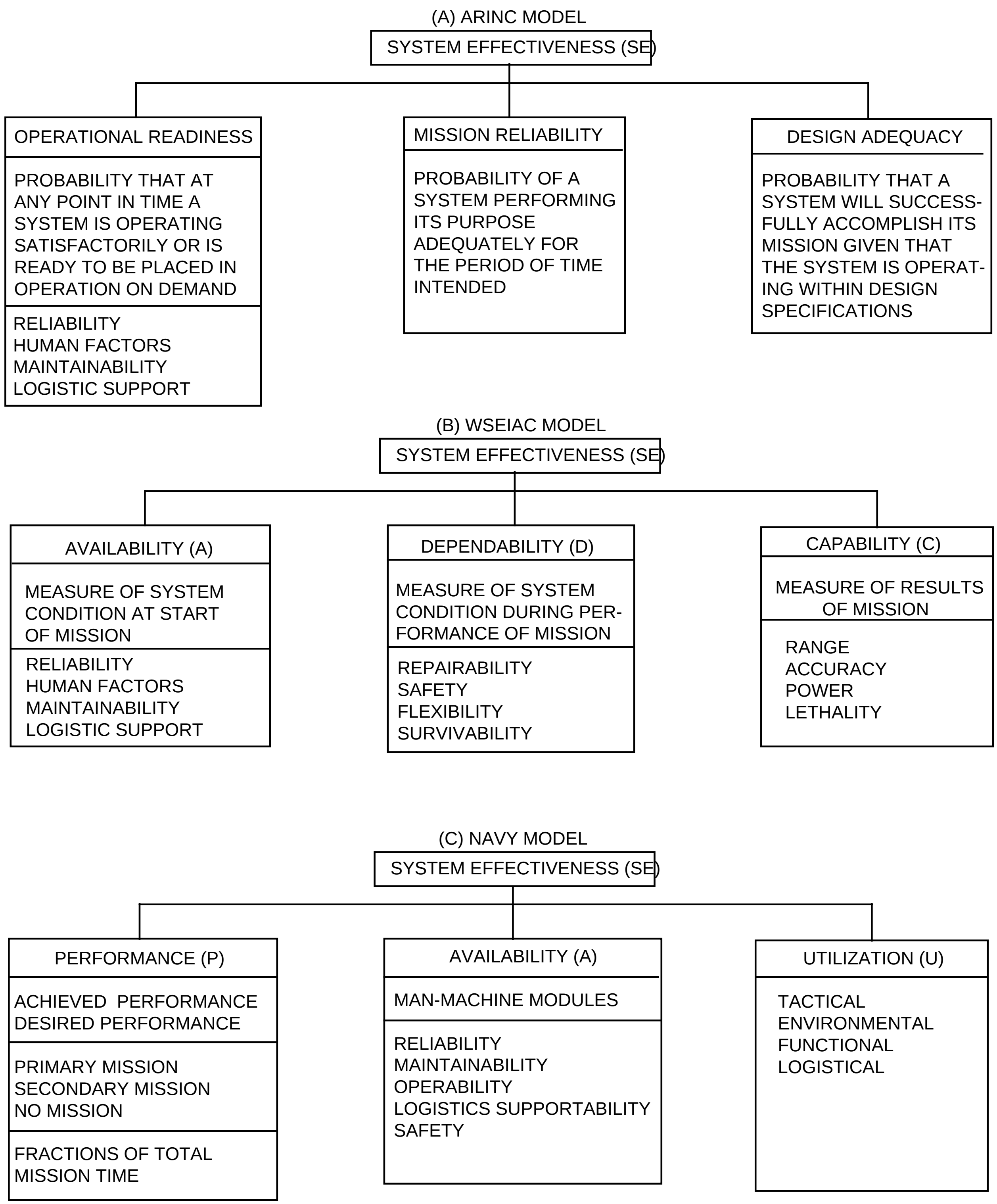


FIGURE 10.2-1: SYSTEM EFFECTIVENESS MODELS

Although these models are relatively simple to describe, their development and application is a rather complex process usually performed by operations research groups and operations analysts utilizing available computerized models (to be discussed later).

10.2.4 An Illustrative Model of a System Effectiveness Calculation

The following simplified example, utilizing the WSEIAC concept, is provided in order to show how R&M parameters are used in system effectiveness calculations.

Problem Statement

The system to be considered consists of a helicopter and its communication equipment. It is to operate in a limited warfare environment where rapid movement of supplies upon request is important. The mission of the system is that upon random call of transporting supplies from a central supply to operational activities within a radius of one-half hour flying time and providing vertical underway replenishment of needed spares. Once the helicopter has reached the target area, proper functioning of the communication equipment enhances the chances of a successful delivery of the supplies in terms of safe delivery, timely delivery, etc. Some major assumptions which are inherent in this example are:

(1) A call for supplies is directed to a single helicopter. If this craft is not in flyable condition (i.e., it is in process of maintenance), the mission will not be started. A flyable craft is defined as one which is in condition to take off and fly with a standard supply load.

(2) The flight time required to reach the target area is one-half hour.

(3) The communication equipment cannot be maintained or repaired in flight.

(4) A loaded helicopter which goes down while enroute to, or which does not reach, the target area, has no delivery value.

Model Determination

For purposes of model formulation, the system condition is divided into three states:

(1) State 1: Helicopter flyable, communication equipment operable

(2) State 2: Helicopter flyable, communication equipment nonoperable

(3) State 3: Helicopter nonflyable

The WSEIAC model for effectiveness is given by the following equation:

$$SE = ADC$$

where A, D and C are defined as follows:

(1) The availability vector is a three-element, row vector, i.e.,

$$A = (a_1 , a_2 , a_3)$$

where a_i is the probability that the helicopter will be in State i at the time of call.

(2) The dependability matrix is a 3x3 square matrix, i.e.,

$$D = \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}$$

where d_{ij} is the probability that if the helicopter is in State i at the time of call it will complete the mission in State j.

(3) The capability vector is a three-element column vector, i.e.,

$$C = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

where c_i is the probability that if the helicopter is in State i at the time of arrival at the target area the supplies can be successfully delivered. (For multi-capability items, C would be a multi-column matrix.)

Determination of Model Elements

Past records indicate that the average time between maintenance activities (including preventive and failure initiated maintenance) for this type of helicopter is 100 hours and the average duration (including such variables as maintenance difficulty, parts availability, manpower, etc.) of a maintenance activity is ten hours. Comparable data for the communication equipment shows an average time between maintenance activities of 500 hours and an average duration of a maintenance activity of five hours.

From the preceding data the elements of A can be determined.

$$A_1 = \text{P(helicopter flyable)} \bullet \text{P(communication equipment operable)}$$

$$= \left(\frac{100}{100 + 10}\right)\left(\frac{500}{500 + 5}\right) = 0.9$$

$$A_2 = \text{P(helicopter flyable)} \bullet \text{P(communication equipment not operable)}$$

$$= \left(\frac{100}{100 + 10}\right)\left(\frac{5}{500 + 5}\right) = 0.009$$

$$A_3 = \text{P(helicopter not flyable)} = \left(\frac{10}{100 + 10}\right) = 0.091$$

Data from past records indicates that the time between failures of the communication equipment during flight is exponentially distributed with a mean of 500 hours. Also, the probability that a helicopter in flight will not survive the one-half hour flight to its destination is 0.05 (includes probability of being shot down, mechanical failures, etc.). Then the elements of the D matrix may be calculated as follows:

(1) If the system begins in State 1:

$$d_{11} = \text{P(helicopter will survive flight)} \bullet \text{P(communication equipment will remain operable)}$$

$$= (1 - 0.05) \exp\left[\left(-\frac{1/2}{500}\right)\right] = 0.94905$$

$$d_{12} = \text{P(helicopter will survive flight)} \bullet \text{P(communication equipment will fail during flight)}$$

$$= (1 - 0.05) \left[1 - \exp\left(-\frac{1/2}{500}\right)\right] = 0.00095$$

d_{13} = P(helicopter will not survive the flight) = 0.05000

(2) If the system begins in State 2:

d_{21} = 0 because the communication equipment cannot be repaired in flight

d_{22} = P(helicopter will survive flight) = 0.95000

d_{23} = P(helicopter will not survive the flight) = 0.05000

(3) If the system begins in State 3:

d_{31} = d_{32} = 0 because the mission will not start

d_{33} = 1, i.e., if the helicopter is not flyable, it will remain nonflyable with reference to a particular mission

Experience and technical judgment have determined the probability of successful delivery of supplies to be c_i if the system is in State i at the time of arrival in the target area, where

$c_1 = 0.95 \quad c_2 = 0.80 \quad c_3 = 0$

Determination of Effectiveness

The effectiveness of the subject system becomes

$$E = \begin{bmatrix} 0.900 & 0.009 & 0.091 \end{bmatrix} \begin{bmatrix} 0.94905 & 0.00095 & 0.05 \\ 0 & 0.95 & 0.05 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.95 \\ 0.8 \\ 0 \end{bmatrix} = 0.82$$

which means that the system has a probability of 0.82 of successful delivery of supplies upon random request.

The effectiveness value attained provides a basis for deciding whether improvement is needed. The model also provides the basis for evaluating the effectiveness of alternative systems considered.

10.3 System R&M Parameters

In this section we are concerned with those system effectiveness submodels, e.g., availability, dependability, operational readiness, which can be exercised to specify, predict, allocate, optimize, and measure system R&M parameters.

Four types of parameters and examples of specific R&M terms applicable to their specification and measurement, are shown in Table 10.3-1. Each will be discussed in more detail in the following paragraphs.

TABLE 10.3-1: SYSTEM R&M PARAMETERS

OBJECTIVES	EXAMPLE TERMS
• READINESS OR AVAILABILITY	R: Mean Time Between Downing Events M: Mean Time to Restore System
• MISSION SUCCESS	R: Mission Time Between Critical Failures M: Mission Time to Restore Function
• MAINTENANCE MANPOWER COST	R: Mean Time Between Maintenance Actions M: Direct Man-hours per Maintenance Action
• LOGISTIC SUPPORT COST	R: Mean Time Between Removals M: Total Parts Cost per Removal

Operational Readiness R&M Parameters - These parameters will define the R&M contribution to the readiness measurement of the system or unit. R&M by itself does not define readiness; there are many other factors relating to personnel, training, supplies, etc., that are necessarily included in any real measure of readiness. The context of readiness includes many factors beyond the realm of equipment capability and equipment R&M achievements. R&M parameters of this type concern themselves with the likelihood of failures occurring that would make a ready system no longer ready and with the effort required to restore the system to the ready condition. Examples of this type of parameter are “mean time between downing events” for reliability and “mean time to restore system” for maintainability.

Mission Success R&M Parameters - These parameters are similar to the classical reliability discussion that is found in most reliability text books. They relate to the likelihood of failures occurring during a mission that would cause a failure of that mission and the efforts that are directed at correcting these problems during the mission itself. Examples would be “mission time between critical failures (MTBCF)” for reliability and “mission time to restore function” for maintainability.

Maintenance Manpower Cost R&M Parameters - Some portion of a system's maintenance manpower requirement is driven by the system's R&M achievement. This category of system R&M parameters concerns itself with how frequently maintenance manpower is required and, once it is required, how many man-hours are needed. Examples of this type of parameter are "mean time between maintenance actions" for reliability and "direct man-hours to repair" for maintainability. Note that the maintainability example does not address the clock hours to complete the repair. Time to restore the system, i.e., the system downtime, is not as significant to the people concerned with manpower needs as the total man-hours required.

Logistic Support Cost R&M Parameters - In many systems, this type of R&M parameter might be properly titled as "material cost" parameters. These parameters address the aspect of R&M achievement that requires the consumption of material. Material demands also relate to the readiness or availability of the system. Examples are "mean time between removals" for reliability and "total parts cost per removal" for maintainability.

Let us examine some of the techniques for using reliability data, reduced to parameters such as those just discussed, for making reliability predictions.

10.3.1 Parameter Translation Models

Frequently it is necessary to convert various reliability parameters from one set of environmental conditions to a different set of environmental conditions. Extensive reliability data may have been generated or gathered in a given environment while the equipment may be subsequently slated for use in an entirely different environment. In other cases, the customers may define a reliability parameter differently than the manufacturer does, or he may use an entirely different figure-of-merit as the basis for acceptance. The intent of this section is to address these areas of concern.

10.3.1.1 Reliability Adjustment Factors

"What if" questions are often asked regarding reliability figures of merit for different operating conditions. For example, what reliability could be expected from a product in a ground fixed environment that is currently experiencing a 700 hour MTBF in an airborne environment. Tables have been derived to make estimates of the effects of quality levels, environments and temperatures enabling rapid conversions between environments. The database upon which these tables are based was a grouping of approximately 18,000 parts from a number of equipment reliability predictions performed on various military contracts. Ratios were developed using this database and the MIL-HDBK-217F algorithms. The relative percentages of each part type in the database are shown in Figure 10.3-1.

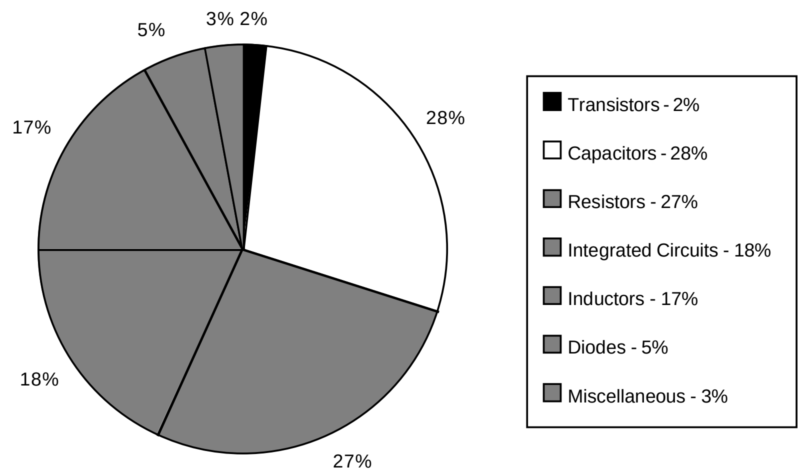


FIGURE 10.3-1: PART DATABASE DISTRIBUTION

(Source: Reliability Toolkit: Commercial Practices Edition, Rome Laboratory and Reliability Analysis Center, Rome, NY 1995).

The following tables, 10.3-2 through 10.3-4, provide a means of converting a known reliability value, expressed as an MTBF, from one set of conditions to another.

TABLE 10.3-2: PART QUALITY FACTORS (MULTIPLY SERIES MTBF BY)

		To Quality Class			
	Part Quality	**Space**	**Military**	**Ruggedized**	**Commercial**
	Space	X	0.8	0.5	0.2
From	**Full Military**	1.3	X	0.6	0.3
Quality	**Ruggedized**	2.0	1.7	X	0.4
Class	**Commercial**	5.0	3.3	2.5	X

Space - Extra Testing Beyond Full Military
Military - Standardized 100% Chip Testing
Ruggedized - Selected 100% Chip Testing
Commercial - Vendor Discretion Testing

TABLE 10.3-3: ENVIRONMENTAL CONVERSION FACTORS (MULTIPLY SERIES MTBF BY)

From Environment \ To Environment	G_B	G_F	G_M	N_S	N_U	A_{IC}	A_{IF}	A_{UC}	A_{UF}	A_{RW}	S_F
G_B	X	0.5	0.2	0.3	0.1	0.3	0.2	0.1	0.1	0.1	1.2
G_F	1.9	X	0.4	0.6	0.3	0.6	0.4	0.2	0.1	0.2	2.2
G_M	4.6	2.5	X	1.4	0.7	1.4	0.9	0.6	0.3	0.5	5.4
N_S	3.3	1.8	0.7	X	0.5	1.0	0.7	0.4	0.2	0.3	3.8
N_U	7.2	3.9	1.6	2.2	X	2.2	1.4	0.9	0.5	0.7	8.3
A_{IC}	3.3	1.8	0.7	1.0	0.5	X	0.7	0.4	0.2	0.3	3.9
A_{IF}	5.0	2.7	1.1	1.5	0.7	1.5	X	0.6	0.4	0.5	5.8
A_{UC}	8.2	4.4	1.8	2.5	1.2	2.5	1.6	X	0.6	0.8	9.5
A_{UF}	14.1	7.6	3.1	4.4	2.0	4.2	2.8	1.7	X	1.4	16.4
A_{RW}	10.2	5.5	2.2	3.2	1.4	3.1	2.1	1.3	0.7	X	11.9
S_F	0.9	0.5	0.2	0.3	0.1	0.3	0.2	0.1	0.1	0.1	X

Environmental Factors as Defined in MIL-HDBK-217

G_B - Ground Benign; G_F - Ground Fixed; G_M - Ground Mobile; N_S - Naval Sheltered; N_U - Naval Unsheltered; A_{IC} - Airborne Inhabited Cargo; A_{IF} - Airborne Inhabited Fighter; A_{UC} - Airborne Uninhabited Cargo; A_{UF} - Airborne Uninhabited Fighter; A_{RW} - Airborne Rotary Winged; S_F - Space Flight

CAUTION: Do not apply to MTBCF.

TABLE 10.3-4: TEMPERATURE CONVERSION FACTORS
(MULTIPLY SERIES MTBF BY)

	To Temperature (°C)							
	Temp.	**10**	**20**	**30**	**40**	**50**	**60**	**70**
	10	X	0.9	0.8	0.8	0.7	0.5	0.4
	20	1.1	X	0.9	0.8	0.7	0.6	0.5
	30	1.2	1.1	X	0.9	0.8	0.6	0.5
From Temperature (°C)	**40**	1.3	1.2	1.1	X	0.9	0.7	0.6
	50	1.5	1.4	1.2	1.1	X	0.8	0.7
	60	1.9	1.7	1.6	1.5	1.2	X	0.8
	70	2.4	2.2	1.9	1.8	1.5	1.2	X

10.3.1.2 Reliability Prediction of Dormant Products

In the past, analysis techniques for determining reliability estimates for dormant or storage conditions relied on simple rules of thumb such as: “the failure rate will be reduced by a ten to one factor”, or “the expected failure rate is zero.” A more realistic estimate, based on part count failure results, can be calculated by applying the conversion factors shown for the example in Table 10.3-5. The factors convert operating failure rates by part type to dormant conditions for seven scenarios.

These conversion factors were determined using data from various military contracts and algorithms from both MIL-HDBK-217F and RADC-TR-85-91, “Impact of Nonoperating Periods on Equipment Reliability” (Ref. [34]). Average values for operating and dormant failure rates were developed for each scenario. For example, to convert the reliability of an operating airborne receiver to a ground nonoperating condition, determine the number of components by type, then multiply each by the respective operating failure rate obtained from handbook data, field data, or vendor estimates. The total operating failure rate for each type is then converted using the conversion factors of Table 10.3-5. The dormant estimate of reliability for the example receiver is determined by summing the part results.

TABLE 10.3-5: AIRCRAFT RECEIVER CONVERSION:
AIRBORNE OPERATING TO GROUND DORMANT FAILURE RATE (EXAMPLE)

Device	Qty	λ_O	λ_{TO}	Conversion Factor	λ_D
Integrated Circuit	25	0.06	1.50	.04	.060
Diode	50	0.001	0.05	.01	.001
Transistor	25	0.002	0.05	.02	.001
Resistor	100	0.002	0.20	.03	.006
Capacitor	100	0.008	0.80	.03	.024
Switch	25	0.02	0.50	.10	.050
Relay	10	0.40	4.00	.04	.160
Transformer	2	0.05	0.10	.20	.020
Connector	3	1.00	3.00	.003	.009
Printed Circuit Board	1	0.70	0.70	.01	.007
Totals	---	---	10.9	---	0.338

λ_O = Part (Operating) Failure Rate (Failures per Million Hours)

λ_{TO} = Total Part (Operating) Failure Rate (Failures per Million Hours)

λ_D = Total Part Dormant Failure Rate (Failures per Million Hours)

Mean-Time-Between-Failure (Operating) = 92,000 hours
Mean-Time-Between-Failure (Dormant) = 2,960,000 hours

10.3.2 Operational Parameter Translation

Field operation typically introduces factors which are beyond the control of designers (e.g. maintenance policy). Thus, "design" reliability may not be the same as "operational" reliability. For this reason, it is often necessary to convert, or translate, from "design" to "operational" terms and vice versa. This translation technique is based on RADC-TR-89-299, "Reliability and Maintainability Operational Parameter Translation II" (Ref. [35]) which developed models for the two most common environments, ground and airborne. While these models are based on military use, similar differences can be expected for commercial products. The translation models are summarized in Table 10.3-6.

TABLE 10.3-6: RELIABILITY TRANSLATION MODELS

R_F Selection							
	Communication	Navigation	Computer	Counter Measure	Radar	All Other	Dependent Var. Lower Bound (% of Ind. Var.)*
1. Airborne Fighter Models							
1A. $MTBF_F = \theta_P^{.64} R_F \left(\frac{C}{D}\right)^{-.46}$	2.1	6.5	5.9	4.7	3.6	4.3	48
1B. $MTBM_F = \theta_P^{.64} R_F \left(\frac{C}{D}\right)^{-.57}$	1.1	2.7	1.9	2.8	1.7	2.0	24
1C. $MTBR_F = \theta_P^{.62} R_F \left(\frac{C}{D}\right)^{-.77}$	1.8	4.4	3.0	5.9	2.5	3.2	34
1D. $MTBF_F = \theta_D^{.76} R_F \left(\frac{C}{D}\right)^{-.34}$	2.1	5.0	5.3	3.7	5.1	2.2	79
1E. $MTBM_F = \theta_D^{.75} R_F \left(\frac{C}{D}\right)^{-.44}$	1.4	2.2	1.8	2.4	2.8	.90	36
1F. $MTBR_F = \theta_D^{.77} R_F \left(\frac{C}{D}\right)^{-.65}$	1.6	4.0	2.2	3.4	3.0	.83	49
2. Airborne Transport Models	**R_F, Uninhabited Equipment**			**R_F, Inhabited Equipment**			
2A. $MTBF_F = \theta_P^{.73} R_F \left(\frac{C}{D}\right)^{-.46}$	2.7			2.5			50
2B. $MTBM_F = \theta_P^{.69} R_F \left(\frac{C}{D}\right)^{-.57}$	1.6			1.4			26
2C. $MTBR_F = \theta_P^{.66} R_F \left(\frac{C}{D}\right)^{-.77}$	2.1			2.3			35
2D. $MTBF_F = \theta_D^{1.0} R_F \left(\frac{C}{D}\right)^{-.34}$	.58			.39			91
2E. $MTBM_F = \theta_D^{1.1} R_F \left(\frac{C}{D}\right)^{-.44}$	.13			.09			44
2F. $MTBR_F = \theta_D^{.88} R_F \left(\frac{C}{D}\right)^{-.65}$	.78			.60			72
3. Ground System Models	**R_F, Fixed Equipment**			**R_F, Mobile Equipment**			
3A. $MTBF_F = \theta_P^{.60} R_F$	27			4.8			90
3B. $MTBM_F = \theta_P^{.67} R_F$	11			1.8			49
3C. $MTBR_F = \theta_P^{.50} R_F$	91			18			80

*The field numeric (i.e., MTBFF, MTBMF or MTBRF) is always taken to be the lesser of (1) the calculated value from Column 1 or, (2) the percentage shown of the independent variable (i.e., θ_P or θ_D).

10.3.2.1 Parameter Definitions

- Mean-time-between-failure-field ($MTBF_F$) includes inherent maintenance events which are caused by design and manufacturing defects.

$$MTBF_F = \frac{\text{Total Operating Hours or Flight Hours}}{\text{Inherent Maintenance Events}}$$

- Mean-time-between-maintenance-field ($MTBM_F$) consists of inherent, induced and no defect found maintenance actions.

$$MTBM_F = \frac{\text{Total Operating Hours or Flight Hours}}{\text{Total Maintenance Events}}$$

- Mean-time-between-removals-field ($MTBR_F$) includes all removals of the equipment from the system.

$$MTBR_F = \frac{\text{Total Operating Hours or Flight Hours}}{\text{Total Equipment Removals}}$$

- θ_P = the predicted MTBF (i.e., estimated by failure rates of the part population)
- θ_D = the demonstrated MTBF (i.e., controlled testing)
- R_F = the equipment type or application constant
- C = the power on-off cycles per mission or operating event
- D = the mission duration or operating event

10.3.2.2 Equipment Operating Hour to Flight Hour Conversion

For airborne categories - $MTBF_F$ represents the mean-time-between-failure in equipment operating hours. To obtain $MTBF_F$ in terms of flight hours (for both fighter and transport models), divide $MTBF_F$ by 1.2 for all categories except countermeasures. Divide by .8 for countermeasures equipment.

Example 1:

Estimate the MTBM of a fighter radar given a mission length of 1.5 hours, two radar shutdowns per mission and a predicted radar MTBF of 420 hours. Using Model 1B in Table 10.3-6,

$$MTBM_F = \theta_P^{.64} R_F \left(\frac{C}{D}\right)^{-.57} = (420 \text{ hr.})^{.64} \; 1.7 \left(\frac{2 \text{ cyc.}}{1.5 \text{ hr.}}\right)^{-.57}$$

$MTBM_F$ = 69 equipment operating hours between maintenance.

SECTION 10: SYSTEMS RELIABILITY ENGINEERING

Since this is below the dependent variable lower bound of (.24)(420) = 101 hours, the calculated $MTBM_F$ is correct. Since this equipment is often turned on for pre- and post-flight checkout, the number of flight hours between maintenance is somewhat less than the actual equipment operating hours. The number of flight hours between maintenance is approximately 69/1.2 = 58 hours.

Example 2:

Estimate the MTBF of a commercial airline navigation unit used on an 8 hour flight and shut down after the flight. The predicted MTBF for the navigation unit is 2,000 hours. Using model 2A for inhabited environment,

$$MTBF_F = \theta_P^{.73} R_F \left(\frac{C}{D}\right)^{-.46}$$

$$= (2{,}000)^{.73}\ 2.5 \left(\frac{1 \text{ cycle}}{8 \text{ hours}}\right)^{-.46}$$

$MTBF_F$ = 1,672 hours between failure

The number of flight hours between failure is estimated to be 1,672/1.2 = 1,393 hours. However, in accordance with the footnote of Table 10.3-6, we calculate a value of (.50)(2000) = 1000 hours using the dependent variable bound. Since this is less than the previous calculation, this is the value to be used.

10.3.3 Availability, Operational Readiness, Mission Reliability, and Dependability - Similarities and Differences

As can be seen from their definitions in Table 10.3-7, availability and operational readiness refer to the capability of a system to perform its intended function when called upon to do so. This emphasis restricts attention to probability “at a point in time” rather than “over an interval of time.” Thus, they are point concepts rather than interval concepts. To differentiate between the two: availability is defined in terms of operating time and downtime, where downtime includes active repair time, administrative time, and logistic time; whereas, operational readiness includes all of the availability times plus both free time and storage time, i.e., all calendar time.

TABLE 10.3-7: DEFINITIONS OF KEY R&M SYSTEM PARAMETERS

AVAILABILITY: A measure of the degree to which an item is in an operable and committable state at the start of a mission when the mission is called for at an unknown (random) time. (Item state at start of a mission includes the combined effects of the readiness-related system R&M parameters but excludes mission time.)

OPERATIONAL READINESS: The ability of a military unit to respond to its operation plan(s) upon receipt of an operations order. (A function of assigned strength, item availability, status or supply, training, etc.)

MISSION RELIABILITY: The ability of an item to perform its required functions for the duration of a specified "mission profile."

DEPENDABILITY: A measure of the degree to which an item is operable and capable of performing its required function at any (random) time during a specified mission profile, given item availability at the start of the mission. (Item state during a mission includes the combined effects of the mission-related system R&M parameters but excludes non-mission time.) (This definition is different than the definition of dependability as it appears in IEC documents.)

MEAN-TIME-BETWEEN-DOWNING-EVENTS (MTBDE): A measure of the system reliability parameter related to availability and readiness. The total number of system life units divided by the total number of events in which the system becomes unavailable to initiate its mission(s) during a stated period of time.

MEAN-TIME-TO-RESTORE-SYSTEM (MTTRS): A measure of the system maintainability parameters related to availability and readiness: the total corrective maintenance time associated with downing events divided by the total number of downing events during a stated period of time. (Excludes time for off-system maintenance and repair of detached components.)

MISSION-TIME-BETWEEN-CRITICAL-FAILURES (MTBCF): A measure of mission reliability: the total amount of mission time divided by the total number of critical failures during a stated series of missions.

MISSION-TIME-TO-RESTORE-FUNCTIONS (MTTRF): A measure of mission maintainability: the total corrective critical failure maintenance time divided by the total number of critical failures during the course of a specified mission profile.

MEAN-TIME-BETWEEN-MAINTENANCE-ACTIONS (MTBMA): A measure of the system reliability parameter related to demand for maintenance manpower: the total number of system life units divided by the total number of maintenance actions (preventive and corrective) during a stated period of time.

DIRECT-MAINTENANCE-MAN-HOURS-PER-MAINTENANCE-ACTION (DMMH/MA): A measure of the maintainability parameter related to item demand for maintenance manpower: the sum of direct maintenance man-hours divided by the total number of maintenance actions (preventive and corrective) during a stated period of time.

Also note that the concepts of availability and operational readiness do not include mission time.

Dependability, although it is a point concept like availability and operational readiness, differs from those concepts in that it is concerned with the degree (or probability) that an item is operable at some point (time) during the mission profile, given its (point) availability at the start of the mission.

Mission reliability, on the other hand, is concerned with the ability of a system to continue to perform without failure for the duration of a specified mission time; in other words, the probability of successful operation over some interval of time rather than at a specific point in time. Thus, mission reliability is an interval concept rather than a point concept. It should be pointed out that mission reliability is also conditional upon the system being operable at the beginning of the mission or its (point) availability.

Further note that dependability and mission reliability do not include non-mission time.

Hopefully, the mathematical models and examples which follow will help to further clarify these concepts.

10.4 System, R&M Modeling Techniques

It was previously pointed out in Section 5 that mathematical models represent an efficient, shorthand method of describing an event and the more significant factors which may cause or affect the occurrence of the event. Such models are useful to engineers and designers since they provide the theoretical foundation for the development of an engineering discipline and a set of engineering design principles which can be applied to cause or prevent the occurrence of an event.

At the system level, models such as system effectiveness models (and their R&M parameter submodels) serve several purposes:

(1) To evaluate the effectiveness of a system of a specific proposed design in accomplishing various operations (missions) for which it is designed and to calculate the effectiveness of other competing designs, so that the decision maker can select that design which is most likely to meet specified requirements,

(2) To perform trade-offs among system characteristics, performance, reliability, maintainability, etc., in order to achieve the most desirable balance among those which result in highest effectiveness,

(3) To perform parametric sensitivity analyses in which the numerical value of each parameter is varied in turn and to determine its effect on the numerical outputs of the model. Parameters that have little or no effect can be treated as constants and the model simplified accordingly. Parameters to which the model outputs show large sensitivity are then examined in detail, since small improvements in the highly sensitive

parameters may result in substantial improvements in system effectiveness at very acceptable cost,

(4) To “flag” problem areas in the design which seriously limit the ability of the design to achieve the desired level of system R&M or system effectiveness.

The evaluation of system effectiveness and its R&M parameters is an iterative process that continues through all life cycle phases of a system. In each of these phases, system effectiveness is continually being “measured” by exercising the system effectiveness models. In the early design stage, system effectiveness and R&M predictions are made for various possible system configurations. When experimental hardware is initially tested, first real life information is obtained about performance, reliability, and maintainability characteristics, and this information is fed into the models to update the original prediction and to further exercise the models in an attempt to improve the design. This continues when advanced development hardware is tested to gain assurance that the improvements in the system design are effective or to learn what other improvements can still be made before the system is fully developed, type classified, and deployed for operational use. Once in operation, field data starts to flow in and the models are then used to evaluate the operational effectiveness of the system as affected by the field environment, including the actual logistic support and maintenance practices provided in the field. The models again serve to disclose or “flag” problem areas needing improvement.

One may summarize the need for system R&M models as follows:

They provide insight, make an empirical approach to system design and synthesis economically feasible, and are a practical method for circumventing a variety of external constraints. Furthermore, the models aid in establishing requirements, provide an assessment of the odds for successful mission completion, isolate problems to definite areas, and rank problems to their relative seriousness of impact on the mission. They also provide a rational basis for evaluation and choice of proposed system configurations and for proposed solutions to discovered problems.

Thus, system R&M models are an essential tool for the quantitative evaluation of system effectiveness and for designing effective weapon systems. Figure 10.4-1 identifies eight principal steps involved in system effectiveness evaluation. Step 1 is mission definition, Step 2 is system description, Step 3 is selection of figure of merit, and Step 4 is the identification of accountable factors that impose boundary conditions and constraints on the analysis to be conducted. After completing these four Steps, it becomes possible to proceed with Step 5, the construction of the mathematical models. To obtain numerical answers from the models, numerical values of all parameters included in the models must be established or estimated (Step 7). To do this, good and reliable data must first be acquired from data sources, tests, etc. (Step 6). In the final Step 8, the models are exercised by feeding in the numerical parametric values to obtain system effectiveness estimates and to perform optimizations. Ref. [7] illustrates in more detail the whole process of system effectiveness evaluations, beginning with the military

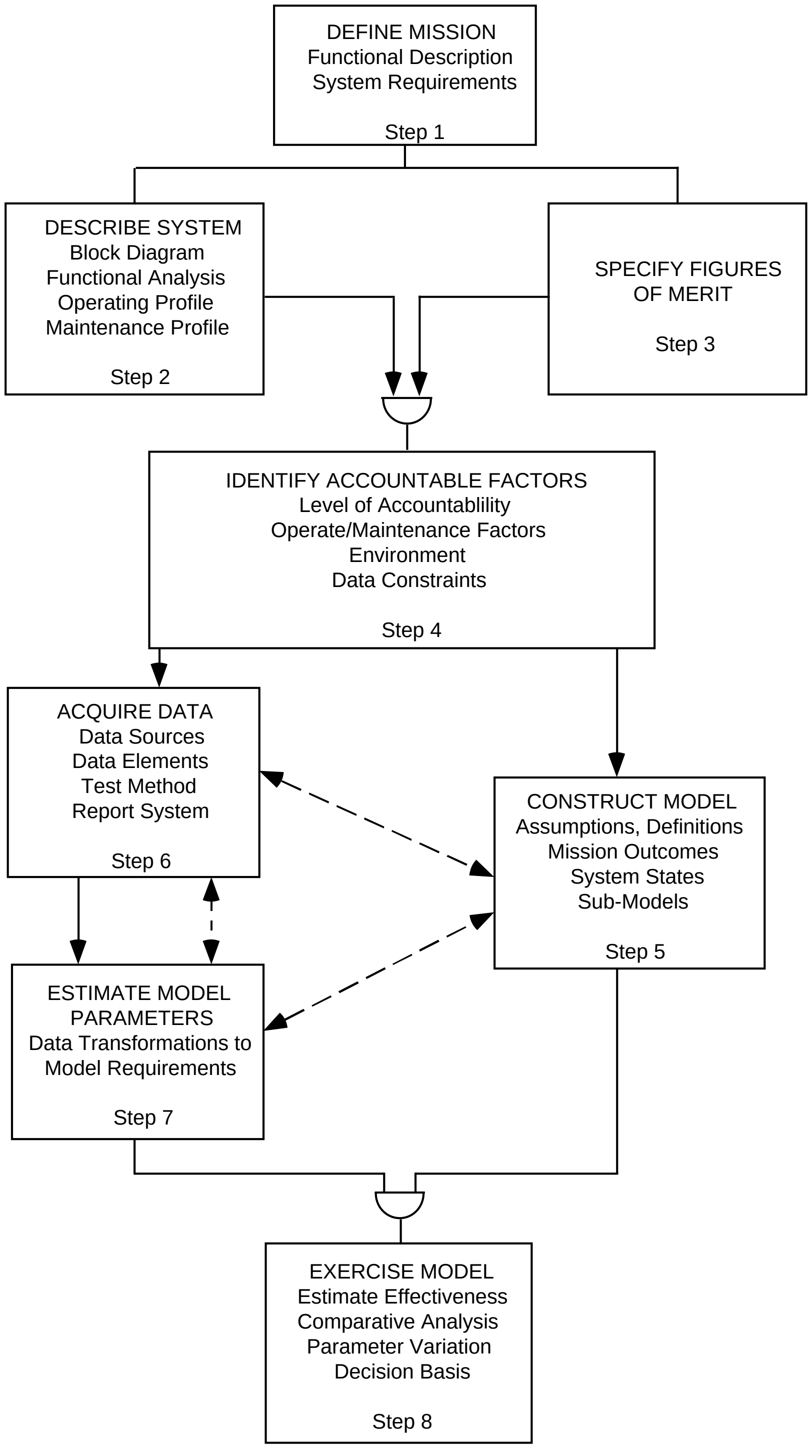


FIGURE 10.4-1: PRINCIPAL STEPS REQUIRED FOR EVALUATION OF SYSTEM EFFECTIVENESS

operational requirements and progressing through the exercising of the system effectiveness model(s) to the decision-making stage.

In terms of system R&M parameter models, reliability and maintainability define system availability and/or operational readiness. Reliability determines the state probabilities of the system during the mission, i.e., the system dependability. If repairs can be performed during the mission, maintainability also becomes a factor in dependability evaluations; this case is often referred to as “reliability with repair.” Then, there is the impact of logistic support on the downtime and turnaround time of the system, since shortcomings in the logistic support may cause delays over and above the maintenance time as determined by the system maintainability design. Finally, there are the performance characteristics of the system that are affected by the state in which the system may be at any point in time during a mission, i.e., by the system dependability.

Submodels of availability, operational readiness, downtime distributions, dependability, etc., are required to obtain the numerical answers that may be fed into an overall system effectiveness model, if such can be constructed. Some of these submodeling techniques will now be discussed.

10.4.1 Availability Models

The concept of availability was originally developed for repairable systems that are required to operate continuously, i.e., round-the-clock, and are at any random point in time either operating or “down” because of failure and are being worked upon so as to restore their operation in minimum time. In this original concept, a system is considered to be in only two possible states -- operating or in repair -- and availability is defined as the probability that a system is operating satisfactorily at any random point in time, t, when subject to a sequence of “up” and “down” cycles which constitute an alternating renewal process.

Availability theory was treated quite extensively in Section 5; this section will concentrate on final results and illustrative examples of the various models.

10.4.1.1 Model A - Single Unit System (Point Availability)

Consider first a single unit system or a strictly serial system that has a reliability, R(t); its availability, A(t), that it will be in an “up” state (i.e., will be operating) at time, t, when it started in an “up” condition at t = 0 is given by:

$$A(t) = \frac{\mu}{\lambda + \mu} + \left\{ \frac{\lambda}{\lambda + \mu} \exp\left[-(\lambda + \mu)t \right] \right\} \quad (10.13)$$

where:

λ is the failure rate and μ is the repair rate

If it started in a “down” state at t = 0

$$A(t) = \frac{\mu}{\lambda + \mu} - \left\{ \frac{\lambda}{\lambda + \mu} \exp\left[-(\lambda + \mu)t \right] \right\} \qquad (10.14)$$

This assumes that the probability density functions for failures and repairs are exponentially distributed and given by, respectively:

$$f(t) = \lambda e^{-\lambda t} \qquad (10.15)$$

$$g(t) = \mu e^{-\lambda t} \qquad (10.16)$$

We may write Equation 10.13 also in terms of the reciprocal values of the failure and repair rates, i.e., in terms of the MTBF and the MTTR, remembering, however, that both time-to-failure and time-to-repair must be exponentially distributed for the equation to hold.

$$A(t) = \frac{MTBF}{MTBF + MTTR} + \left\{ \frac{MTTR}{MTBF + MTTR} \bullet \exp\left[-\left(\frac{1}{MTBF} + \frac{1}{MTTR} \right) t \right] \right\} \qquad (10.17)$$

When we study this equation we see that as t increases the second term on the right diminishes and that availability in the limit becomes a constant, i.e.,

$$\lim_{t \to \infty} A(t) = A_s = \frac{MTBF}{MTBF + MTTR} \qquad (10.18)$$

We call this the steady-state availability or inherent uptime ratio of a serial system. It is equivalent to the intrinsic availability, A_i, discussed in Section 5.

Figure 10.4-2 shows plots of A(t), instantaneous availability, and A_i or A_s (steady state availability) for a single system having a failure rate, (λ), of 0.01 failures/hour and a repair rate (μ), of 1 repair/hour.

Note that the transient term decays rather rapidly; it was shown in Section 5 that the transient term becomes negligible for

$$t \geq \frac{4}{\lambda + \mu} \qquad (10.19)$$

An important point to be made is that Eq. (10.18) holds regardless of the probability distribution of time-to-failure and time-to-repair.

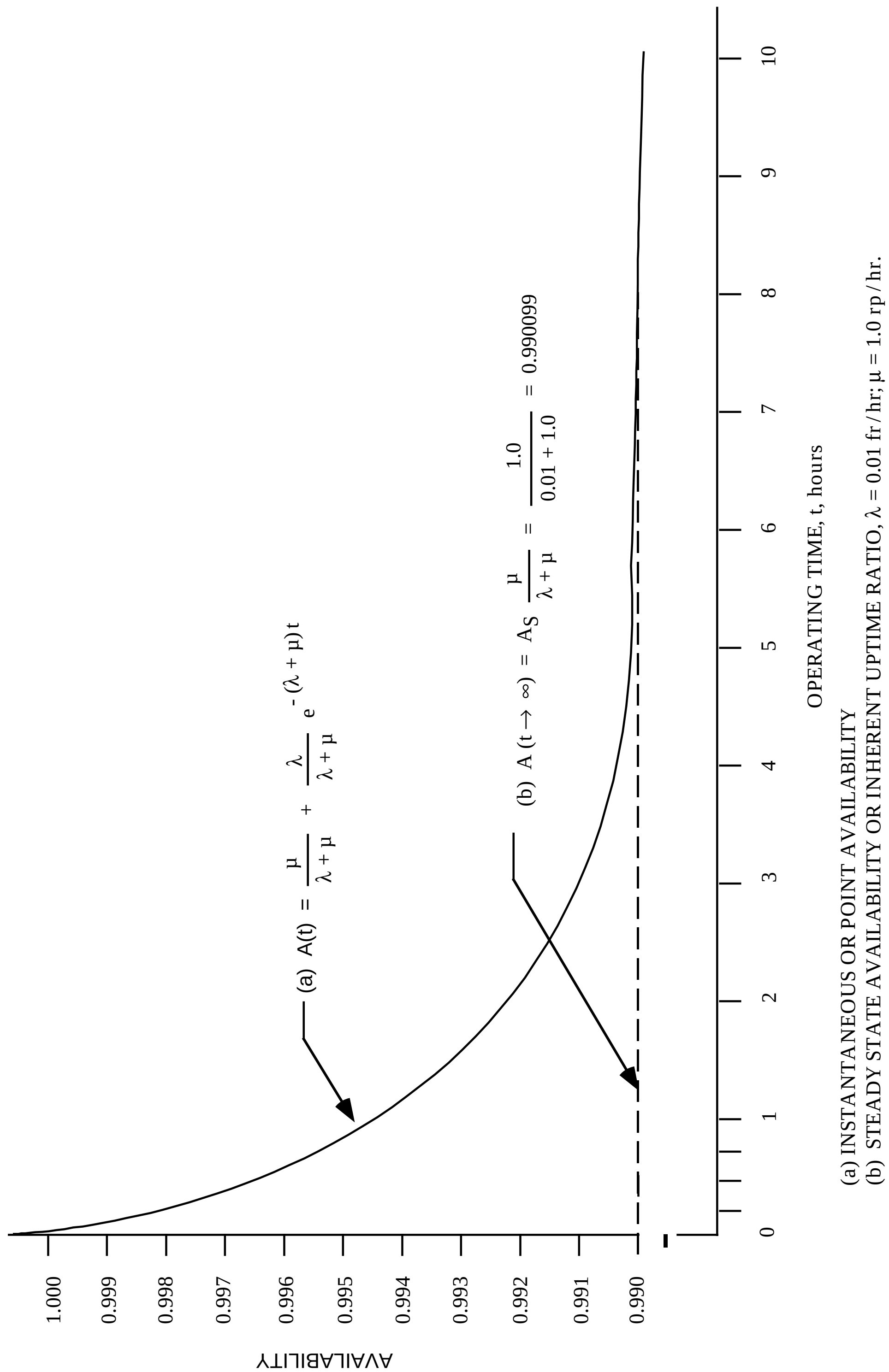


FIGURE 10.4-2: THE AVAILABILITY OF A SINGLE UNIT

Looking again at Eq. (10.18), we may divide the numerator and denominator by the MTBF and write the steady state availability as follows:

$$A = 1/(1 + \alpha) \tag{10.20}$$

where:

α = MTTR/MTBF, the maintenance time ratio (MTR), or alternatively,

α = λ/μ which the reader may recognize from queuing theory as the "utilization" factor. Thus, the availability, A, does not depend upon the actual values of MTBF or MTTR or their reciprocals but only on their ratio.

Since there is a whole range of MTBF ($1/\lambda$) and MTTR ($1/\mu$) values which can satisfy a given availability requirement, the system designer has the option of trading off MTBF and MTTR to achieve the required system availability within technological and cost constraints. This will be discussed later.

Another observation to be made from Eq. (10.20) is that if α, which is equal to MTTR/MTBF, or λ/μ, is less than 0.10, then A_i can be approximated by 1 - MTTR/MTBF, or 1 - λ/μ.

Thus far we have discussed inherent or intrinsic availability which is the fundamental parameter used in equipment/system design. However, it does not include preventive maintenance time, logistic delay time, and administrative time. In order to take these factors into account, we need several additional definitions of availability.

For example, achieved availability, A_a, includes preventive maintenance and is given by the formula:

$$A_a = \frac{\text{MTBM}}{\text{MTBM} + \overline{M}} \tag{10.21}$$

where $\overline{M}$ is the mean active corrective and preventive maintenance time and MTBM is the mean interval between corrective and preventive maintenance actions equal to the reciprocal of the frequency at which these actions occur, which is the sum of the frequency or rate (λ) at which corrective maintenance actions occur and the frequency or rate (f) at which preventive maintenance actions occur.

Therefore,

$$\text{MTBM} = 1/(\lambda + f)$$

Operational availability, A_o, includes, in addition to A_a, logistic time, waiting time, and administrative time, so that the total mean downtime MDT becomes:

$$MDT = \overline{M} + \text{Mean Logistic Time} + \text{Mean Administrative Time}$$

and adds to the uptime the ready time, RT, i.e.,

$$A_o = \frac{MTBM + RT}{MTBM + RT + MDT} \tag{10.22}$$

It is important to realize that RT is the system average ready time (available but not operating) in a complete operational cycle, the cycle being MTBM + MDT + RT.

Example 3: Illustration of Availability Calculations

The following example is provided to clarify the concepts in the subsection. A ground radar system was found to have the following R&M parameters. Determine A_i, A_a, and A_o:

MTBF = 100 hours

MTTR = 0.5 hour

Mean active preventive maintenance time = 0.25 hours

Mean logistic time = 0.3 hour

Mean administrative time = 0.4 hours

MTBM = 75 hours for either corrective or preventive maintenance actions

Mean ready time = 20 hours

Intrinsic or Inherent Availability = A_i

$$A_i = \frac{MTBF}{MTBF + MTTR} = \frac{100}{100 + 0.5} = 0.995$$

Achieved Availability = A_a

$$A_a = \frac{MTBM}{MTBM + \overline{M}} = \frac{75}{75 + 0.5 + 0.25} = 0.99$$

Operational Availability = A_o

$$A_o = \frac{\text{MTBM} + \text{RT}}{\text{MTBM} + \text{RT} + \text{MDT}} = \frac{75 + 20}{75 + 20 + 0.5 + 0.25 + 0.3 + 0.4} = \frac{95}{96.45} = 0.985$$

10.4.1.2 Model B - Average or Interval Availability

What we discussed in the previous section is the concept of point availability which is the probability that the system is “up” and operating at any point in time. Often, however, one may be interested in knowing what percent or fraction of a time interval (a,b) a system can be expected to operate. For example, we may want to determine the availability for some mission time. This is called the interval or average availability, A_{AV} , of a system and is given by the time average of the availability function A(t) averaged over the interval (a,b):

$$A_{AV}(a,b) = \left[\, 1/(b-a) \,\right] \int_b^a A(t)\,dt \tag{10.23}$$

For instance, if we want to know the fraction of time a system such as shown in Figure 10.4-2 will be operating counting from t = 0 to any time, T, we substitute A(t) of Eq. (10.13) into Eq. (10.23) and perform the integration. The result is:

$$A_{AV}(T) = \frac{1}{T}\left[\int_0^T \frac{m}{1+m}\,dt + \int_0^T \frac{l}{1+m}\exp[-(1+m)t]\,dt\right] \tag{10.24}$$

$$= \frac{\mu}{\lambda + \mu} + \frac{\lambda}{T(\lambda + \mu)^2}\left\{\ 1 - \exp\left[\,-(\lambda + \mu)T\right]\right\}$$

Figure 10.4-3 shows the relationship of A(t) to $A_{AV}(t)$ for the exponential case. Note that in the limit in the steady state we again get the availability A of Eq. (10.18), i.e.,

$$\lim_{t \to \infty} A_{AV}(t) = \mu/(\lambda + \mu) = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} \tag{10.25}$$

But in the transient state of the process, as shown in the figure for an interval (0, T), before equilibrium is reached $A_{AV}(t)$ is in the exponential case larger than A(t) for an interval (0, t). This is not true for all distributions, since A(t) and $A_{AV}(t)$ may be subject to very large fluctuations in the transient state.

From Eq. (10.24) we may also get the average or expected “on” time in an interval (0, t) by multiplying $A_{AV}(t)$ and t, the length of the time interval of interest. Ref. [8], pp. 74-83, contains an excellent mathematical treatment of the pointwise and interval availability and related concepts.

Unavailability (U) is simply one minus availability (1-A).

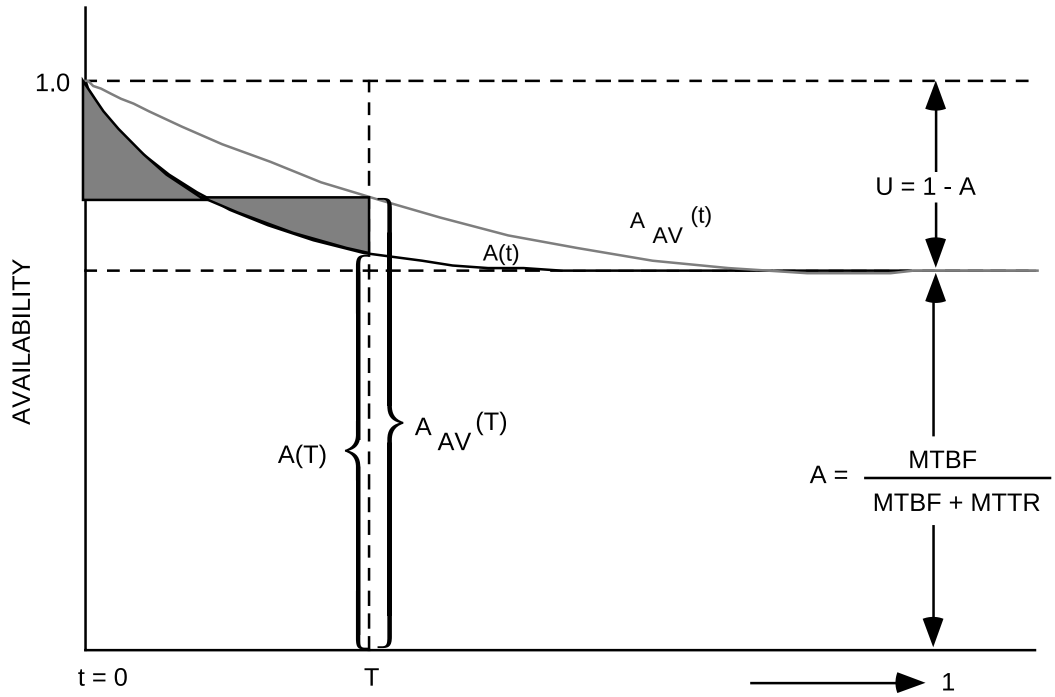


FIGURE 10.4-3: AVERAGE AND POINTWISE AVAILABILITY

Example 4: Average Availability Calculation

Using our ground radar example from the previous subsection, calculate A_{AV} for a mission time of 1 hour.

MTBF = 100 hrs. = $1/\lambda$

MTTR = 0.5 hr. = $1/\mu$

T = 1 hr.

$$A_{AV}(T) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{T(\lambda + \mu)^2}\left\{1 - \exp[-(\lambda + \mu)T]\right\}$$

$$= \frac{2}{2.01} + \frac{0.01}{1(2.01)^2}\left\{1 - \exp[-(2.01)(1)]\right\}$$

$$= \quad 0.995 + 0.0025\,(1 - 0.134)$$

$$= \quad 0.9972$$

and its expected "on" time for a 1-hr. mission would be (0.9972)(60) = 59.8 minutes.

10.4.1.3 Model C - Series System with Repairable/Replaceable Units

When a series system consists of N units (with independent unit availabilities) separately repairable or replaceable whenever the system fails because of any one unit failing, the steady state availability is given by:

$$A = \prod_{i=1}^{N} A_i \tag{10.26}$$

$$= \prod_{i=1}^{N} \left(\frac{1}{1 + \frac{MTTR_i}{MTBF_i}} \right) \tag{10.27}$$

$$= \prod_{i=1}^{N} \left(\frac{1}{1 + \lambda_1 / \mu_i} \right) \tag{10.28}$$

$$= \prod_{i=1}^{N} \left(\frac{1}{1 + \alpha_i} \right) \tag{10.29}$$

where:

$$\alpha_i = \frac{MTTR_i}{MTBF_i} = \frac{\lambda_i}{\mu_i}$$

Furthermore, if each $\frac{MTTR_i}{MTBF_i}$ is much less than 1, which is usually the case for most practical systems, Eq. (10.29) can be approximated by:

$$A = (1 + \Sigma\, \alpha_i)^{-1} \tag{10.30}$$

Caution is necessary in computing α_i, since Eq. (10.30) applies to the availability of the whole system. Thus, when the units are replaceable as line replaceable units or system replaceable units, the $MTTR_i$ is the mean time required to replace the unit with a good one at the system maintenance level and is not the mean repair time of the failed removed unit. On the other hand, if failed units are not replaced but are repaired at the system level, $MTTR_i$ is the mean-time-to-repair of the unit, which becomes also the downtime for the system. Thus, when computing the A_S of the units and the availability A_S of the system, all MTTRs must be those repair times that the system experiences as its own downtime. The $MTTR_i$ of the i^{th} unit is thus the system mean repair time when the i^{th} unit fails.

If we compare Eq. (10.30) with Eq. (10.20) in Model A we find that they are identical. The system maintenance time ratio (MTR) is:

$$\alpha = MTTR/MTBF \tag{10.31}$$

But the serial system's MTTR as shown in Section 4 is given by:

$$MTTR = \Sigma\lambda_i (MTTR_i)/\Sigma\lambda_i \tag{10.32}$$

while its MTBF is

$$\begin{aligned} MTBF = \; & (\Sigma\lambda_i)^{-1} \\ = \; & \Sigma\lambda_i (MTTR_i)\Sigma\lambda_i/\Sigma\lambda_i \\ = \; & \Sigma\lambda_i (MTTR_i) = \Sigma\alpha_i \end{aligned} \tag{10.33}$$

where:

$$\lambda_i = \frac{1}{MTBF_i}$$

In other words, the system MTR is the sum of the unit MTRs. The MTR is actually the average system downtime per system operating hour. Conceptually, it is very similar to the maintenance ratio (MR) defined as maintenance man-hours expended per system operating hour. The difference is that in the MTR one looks only at system downtime in terms of clock hours of system repair, whereas in the MR one looks at all maintenance man-hours expended at all maintenance levels to support system operation.

Eq. (10.30) can be still further simplified if $\sum_{i=1}^{N} \lambda_i/\mu_i < 0.1$

In that case

$$A \approx 1 - \sum_{i=1}^{N} \lambda_i/\mu_i \tag{10.34}$$

or the system availability is equal to 1 - (the sum of the unit MTRs).

Let us work some examples.

Example 5:

Figure 10.4-4 represents a serial system consisting of 5 statistically independent subsystems, each with the indicated MTBF and MTTR. Find the steady state availability of the system.

Note that for the system, we cannot use any of the simplifying assumptions since, for example, subsystems 3 and 4 have MTRs of 0.2 and 0.1, respectively, which are not << than 1.

Also $\sum_{i=1}^{N} \lambda_i/\mu_i = 0.33$ which is not < 0.1.

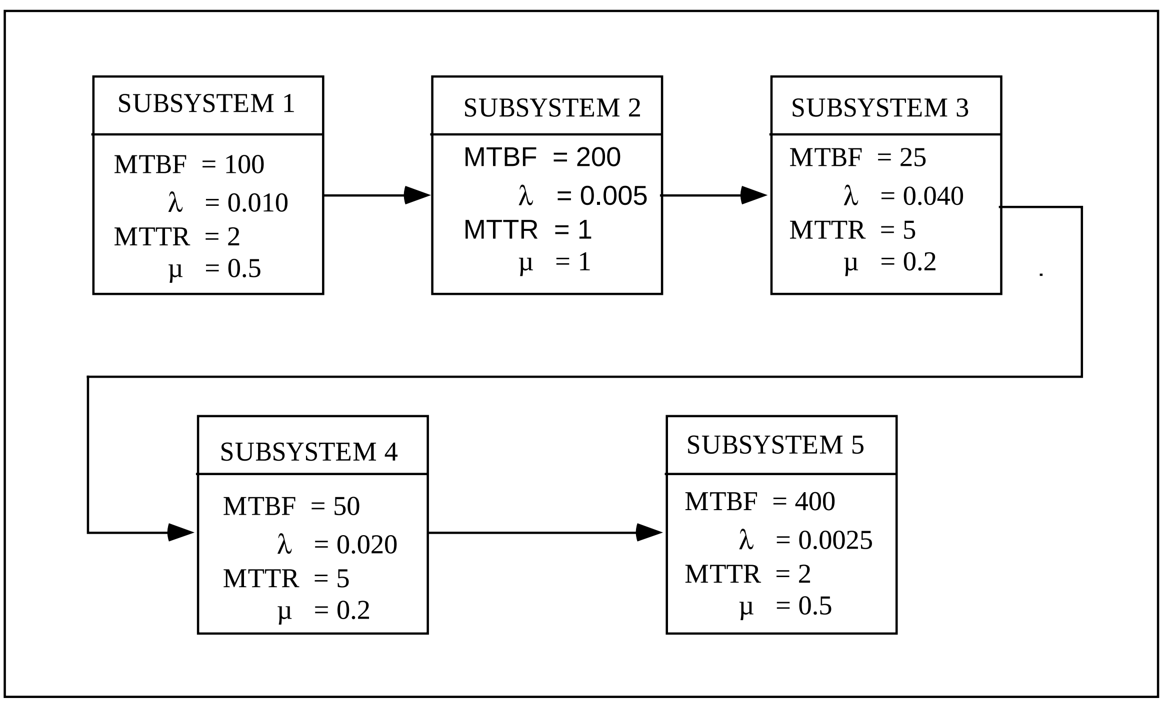


FIGURE 10.4-4: BLOCK DIAGRAM OF A SERIES SYSTEM

Therefore, we must use the basic relationship, Eq. (10.27).

$$A = \prod_{i=1}^{N} \left(\frac{1}{1 + \frac{MTTR_i}{MTBF_i}} \right)$$

$$= \left(\frac{1}{1 + 2/100}\right)\left(\frac{1}{1 + 1/200}\right)\left(\frac{1}{1 + 5/25}\right)\left(\frac{1}{1 + 5/50}\right)\left(\frac{1}{1 + 2/400}\right)$$

$$= (0.98039)\ (0.99502)\ (0.83333)\ (0.90909)\ (0.99502) = 0.73534$$

Example 6:

Now let us look at a similar series system, consisting of 5 statistically independent subsystems having the following MTBFs and MTTRs, as shown in the table below.

Subsystem	MTBF	MTTR	α	A
1	100	0.5	0.005	0.995
2	200	1	0.005	0.995
3	300	0.75	0.0025	0.9975
4	350	1.5	0.0043	0.9957
5	500	2	0.004	0.996

In this case, each α_i is << than 1 and $\sum_{i=1}^{5} \alpha_i < .1$, so that we can use the simplified Eq. (10.34).

$$A \approx 1 - \sum_{i=1}^{5} \lambda_i/\mu_i = 1 - 0.0208 = 0.9792$$

Of course, the power and speed of modern hand-held calculators and personal computers tend to negate the benefits of the simplifying assumptions.

10.4.1.4 Model D - Redundant Systems

(See Section 7.5 for a more detailed description of the mathematical models used to calculate the reliability of systems incorporating some form of redundancy). In this model, the availability of some redundant systems is considered. First we deal with two equal, independent units in a parallel redundant arrangement with each unit being separately repairable or replaceable while the other continues operating. Thus, the system is “up” if both or any one of the two units operates. (See Section 7.5 for a more detailed description of the mathematical models used to calculate the reliability of systems incorporating some form of redundancy).

If we define the unavailability U of a unit as

$$U = 1 - A = \text{MTTR}/(\text{MTBF} + \text{MTTR}) \tag{10.35}$$

then the probability that the system is unavailable is the probability that both units are down at the same time, which is

$$U_{system} = U^2 \tag{10.36}$$

and system availability is

$$A_{system} = 1 - U^2 \tag{10.37}$$

Further, using the binomial expansion

$$(A + U)^2 = A^2 + 2AU + U^2 = 1 \tag{10.38}$$

we find that we may write Eq. (10.38) also in the form

$$A_{system} = A^2 + 2AU \tag{10.39}$$

which gives us the probability A^2 that both units are operating at any point in time and the probability 2AU that only one unit is working. Over a period of time T, the system will on the average be operating for a time TA^2 with both units up, while for 2TAU only one unit will be up. If the performance of the system is P_1 when both units are up and P_2 when only one unit is up, the system output or effectiveness, SE, over T is expected to be

$$SE = P_1 TA^2 + 2P_2 TAU \tag{10.40}$$

Assume a ship has two engines which are subject to on-board repair when they fail. When both engines work, the ship speed is 30 nmi/hour, and when only one engine works it is 20 nmi/hour. Let an engine MTBF be 90 hr. and let its MTTR be 10 hr., so that the availability of an engine is A = 0.9 and its unavailability is U = 0.1. Over a 24-hour cruise the ship will be expected to travel on the average

$$SE = 30 \bullet 24 \bullet .81 + 2 \bullet 20 \bullet 24 \bullet 0.9 \bullet 0.1 = 583.2 + 86.4 = 669.6 \text{ nmi.}$$

The expected time for the ship to be found idle with both engines out for a 24-hour cruise is:

$$T_{idle} = 24U^2 = 24(0.01) = 0.24 \text{ hr} \tag{10.41}$$

For three units in parallel we get

$$(A + U)^3 = A^3 + 3A^2 U + 3AU^2 + U^3 = 1 \tag{10.42}$$

If the system goes down only if all three units are down, system availability is:

$$A_{system} = A^3 + 3A^2 U + 3AU^2 = 1 - U^3 \tag{10.43}$$

but if at least two units are needed for system operation since a single unit is not sufficient, system availability becomes

$$A_{system} = A^3 + 3A^2 U \tag{10.44}$$

In general, for a system with n equal, redundant units, we expand the binomial term

$$(A + U)^n = 1, \text{ or}$$

$$A^n + \left(nA^{n-1}U\right) + \left(\frac{n(n-1)}{2!} A^{n-2} U^2\right) + \left(\frac{n(n-1)(n-2)}{3!} A^{n-3} U^3\right) + \ldots + U^n = 1 \tag{10.45}$$

which yields the probabilities of being in any one of the possible states. Then, by adding the probabilities of the acceptable states, we obtain the availability of the system. As stated earlier, the units must be independent of each other, both in terms of their failures and in terms of their repairs or replacements, with no queuing up for repair.

Reference [9] contains, throughout the text, extensive tabulations of availability and related measures of multiple parallel and standby redundant systems for cases of unrestricted as well as restricted repair when failed redundant units must queue up and wait until their turn comes to get repaired.

Returning briefly to Eq. (10.36), when the two redundant units are not equal but have unavailabilities $U_1 = 1 - A_1$ and $U_2 = 1 - A_2$, system unavailability becomes:

$$U_{system} = U_1 U_2 \tag{10.46}$$

and availability

$$A_{system} = 1 - U_1 U_2 \tag{10.47}$$

Again, we may expand the multinomial

$$(A_1 + U_1)(A_2 + U_2) = A_1A_2 + A_1U_2 + A_2U_1 + U_1U_2 \tag{10.48}$$

and may write system availability in the form

$$A_{system} = A_1A_2 + A_1U_2 + A_2U_1 \tag{10.49}$$

For n unequal units we expand the term

$$\sum_{i=1}^{n} (A_i + U_i) = 1 \tag{10.50}$$

and add together the probabilities of acceptable states and other effectiveness measures, as illustrated in the ship engines example.

This approach is analogous to that shown in Section 5 (k out of n configuration) for reliability.

It can be shown that the limiting expression for an n equipment parallel redundant system reduces to the binomial form if there are as many repairmen as equipments. This is equivalent to treating each equipment as if it had a repairman assigned to it or to saying that a single repairman is assigned to the system but that the probability of a second failure occurring while the first is being repaired is very small. The expression for steady state availability is

$$A\left[1/n\right] = 1 - (1 - A)^n \tag{10.51}$$

where n is the number of redundant equipments and l/n indicates that at least 1 equipment must be available for the system to be available.

In general where at least m out of n redundant equipments must be available for the system to be available:

$$\begin{aligned} A\left[m/n\right] &= \sum_{i=m}^{n} \binom{n}{i} A^i (1 - A)^{n-i} \\ &= \sum_{i=m}^{n} \frac{n!}{(n-1)!i!} \left(\frac{\mu}{\mu+\lambda}\right)^i \left(\frac{\lambda}{\mu+\lambda}\right)^{n-1} \end{aligned} \tag{10.52}$$

Table 10.4-1 (Ref. [10]) provides expressions for the instantaneous and steady state availability for 1, 2, and 3 equipments, parallel and standby redundancy, and single and multiple repair maintenance policies.

Single repair means that failed units can be repaired one at a time. If a unit fails, repairs are immediately initiated on it. If more than one unit is down, repairs are initiated on a single unit until it is fully operational; then, repairs are initiated on the second failed unit. For the case of ***multiple repair***, all failed units can have repair work initiated on them as soon as failure occurs, and the work continues until each unit is operational. Also, a repair action on one unit is assumed to be independent of any other unit.

One case not yet addressed is the case of redundant units when repairs cannot be made until complete system failure (all redundant units have failed). The steady state availability can be approximated by (see Ref. [25] for deriving exact expressions):

$$A = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}} \tag{10.53}$$

where:

MTTF = mean time to failure for redundant system

and

MTTR = mean time to restore all units in the redundant system

In the case of an n-unit ***parallel*** system

$$\text{MTTF} = \sum_{n=1}^{n} \frac{1}{i\lambda} \tag{10.54}$$

and

$$\text{MTTR} = \frac{m}{\mu} \tag{10.55}$$

TABLE 10.4-1: AVAILABILITY OF SOME REDUNDANT SYSTEMS BASED ON EXPONENTIAL FAILURE AND REPAIR DISTRIBUTIONS

No. of Equipments	Conditions: Type Redundancy	Conditions: Repair	Instantaneous Availability Model	Definitions of Constants for Instantaneous Availability Model	Steady State Availability: Model	Steady State Availability: A_{system} for $\lambda = 0.01$, $\mu = 0.2$
1	---	---	$A(t) = \frac{\mu}{\mu+\lambda} + \frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)t}$	———	$\frac{\mu}{\mu+\lambda}$	0.95
2	Standby	Single	$A(t) = \frac{\mu^2+\mu\lambda}{\mu^2+\mu\lambda+\lambda^2} - \frac{\lambda^2(s_2e^{s_1t} - s_1e^{s_2t})}{s_1s_2(s_1-s_2)}$	$s_1 = -(\lambda+\mu) - \sqrt{\mu\lambda}$ $s_2 = -(\lambda+\mu) + \sqrt{\mu\lambda}$	$\frac{\mu^2+\mu\lambda}{\mu^2+\mu\lambda+\lambda^2}$	0.998
		Multiple	$A(t) = \frac{2\mu^2+2\mu\lambda}{2\mu^2+2\mu\lambda+\lambda^2} - \frac{\lambda^2(s_2e^{s_1t} - s_1e^{s_2t})}{s_1s_2(s_1-s_2)}$	$s_1 = -\frac{1}{2}\left[(2\lambda+3\mu) + \sqrt{\mu^2+4\mu\lambda}\right]$ $s_2 = -\frac{1}{2}\left[(2\lambda+3\mu) - \sqrt{\mu^2+4\mu\lambda}\right]$	$\frac{2\mu^2+2\mu\lambda}{2\mu^2+2\mu\lambda+\lambda^2}$	0.999
	Parallel	Single	$A(t) = \frac{\mu^2+2\mu\lambda}{\mu^2+2\mu\lambda+2\lambda^2} - \frac{2\lambda^2(s_2e^{s_1t} - s_1e^{s_2t})}{s_1s_2(s_1-s_2)}$	$s_1 = -\frac{1}{2}\left[(3\lambda+2\mu) + \sqrt{\lambda^2+4\mu\lambda}\right]$ $s_2 = -\frac{1}{2}\left[(3\lambda+2\mu) - \sqrt{\lambda^2+4\mu\lambda}\right]$	$\frac{\mu^2+2\mu\lambda}{\mu^2+2\mu\lambda+2\lambda^2}$	0.996
		Multiple	$A(t) = \frac{\mu^2+2\mu\lambda}{\mu^2+2\mu\lambda+\lambda^2} - \frac{2\lambda^2(s_2e^{s_1t} - s_1e^{s_2t})}{s_1s_2(s_1-s_2)}$	$s_1 = -2(\mu+\lambda)$ $s_2 = -(\mu+\lambda)$	$\frac{\mu^2+2\mu\lambda}{\mu^2+2\mu\lambda+\lambda^2}$	0.998
3	Standby	Single	$A(t) = \frac{\mu^3+\mu^2\lambda+\mu\lambda^2}{\mu^3+\mu^2\lambda+\mu\lambda^2+\lambda^3} + \frac{\lambda^3[s_2s_3(s_2-s_3)e^{s_1t} - s_1s_3(s_1-s_3)e^{s_2t} + s_1s_2(s_1-s_2)e^{s_3t}]}{s_1s_2s_3(s_1-s_2)(s_1-s_3)(s_2-s_3)}$	s_1, s_2, and s_3 correspond to the three roots of $s^3 + s^2(3\lambda+3\mu) + s(3\lambda^2+4\mu\lambda+3\mu^2) + (\lambda^3+\mu\lambda^2+\lambda\mu^2+\mu^3)$	$\frac{\mu^3+\mu^2\lambda+\lambda^2\mu}{\mu^3+\mu^2\lambda+\lambda^2\mu+\lambda^3}$	0.9999
		Multiple	$A(t) = \frac{6\mu^3+6\mu^2\lambda+3\mu\lambda^2}{6\mu^3+6\mu^2\lambda+3\mu\lambda^2+\lambda^3} + \frac{\lambda^3[s_2s_3(s_2-s_3)e^{s_1t} - s_1s_3(s_1-s_3)e^{s_2t} + s_1s_2(s_1-s_2)e^{s_3t}]}{s_1s_2s_3(s_1-s_2)(s_1-s_3)(s_2-s_3)}$	s_1, s_2, and s_3 correspond to the three roots of $s^3 + s^2(3\lambda+3\mu) + s(3\lambda^2+9\mu\lambda+11\mu^2) + (\lambda^3+3\mu\lambda^2+6\mu^2\lambda+6\mu^3)$	$\frac{6\mu^3+6\mu^2\lambda+3\mu\lambda^2}{6\mu^3+6\mu^2\lambda+3\mu\lambda^2+\lambda^3}$	0.99998
	Parallel	Single	$A(t) = \frac{\mu^3+3\mu^2\lambda+6\mu\lambda^2}{\mu^3+3\mu^2\lambda+6\mu\lambda^2+6\lambda^3} + \frac{6\lambda^3[s_2s_3(s_2-s_3)e^{s_1t} - s_1s_3(s_1-s_3)e^{s_2t} + s_1s_2(s_1-s_2)e^{s_3t}]}{s_1s_2s_3(s_1-s_2)(s_1-s_3)(s_2-s_3)}$	s_1, s_2, and s_3 correspond to the three roots of $s^3 + s^2(6\lambda+3\mu) + s(11\lambda^2+9\mu\lambda+3\mu^2) + (6\lambda^3+6\mu\lambda^2+3\mu^2+\mu^3)$	$\frac{\mu^3+3\mu^2\lambda+6\mu\lambda^2}{\mu^3+3\mu^2\lambda+6\mu\lambda^2+6\lambda^3}$	0.9993
		Multiple	$A(t) = \frac{\mu^3+3\mu^2\lambda+3\mu\lambda^2}{(\mu+\lambda)^3} + \frac{6\lambda^3[s_2s_3(s_2-s_3)e^{s_1t} - s_1s_3(s_1-s_3)e^{s_2t} + s_1s_2(s_1-s_2)e^{s_3t}]}{s_1s_2s_3(s_1-s_2)(s_1-s_3)(s_2-s_3)}$	s_1, s_2, and s_3 correspond to the three roots of $s^3 + s^2(6\lambda+6\mu) + s[11(\mu+\lambda)^2] + 6(\mu+\lambda)^3$	$\frac{\mu^3+3\mu^2\lambda+3\mu\lambda^2}{\mu^3+3\mu^2\lambda+3\mu\lambda^2+\lambda^3}$	0.9999

NOTES:

1. A(t) is the probability of a system being available at time t: A(t) is a function of μ and λ the repair and failure rates. For all functions, the probability of a system being available at time zero is unity. The units of μ and λ must be the same as for t.
2. Instantaneous availability. The probability that the system will be available at any instant in time.
3. Mission availability. Expected availability for a given mission period. This value can be derived from the general model by computing the average value of A(t) for the mission period. Mathematically, this is $A = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} A(t)dt$. Usually t_1 is considered as zero.
4. Steady state availability. The portion of up-time expected for continuous operation.

where:
m = 1, for the multiple repairs case

and

m = n, for the single repair case, or

$$A\,(1/n) = \frac{\sum_{i=1}^{n} \frac{1}{i\lambda}}{\sum_{i=1}^{n} \frac{1}{i\lambda} + \frac{m}{\mu}} \qquad (10.56)$$

In the case of an n-unit ***standby*** system with one active and n-1 standby units

$$\text{MTTF} = \frac{n}{\lambda} \qquad (10.57)$$

and

$$\text{MTTR} = \frac{m}{\lambda} \qquad (10.58)$$

where:
m = 1, for the multiple repairs case

and

m = n, for the single repair case.

Then

$$A = \frac{n/\lambda}{n/\lambda + m/\lambda} \qquad (10.59)$$

Following are some examples utilizing the concepts presented in this section.

Example 7:

In the case of a 2-unit parallel system with λ = 0.01 fr/hr and μ = 1.0 rp/hr, if the system does not undergo repairs until both units fail, the system's steady-state availability is by Eq. (10.56).

$$A[1/2] = \frac{\sum_{n=1}^{2} \frac{1}{n\lambda}}{\sum_{n=1}^{2} \frac{1}{n\lambda} + \frac{m}{\mu}}$$

With single repair (Case 1)

$$A[1/2] = \frac{\frac{1}{\lambda} + \frac{1}{2\lambda}}{\frac{1}{\lambda} + \frac{1}{2\lambda} + \frac{2}{\mu}} = \frac{\frac{1}{0.01} + \frac{1}{2(0.01)}}{\frac{1}{0.01} + \frac{1}{2(0.01)} + 2} = 150/152 = 0.9868$$

With multiple repairs (Case 2)

$$A(1/2) = \frac{\frac{1}{\lambda} + \frac{1}{2\lambda}}{\frac{1}{\lambda} + \frac{1}{2\lambda} + \frac{1}{\mu}} \quad \text{or} \quad A(1/2) = \frac{\frac{1}{0.01} + \frac{1}{2(0.01)}}{\frac{1}{0.01} + \frac{1}{2(0.01)} + \frac{1}{1}}$$

$$A(1/2) = 0.9934$$

If repairs are initiated each time a unit fails, with multiple repairs when both units fail (Case 3) then from Table 10.4-1.

$$A(1/2) = \frac{\mu^2 + 2\lambda\mu}{\mu^2 + 2\mu\lambda + \lambda^2} \quad \text{or} \quad A(1/2) = \frac{(1)^2 + 2(0.01)(1)}{(1)^2 + 2(1)(0.01) + (0.01)^2}$$

and

$$A(1/2) = 0.9999$$

Looking at the three cases of this example

	Availability	Average Downtime in 10,000 hours
Case 1	0.9868	132 hrs.
Case 2	0.9934	66 hrs.
Case 3	0.9999	1 hr.

We can see that the maintenance philosophy plays a significant role. For example, Cases 1 and 2 may not be acceptable for a crucial system such as a ballistic missile early warning system.

<u>Example 8</u>:

We have three redundant equipments, each with an availability of 0.9. What is the availability of the configuration if two of the three equipments must be available at anytime?

(a) From Eq. (10.45)

$$A^3 + 3A^2U + 3AU^2 + U^3 = 1$$

$$A^3 + 3A^2U = (0.9)^3 + 3\,(0.9)^2(0.1)$$

$$= 0.729 + 0.243 = 0.972$$

(b) From Eq. (10.52)

$$A(\,2/3\,) = \frac{3!}{(3-2)!\,2!}\,(0.9)^2(0.1)^{3-2} + \frac{3!}{(3-3)!3!}\,(0.9)^3(0.1)^{3-3}$$

$$= 3(0.9)^2(0.1) + (0.9)^3 = 0.972$$

Example 9:

Given three standby equipments with multiple repair capability, the MTBF of each equipment is 1000 hours and the repair rate is 0.02/hr. What is the expected steady state availability (A_{ss})?

From Table 10.4-1, we see that the appropriate formula is

$$A_{ss} = \frac{6\mu^3 + 6\mu^2\lambda + 3\mu\lambda^2}{6\mu^3 + 6\mu^2\lambda + 3\mu\lambda^2 + \lambda^3}$$

$$\lambda = 1/1000 = 0.001/\text{hr}$$

$$\mu = 0.02/\text{hr}$$

Substituting these values

$$A_{ss} = \frac{6(0.02)^3 + 6(0.02)^2 (0.001) + 3(0.02)(0.001)^2}{6(0.02)^3 + 6(0.02)^2 (0.001) + 3(0.02)(0.001)^2 + (0.001)^3}$$

$$= \frac{6(0.000008) + 6(0.0004)(0.001) + (0.06)(0.000001)}{6(0.000008) + 6(0.0004)(0.001) + (0.06)(0.000001) + (0.001)^3}$$

$$= \frac{0.000048000 + 0.00000240 + 0.00000006}{0.000048000 + 0.00000240 + 0.000000060 + 000000001}$$

$$= \frac{5.046 \times 10^{-5}}{5.0461 \times 10^{-5}} = 0.99998$$

Example 10:

Given two standby equipments in an early warning ground radar system. The equipments are operated in parallel and have a single repair capability. The MTBF of each equipment is 100 hours and the repair rate is 2/hr. What is the expected steady state availability?

From Table 10.4-1, the appropriate equation is:

$$A_{ss} = \frac{\mu^2 + \mu\lambda}{\mu^2 + \mu\lambda + \lambda^2} = \frac{(2)^2 + 2(0.01)}{(2)^2 + 2(0.01) + (0.01)^2} = \frac{4.02}{4.0201} = 0.999975$$

Example 11:

Let us return to the example of the previous section, Figure 10.4-4, in which we had a series system consisting of five subsystems with the following R&M parameters:

Subsystem	λ	μ	A (previously calculated)
1	0.01	0.5	0.98039
2	0.005	1	0.99502
3	0.04	0.2	0.83333
4	0.02	0.2	0.90909
5	0.0025	0.5	0.99502

It was previously found that the availability of this system was $\prod_{i=1}^{5} A_i = 0.73534$

Suppose that we would like to raise the system availability to 0.95 by using redundant parallel subsystems with multiple repair for subsystems 3 and 4 (the two with lowest availability). How many redundant subsystems would we need for each subsystem?

We have the situation

$$A_1 \bullet A_2 \bullet A_3 \bullet A_4 \bullet A_5 = 0.95$$

$$A_3 \bullet A_4 = \frac{0.95}{A_1 A_2 A_5} = \frac{0.95}{(0.98039)(0.99502)(0.99502)} = \frac{0.95}{0.97065} \approx 0.98$$

This means that the product of the improved availabilities ($A_3 A_4$) of subsystems 3 and 4 must be approximately 0.98. As a first cut, let us assume equal availability for improved subsystems 3 and 4. This means that each must have an availability of 0.99 for their product to be 0.98.

Eq. (10.51) is the general expression for improvement in availability through redundancy

$$A(1/n) = 1 - (1 - A)^n$$

where A(1/n) is the improved availability with n redundant units. Let us call this A'. Then,

$$A' = 1 - (1 - A)^n$$

and

$$1 - A' = (1 - A)^n$$

Taking the natural logarithm of both sides of the equation

$$\ln (1 - A') = n \ln (1 - A)$$

$$n = \frac{\ln(1-A')}{\ln(1-A)} \tag{10.60}$$

which is a general expression that can be used to determine the number of redundant subsystems required to achieve a desired subsystem availability (A').

Let us look at improved subsystem 3:

$$A' = 0.99$$

$$A = 0.83333$$

$$n = \frac{\ln(1-0.99)}{\ln(1-0.8333)} = \frac{\ln(0.01)}{\ln(0.16667)} = \frac{-4.605}{-1.79}$$

$= 2.57$, which is rounded up to 3 redundant subsystems (total).

Similarly for subsystem 4:

$$n = \frac{\ln(1-0.99)}{\ln(1-0.90909)} = \frac{\ln(0.01)}{\ln(0.09091)} = \frac{-4.605}{-2.397}$$

$= 1.92$, which is rounded up to 2 redundant subsystems

Thus, in order for the system availability to be raised to 0.95, we need 3 parallel redundant Subsystems 3, and 2 parallel redundant Subsystems 4.

Note that we have not discussed the optimum allocation of failure and repair rates to achieve a given availability; this will be done later in this section.

10.4.1.5 Model E - R&M Parameters Not Defined in Terms of Time

A very different situation in availability modeling is encountered when system "uptime" is not measured in hours of operation or any time parameter but rather in terms of number of rounds fired, miles traveled, actuations or cycles performed, etc. The reliability parameter is then no longer expressed in terms of MTBF but rather in mean-rounds-between-failures (MRBF), mean-miles-between-failures (MMBF), mean-cycles-between-failures (MCBF), etc. The failure rate then also is expressed in number of failures per round, per mile, or per cycle rather than number of failures per operating hour.

For straightforward reliability calculations this poses no problem since the same reliability equations apply as in the time domain, except that the variable time, t, in hours is replaced by the variable number of rounds, number of miles, etc. We may then calculate the reliability of such systems for one, ten, one hundred, or any number of rounds fired or miles traveled, as we wish. The maintainability calculations remain as before, since downtime will always be measured in terms of time and the parameter of main interest remains the MTTR.

However, when it comes to availability, which usually combines two time parameters (i.e., the MTBF and the MTTR into a probability of the system being up at some time, t), a difficult problem arises when the time, t, is replaced by rounds or miles, since the correlation between time and rounds or time and miles is quite variable.

An equation for the steady-state availability of machine guns is given in Reference [11]. This equation is based on a mission profile that at discrete times, t_1, t_2, t_3, etc., requires the firing of N_1, N_2, N_3, etc., bursts of rounds. When the gun fails during a firing, for example at time t, it fires only f rounds instead of N_3 rounds and must undergo repair during which time it is not available to fire; for example, it fails to fire a required N_4 rounds at t_4, and a further N_5 rounds at t_5 before becoming again available (see Figure 10.4-5). Its system availability, A, based on the rounds not fired during repair may be expressed, for the described history, as:

$$A = (N_1 + N_2 + f)/(N_1 + N_2 + N_3 + N_4 + N_5) \tag{10.61}$$

Each sequence of rounds fired followed by rounds missed (not fired) constitutes a renewal process in terms of rounds fired, as shown in Figure 10.4-6, where the gun fails after firing x rounds, fails to fire $\gamma(x)$ rounds in the burst of rounds during which it failed and also misses firing the required bursts of rounds while in repair for an MTTR = M. Assume that the requirements for firing bursts of rounds arrives at random according to a Poisson process with rate r and the average number of rounds per burst is N, then the limiting availability of the gun may be expressed as:

$$A = MRBF/(MRBF + N + \gamma MN) \tag{10.62}$$

where MRBF is the mean number of rounds between failure. The derivation of this formula, developed by R.E. Barlow, is contained in the Appendix of Reference [11]. To calculate A from Eq. (10.62) one must know the MRBF and MTTR of the gun, the average rounds N fired per burst, and the rate γ at which requirements for firing bursts of rounds arrive.

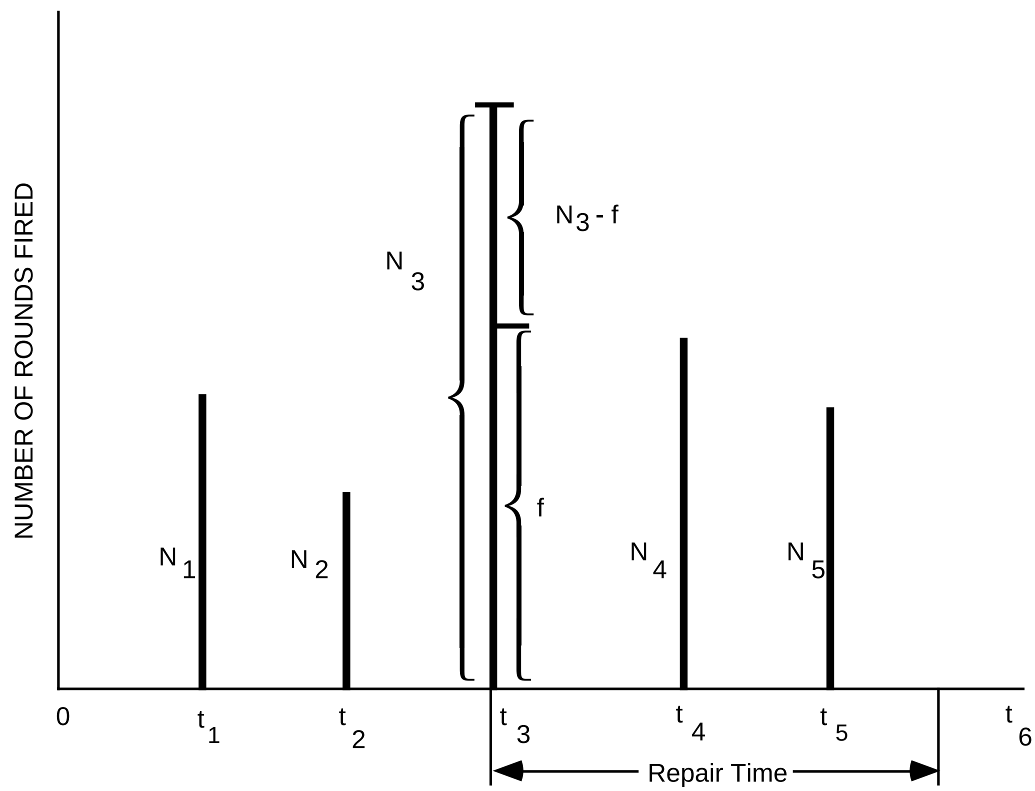


FIGURE 10.4-5: HYPOTHETICAL HISTORY OF MACHINE GUN USAGE

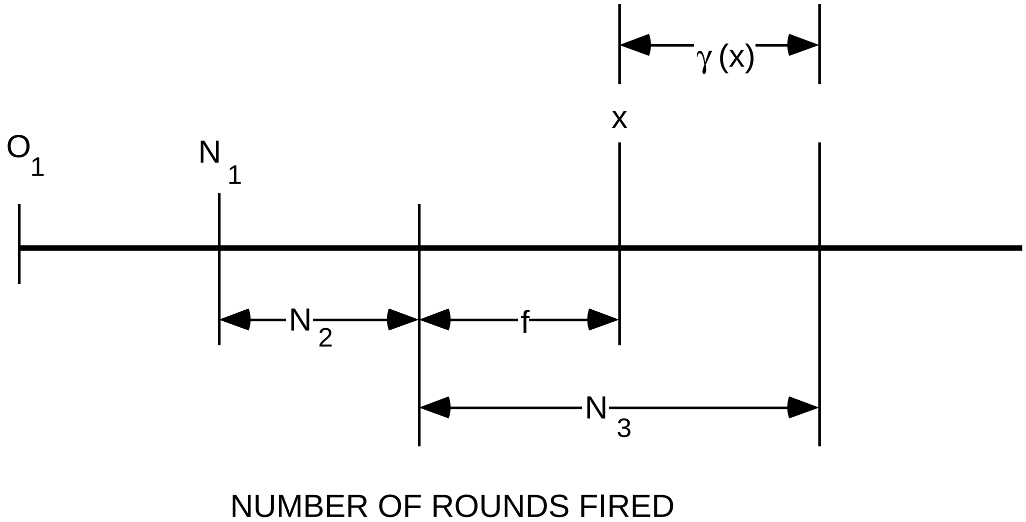


FIGURE 10.4-6: RENEWAL PROCESS IN TERMS OF ROUNDS FIRED

Similar availability equations can be developed for other types of weapons and also for vehicles where the renewal process is in terms of miles traveled. Other approaches to calculating the availability of guns as well as vehicles are found in Reference [12] and are based on calculating from historical field data the maintenance ratios and, via regression analysis, the maintenance time ratios (called the "maintenance clock hour index") that are in turn used in the conventional time based equation of inherent, achieved, and operational availability.

For example, consider a machine gun system in a tank on which historical data are available, showing that 0.014 corrective maintenance man-hours are expended per round fired and that per year 4800 rounds are fired while the vehicle travels for 240 hr per yr. The maintenance ratio (MR) for the gun system is then computed as (Ref. [12], pp. 36-38).

$$MR_{Gun} = \frac{\text{MMH}}{\text{Round}} \bullet \frac{\text{Number of Rounds Fired per Annum}}{\text{Vehicle Operating Hours per Annum}}$$

$$= 0.014 \bullet (4800/240) = 0.28 \qquad (10.63)$$

The dimensions for 0.28 are gun system maintenance man-hours per vehicle operating hour. According to this example, the corrective maintenance time ratio, α (sometimes called the maintenance clock hour index, Ω), is, given by:

$$\alpha_{Gun} = 0.628(0.28)^{0.952} = 0.187 \qquad (10.64)$$

SECTION 10: SYSTEMS RELIABILITY ENGINEERING

The numbers 0.628 and 0.952 are the intercept and the regression coefficients, respectively, obtained by regression analysis as developed in Reference [12], p. 18, Table 1. The dimension for α_{Gun} is gun system downtime per vehicle operating hour. The inherent availability of the gun system is then, according to the conventional time equation, Eq. (10.20).

$$A_i = (1 + \alpha Gun)^{-1} = (1.187)^{-1} = 0.842 \qquad (10.65)$$

This may be interpreted as the gun system being available for 84.2% of the vehicle operating time. Caution is required in using this approach for weapon availability calculations, since in the case where the vehicle would have to be stationary and the gun would still fire rounds, MR and would become infinitely large and the inherent availability of the gun system would become zero.

10.4.2 Mission Reliability and Dependability Models

Although availability is a simple and appealing concept at first glance, it is a point concept, i.e., it refers to the probability of a system being operable at a random point in time. However, the ability of the system to continue to perform reliably for the duration of the desired operating (mission) period is often more significant. Operation over the desired period of time depends, then, on clearly defining system operating profiles. If the system has a number of operating modes, then the operating profile for each mode can be considered.

The term mission reliability has been used to denote the system reliability requirement for a particular interval of time. Thus, if the system has a constant failure rate region so that its reliability R can be expressed as:

$$R = \exp(\lambda t) \qquad (10.66)$$

where:

λ = failure rate = 1/MTBF
t = time for mission

then mission reliability R_M for a mission duration of T is expressed as:

$$R_M = \exp(\lambda T) \qquad (10.67)$$

This reliability assessment, however, is conditional upon the system being operable at the beginning of its mission or its (point) availability.

In order to combine these two concepts, a simplified system effectiveness model may be used where the system effectiveness may be construed simply as the product of the probabilities that the system is operationally ready and that it is mission reliable.

If A is the mean availability of a system at any point in time t_o when we want to use the system and if R_M is the system reliability during mission time T, then system effectiveness E, not including performance, may be defined as:

$$E = AR_M \tag{10.68}$$

Thus, A is a weighting factor, and E represents an assessment of system ability to operate without failure during a randomly chosen mission period.

One concept of dependability used by the Navy (Ref. [13]) takes into account the fact that for some systems a failure which occurs during an operating period t_1 may be acceptable if the failure can be corrected in a time t_2 and the system continues to complete its mission. According to this concept, dependability may be represented by:

$$D = R_M + (1 - R_M)M_o \tag{10.69}$$

where:

- D = system dependability - or the probability that the mission will be successfully completed within the mission time t_1, providing a downtime per failure not exceeding a given time t_2 will not adversely affect the overall mission
- R_M = mission reliability - or the probability that the system will operate without failure for the mission time t_1
- M_o = operational maintainability - or the probability that when a failure occurs, it will be repaired in a time not exceeding the allowable downtime t_2
- t_2 = specified period of time within which the system must be returned to operation

For this model, the exponential approximation of the lognormal maintainability function is used, or

$$M_o = \left(1 - e^{-\mu t_2}\right) \tag{10.70}$$

Then, the system effectiveness is:

$$E = AD = A\left[R_M + (1 - R_M)\,M_o\right] \tag{10.71}$$

In the case where no maintenance is allowed during the mission ($t_2 = 0$ or $M_o = 0$), as in the case of a missile, then this reduces to Eq. (10.68).

$$E = AD = AR_M$$

This concept of dependability is compatible with the WSEIAC model and indeed can be taken into account in the dependability state transition matrices.

Let us examine an airborne system with the following parameters and requirements:

λ = 0.028 failures/hr

μ = 1 repair/hr

Mission time (T) = 8 hours

t_a = 30 minutes to repair a failure during a mission

Thus,

$$A = \frac{\mu}{\mu + \lambda} = \frac{1}{1 + 0.028} = .973 \text{ at the start of the mission}$$

$$R_M = e^{-\lambda T} = e^{-(0.028)(8)} = 0.8 \text{ (mission reliability)}$$

$$M_o = 1 - e^{-\mu t_a} = 1 - e^{-(1)(0.5)} = 0.4 \quad \text{(probability of repairing failure during mission within } \tfrac{1}{2} \text{ hour)}$$

$$\therefore E = A\left[R_M + (1 - R_M)\, M_o\right]$$

$$= 0.973\left[0.8 + (1 - 0.8)(0.4)\right]$$

$$= 0.973\left[0.8 + 0.08\right] = 0.86$$

10.4.3 Operational Readiness Models

Availability, defined as the uptime ratio, is not always a sufficient measure to describe the ability of a system to be committed to a mission at any arbitrary time. In many practical military operations, the concept of operational readiness serves this purpose better. We here define operational readiness as the probability that a system is in an operable condition, i.e., ready to be committed to perform a mission when demands for its use arise. The difference as well as the similarity between availability and operational readiness will become clear by comparing the models developed subsequently with the availability models discussed in the preceding section.

In the development of operational readiness models, one has to consider the usage and the maintenance of the system, i.e., its operating, idle, and repair times. When a call arrives for the system to engage in a mission, the system at such time may be in a state of perfect repair and ready to operate immediately. But it may also be in need of maintenance and not ready. Its state when called upon to operate depends on the preceding usage of the system, i.e., on its preceding mission, in what condition it returned from that mission, and how much time has elapsed since it completed the last mission. Many models can be developed for specific cases, and some are discussed in the following paragraphs.

10.4.3.1 Model A - Based Upon Probability of Failure During Previous Mission and Probability of Repair Before Next Mission Demand

In this model, the assumption is made that if no failures needing repair occurred in the preceding mission, the system is immediately ready to be used again; and, if such failures did occur, the system will be ready for the next mission only if its maintenance time is shorter than the time by which the demand for its use arises. The operational readiness P_{OR} may then be expressed as:

$$P_{OR} = R(t) + Q(t) \bullet P(t_m < t_d) \tag{10.72}$$

where:

$R(t)$	=	probability of no failures in the preceding mission
$Q(t)$	=	probability of one or more failures in the preceding mission
t	=	mission duration
$P(t_m < t_d)$	=	probability that if failures occur, the system maintenance time, t_m, is shorter than the time, t_d, at which the next demand or call for mission engagement arrives

The calculations of R(t) and Q(t) = 1 - R(t) are comparatively simple using standard reliability equations; however, all possible types of failures that need fixing upon return in order to restore in full the system reliability and combat capability must be considered, including any failures in redundant configurations.

As for $P(t_m < t_d)$, one needs to know the probability distributions of the system maintenance time and of call arrivals. Denoting by $f(t_m)$ the probability density function of maintenance time and by $g(t_d)$, the probability density function of time to the arrival of the next call, counted from the instant the system returned from the preceding mission in a state requiring repair, the probability that the system will be restored to its full operational capability before the next call arrives is:

$$P(t_m < t_d) = \int_{t_m}^{\infty} f(t_m) \left[\int_{t_d = t_m}^{\infty} g(t_d) dt_d \right] dt_m \tag{10.73}$$

The integral in the square brackets on the right side of the equation is the probability that the call arrives at t_d after a variable time t_m. When this is multiplied by the density function $f(t_m)$ of the duration of maintenance times and integrated over all possible values of t_m, we get $P(t_m < t_d)$.

Now assume that maintenance time t_m and time to next call arrival t_d are exponentially distributed, with M_1 being the mean time to maintain the system and M_2 the mean time to next call arrival. The probability density functions are thus:

$$f(t_m) = \left[\exp(-t_m/M_1)\right]/M_1 \tag{10.74}$$

$$f(t_d) = \left[\exp(-t_d/M_2)\right]/M_2 \tag{10.75}$$

We then obtain

$$P(t_m < t_d) = \int_0^{\infty} M_1^{-1} \exp(-t_m/M_1) \bullet \left[\int_{t_m}^{\infty} M_2^{-1} \exp(-t_d/M_2) dt_d \right] dt_m$$

$$= \int_0^{\infty} (-M_1^{-1}) \exp\left[-(1/M_1 + 1/M_2) t_m\right] dt_m$$

$$= M_2/(M_1 + M_2) \tag{10.76}$$

In this exponential case, system operational readiness becomes

$$P_{OR} = R(t) + Q(t) \bullet \left[M_2/(M_1 + M_2) \right] \tag{10.77}$$

As a numerical example let us look at a system with a probability of R = 0.8 of returning from a mission of t = 1 hr duration without requiring repair and therefore had a probability of Q = 0.2 that it will require repair. If system mean maintenance time is M_1 = 1 hr and the mean time to next call arrival is M_2 = 2 hr., the operational readiness of the system becomes

$$P = 0.8 + 0.2\,(2/3) = 0.933$$

Comparing this result with the conventional steady-state availability concept and assuming that the system has a mean maintenance time of M_1 = 1 hr and a mean time to failure of M_2 = 5 hr (roughly corresponding to the exponential case of R = 0.8 for a one-hour mission), we obtain a system availability of:

$$A = M_2/(M_1 + M_2) = 5/6 = 0.833$$

which is a result quite different from P_{OR} = 0.933.

10.4.3.2 Model B - Same As Model A Except Mission Duration Time, t is Probabilistic

The operational readiness model of Eq. (10.72) can be extended to the case when mission duration time t is not the same for each mission but is distributed with a density q(t). We then get

$$P_{OR} = \int_0^{\infty} R(t)q(t)dt + P(t_m < t_d) \int_0^{\infty} Q(t)q(t)dt \qquad (10.78)$$

Since the integrals in Eq. (10.78) are fixed numbers, we may write:

$$R = \int_0^{\infty} R(t)q(t)dt \text{ , and}$$

$$Q = \int_0^{\infty} Q(t)q(t)dt \qquad (10.79)$$

and using the symbol P for $P(t_m < t_d)$, i.e., $P = P(t_m < t_d)$, Eq. (10.78) may be written in the form:

$$P_{OR} = R + QP \qquad (10.80)$$

In this equation R is the probability that the system returns without failures from the last mission;

Q = 1 - R is the probability that one or more failures developed in the last mission; and P is the probability that the system will be repaired before the next call arrives if it developed failures. The mission times are variable here with density q(t).

10.4.3.3 Model C - Similar To Model A But Includes Checkout Equipment Detectability

The operational readiness of the system at time t_a is given by:

$$P_{OR}(t_a) = R(t_m) + \left[kM(t_r)\right] \bullet \left[1 - R(t_m)\right] \tag{10.81}$$

where:

$P_{OR}(t_a)$ = probability of system being available for turnaround time, e.g., t_a of 30 minutes, following completion of preceding mission or initial receipt of alert

$R(t_m)$ = probability that the system will survive the specified mission of duration t_m without failure

t_r = specified turnaround time, or maximum downtime for repair required of the system

k = probability that if a system failure occurs it will be detected during the mission or during system checkout following the mission

$M(t_r)$ = probability that a detected system failure can be repaired in time t_r to restore *the* system to operational status

Thus, when mission reliability, mission duration, availability, and turnaround time are specified for the system, the detectability-times-maintainability function for the system is constrained to pass through or exceed the point given by:

$$kM(t_r) \geq \frac{P_{OR}(t_a) - R(t_m)}{\left[1 - R(t_m)\right]}$$

Consider, for example, the following specified operational characteristics for a new weapons system:

Mission Reliability, $R(t_m)$ = 0.80 for t_m of 8 hours

Operational Readiness $P_{OR}(t_a)$ = 0.95 for turnaround time, t_a of 30 minutes, following completion of preceding mission or initial receipt of alert.

From the requirements, the required detectability-maintainability product (kM) is derived as follows:

$$kM(30) = \frac{P_{OR}(30) - R(8)}{1 - R(8)} = \frac{0.95 - 0.8}{1 - 0.8} = 0.75$$

Therefore, kM(30) = 0.75 is the joint probability, given that a system failure has occurred, that the failure will be detected (either during the mission or during post mission checkout) and will be repaired within 30 minutes following completion of the mission.

Assume that k is to be 0.9, i.e., built-in test equipment is to be incorporated to detect at least 90% of the system failures and provide go/no-go failure indication.

Then, the maintainability requirement is:

$$M(30) = \frac{0.75}{k} = \frac{0.75}{0.9} \approx 0.83$$

which means that 83% of all system repair actions detected during the mission or during post mission checkout must be completed within 30 minutes.

Using the exponential approximation, maintainability as a function of repair time is expressed as the probability of repair in time t_r:

$$M(t_r) = 1 - e^{-\mu t_r} = 1 - e^{-t_r / \overline{M}_{ct}} \qquad (10.82)$$

where:

$\overline{M}_{ct}$ = MTTR

μ = repair rate, $1/\overline{M}_{ct}$

t_r = repair time for which M(t) is to be estimated

The required mean time to repair ($\overline{M}_{ct}$) is found from Eq. (10.82) by taking the natural log of both sides:

$$\overline{M}_{ct} = - \frac{t_r}{ln\left[1 - M(t_r)\right]}$$

Substituting t_r = 30 minutes, and M(t_r), which we previously found to be 0.83,

$$\overline{M}_{ct} = -\frac{30}{ln\,(0.17)} = \frac{-30}{-1.77} \approx 17 \text{ minutes}$$

And from M(t_{max}) = 0.95 we find the maximum time for repair of 95% of detected system failures ($M_{max_{ct}}$) as follows:

$$M(t_{max}) = 0.95 = 1 - e^{-M_{max_{ct}}/\overline{M}_{ct}}$$

$$M_{max_{ct}} = -\overline{M}_{ct}\ ln\,(1 - 0.95)$$

$$= -(17)(-3) = 51 \text{ minutes}$$

Thus, these requirements could be established as design requirements in a system development specification.

Detectability Factor, k = 0.90

Mean Time To Repair, $\overline{M}_{ct}$ = 17 minutes

Maximum Time To Repair, $M_{max_{ct}}$ = 51 minutes

10.4.3.4 Model D - For a Population of N Systems

Let N be the total population of systems, e.g., squadron of aircraft. The service facility itself shall be considered as having k channels, each servicing systems at a mean rate μ. The analysis is based on an assumed Poisson distribution of arrivals and on a mean service time which is assumed to be exponentially distributed. This service is performed on a first come, first served basis.

The basic equations (derived in Ref. [11]) are as follows:

$$p_n = \frac{N!}{(N-n)!}\left(\frac{\rho^k}{k!}\right)\left(\frac{\rho}{k}\right)^{n-k} p_o \quad \text{when } n > k \tag{10.83}$$

$$p_n = \frac{N!}{(N-n)!}\frac{\rho^n}{n!}\, p_o \quad \text{when } n \le k \tag{10.84}$$

$$p_o = \left[\sum_{n=k+1}^{k} \frac{N!}{(N-n)!} \frac{\rho^n}{n!} + \sum_{n=0}^{k} \frac{N!}{(N-n)!} \left(\frac{\rho^k}{k!} \right) \left(\frac{\rho}{k} \right)^{n-k} \right]^{-1} \quad (10.85)$$

$$\rho = \frac{\lambda}{\mu} = \frac{\text{Mean arrival rate (failure)}}{\text{Mean service rate}} \quad (10.86)$$

where:

p_i = probability of i units awaiting service
k = number of repair channels or facilities
N = total number of systems

$$P_{OR} = \frac{N - \overline{n}}{N} \quad (10.87)$$

where:

P_{OR} = probability that a system is neither awaiting nor undergoing service.
$\overline{n}$ = average number of systems either awaiting or undergoing service at a given time and is defined by:

$$\overline{n} = \sum_{n=0}^{N} n p_n \quad (10.88)$$

The specific procedure, which will be illustrated by an example, is as follows:

Step 1: Use Eq. (10.85) to solve for p_o

Step 2: Use p_o from Step 1 to help derive p_n for all values of $n \leq k$ by use of Eq. (10.84)

Step 3: Use p_o from Step 1 to help derive p_n for all values of $n > k$ by use of Eq. (10.83)

Step 4: For all values of n, from 0 through N, sum the terms np_n derived from Steps 1 through 3. This, per Eq. (10.88) gives $\overline{n}$, the average number of systems not ready

Step 5: Use Step 4 results and Eq. (10.87) to obtain the operational readiness probability, P_{OR}

SECTION 10: SYSTEMS RELIABILITY ENGINEERING

Example 12: P_{OR} of Interceptor Squadron

An interceptor squadron contains fifteen planes (N = 15) and has available four flight line repair channels (k = 4). Each plane averages 50 operating hours per month out of 24 times 30, or 720 total available hours. Because of five-day, two-shift maintenance each failure requires an average of five clock hours (MTTR) to repair. The plane MTBF is 3.47 operating hours between failures. What is the operational readiness probability for this squadron?

We first compute the utilization factor ρ

$$r = \frac{1}{\rho} \cdot \frac{\text{Operating hours per plane per month}}{\text{Total hours per month}}$$

$$= \frac{(5)(50)}{(3.47)(720)} = \frac{250}{2500} = 0.1$$

Step 1: Use Equation (10.85) to obtain p_o

$$p_o = \left[\sum_{n=k+1}^{N} \frac{N!}{(N-n)!} \frac{r^n}{n!} + \sum_{m=k}^{N} \frac{N!}{(N-n)!} \frac{r^k}{k!} \frac{r^{n-k}}{k} \right]^{-1}$$

$$p_o = \left[\sum_{m=0}^{4} \frac{15!}{(15-n)!} \frac{(0.1)^n}{n!} + \sum_{m=4}^{15} \frac{15!}{(15-n)!} \frac{(0.1)^4}{4!} \frac{(0.1)^{n-4}}{4} \right]^{-1}$$

The calculated results are shown in the following table:

n	Term (1)	Term (2)
0	1.0	- -
1	1.5	- -
2	1.05	- -
3	0.455	- -
4	0.1365	0.03412
5	- -	0.0375
6	- -	0.03753
7	- -	0.0337
8	- -	0.0127
9	- -	0.0189
10	- -	0.0113
11	- -	0.0056
12	- -	0.0022
13	- -	0.0006
14	- -	0.00013
15	- -	0.00000
Sum	4.1415	0.19428

$$p_o = (4.1415 + 0.19428)^{-1} = (4.3358)^{-1} = 0.2306$$

Step 2: Use Equation (10.84) and obtain p_n for n = 1 through 4.

$$P_n = \frac{N!}{(N - n)!} \frac{\rho^n}{n!} p_o$$

Thus,

$$p_1 = \frac{15!}{(15 - 1)!} \frac{(0.1)^1}{1!} (0.23) = 0.3450$$

$$p_2 = \frac{15!}{13!} \frac{(0.1)^2}{2!} (0.23) = 0.2415$$

$$p_3 = \frac{15!}{12!} \frac{(0.1)^3}{3!} (0.23) = 0.10465$$

$$p_4 = \frac{15!}{11!} \frac{(0.1)^4}{4!} (0.23) = 0.0313$$

Step 3: Use Equation (10.83) and obtain p_n for n = 5 through 15.

$$p_n = \frac{N!}{(N - n)!} \left(\frac{\rho^k}{k!}\right) \left(\frac{\rho}{k}\right)^{n-k} p_o$$

Thus,

$$p_5 = \frac{15!}{10!} \left[\frac{(0.1)^1}{4!}\right] \left(\frac{0.1}{4}\right)^1 (0.23) = 0.0086$$

$$p_6 = \frac{15!}{9!} \left(\frac{0.1^4}{4!}\right) \left(\frac{0.1}{4}\right)^2 (0.23) = 0.00214$$

Similarly,

$$p_7 = 0.000486$$

$$p_8 = 0.000097$$

$$p_9 = 0.000017$$

p_{10} through p_{15} are negligible probabilities.

Step 4: Sum the terms np_n for n = 0 through n = 15.

n	p_n	np_n
0	0.2300	0
1	0.3450	0.3450
2	0.2415	0.4830
3	0.1047	0.314100
4	0.0313	0.12500
5	0.0086	0.043000
6	0.00214	0.012850
7	0.000486	0.003400
8	0.000097	0.000776
9	0.000017	0.000153
10	---	---
11	---	---
12	---	---
13	---	---
14	---	---
15	---	---
Total		1.214779

Therefore from Equation (10.88):

$$\overline{n} = \sum_{n=0}^{N} np_n$$

$$= 1.215 \text{ planes which are not ready on the average}$$

Step 5: Using Step 4 results and Equation (5.87), we obtain P_{OR}, the operational readiness probability

$$P_{OR} = \frac{N - \overline{n}}{N} = \frac{15 - 1.215}{15} = \frac{13.785}{15} = 0.919$$

As can be seen, the calculations are rather laborious and best done by a computer. Figures 10.4-7 and 10.4-8 (from Ref. [10]) are a series of curves for N = 15 and N = 20 with k values ranging from 1 to 10 and 1 to 20, respectively. Note that 0.919 checks out the r = 0.1, k = 4 point on Figure 10.4-7.

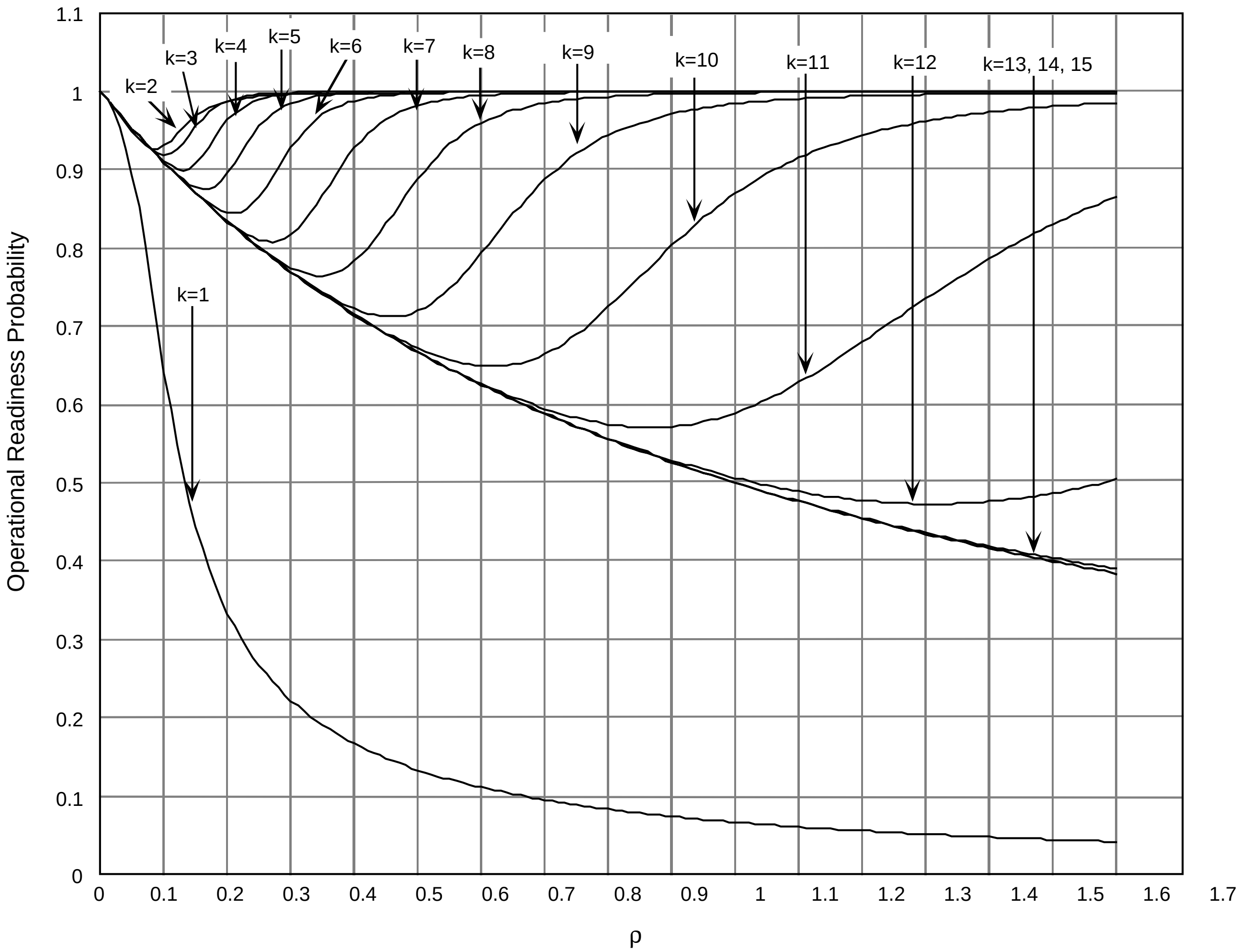


FIGURE 10.4-7: OPERATIONAL READINESS PROBABILITY VERSUS QUEUING FACTOR ρ. FOR POPULATION SIZE N = 15; NUMBER OF REPAIR CHANNELS k

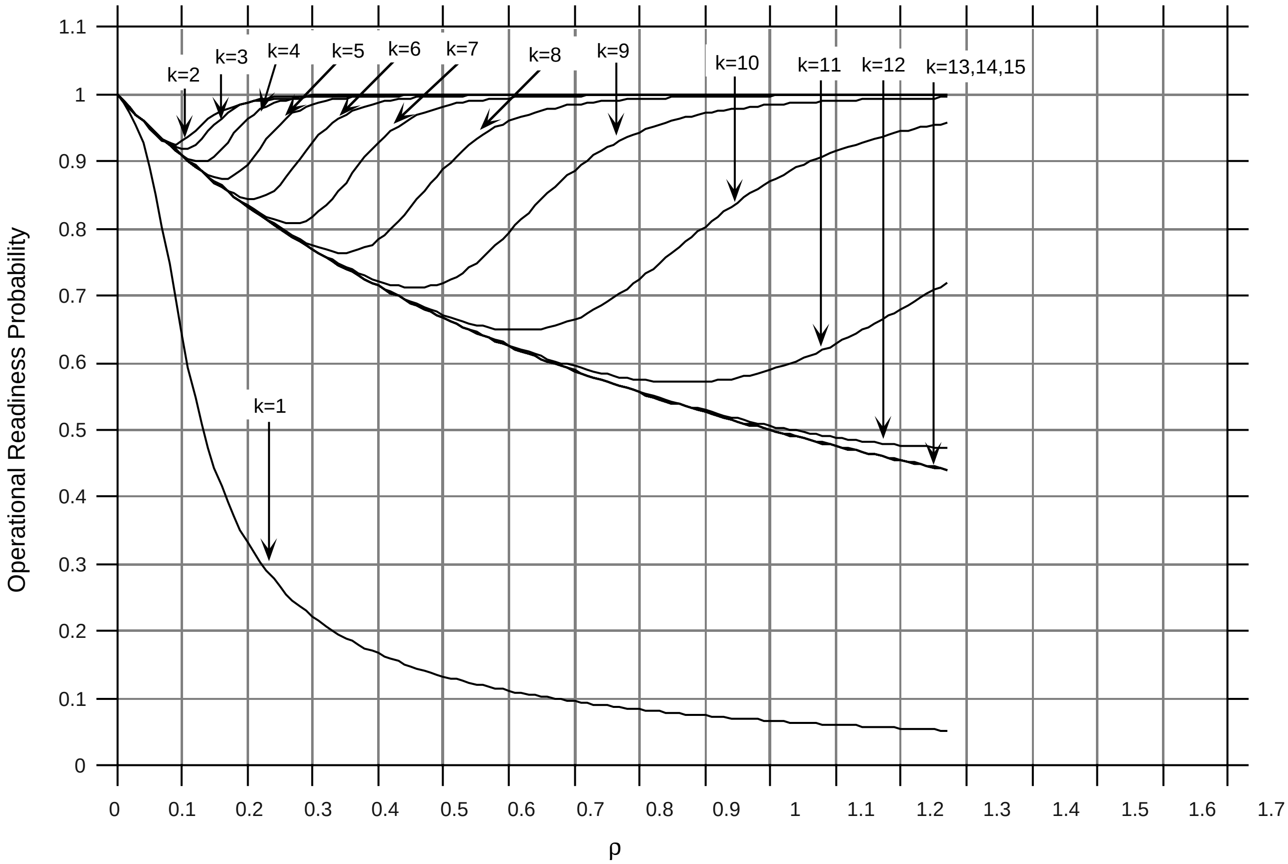


FIGURE 10.4-8: OPERATIONAL READINESS PROBABILITY VERSUS QUEUING FACTOR ρ. FOR POPULATION SIZE N = 20; NUMBER OF REPAIR CHANNELS k

10.5 Complex Models

In summing up the discussion of models, it should be recognized that there may be other measures of system R&M parameters or system effectiveness than those previously discussed. For example, in cases such as manned aircraft models it might be meaningful to combine operational readiness and equipment availability into one index, or we may wish to combine detection probability and availability for a ground radar system to be an index of the probability that a raid launched at any random time will be detected. The important point in selecting an index of system reliability effectiveness is recognizing that it is equivalent to a correct statement of the problem.

When selecting an index of effectiveness we should keep in mind some characteristics without which the index would be of little value. Probably the most important characteristic is that the index be expressed quantitatively. We should be able to reduce it to a number such that comparisons between alternative designs can be made. Furthermore, the index we choose must have a basis in physical reality. Thus it should be descriptive of the real problem, not exaggerated or oversimplified. Yet at the same time the index should be simple enough to allow for mathematical manipulation to permit evaluating alternatives.

In complex system effectiveness mathematical models, an attempt is made to relate the impact of system reliability, maintainability, and performance to the mission profiles, scenario, use, and logistic support. Only in simple situations can a meaningful single model be developed that will relate all these parameters and yield a single quantitative measure of system effectiveness. Numerous complex computerized models exist and, as a matter of fact, every major company in the aerospace business has developed a multitude of such models.

10.6 Trade-off Techniques

10.6.1 General

A trade-off is a rational selection among alternatives in order to optimize some system parameter that is a function of two or more variables which are being compared (traded off). Examples of system trade-offs involve performance, reliability, maintainability, cost, schedule, and risk. A trade-off may be quantitative or qualitative. Insofar as possible, it is desirable that trade-offs be based on quantifiable, analytic, or empirical relationships. Where this is not possible, then semi-quantitative methods using ordinal rankings or weighting factors are often used.

The methodology for structuring and performing trade-off analyses is part of the system engineering process described in Section 4. The basic steps, summarized here are:

(1) Define the trade-off problem and establish the trade-off criteria and constraints

(2) Synthesize alternative design configurations

(3) Analyze these alternative configurations

(4) Evaluate the results of the analyses with respect to the criteria, eliminating those which violate constraint boundaries

(5) Select the alternative which best meets criteria and constraint boundaries or iterate the design alternatives, repeating Steps 2 through 5 to obtain improved solutions

System effectiveness and cost effectiveness models provide the best tools for performing trade-off studies on the system level. Through the computerized models, any changes in any of the multitude of reliability, maintainability, performance, mission profile, logistic support, and other parameters can be immediately evaluated as to their effect on the effectiveness and total cost of a system. Thus, cost effectiveness modeling and evaluation, besides being used for selecting a specific system design approach from among several competing alternatives, is a very powerful tool for performing parametric sensitivity studies and trade-offs down to component level when optimizing designs to provide the most effective system for a given budgetary and life cycle cost constraint or the least costly system for a desired effectiveness level.

At times, however, especially in the case of the more simple systems, trade-offs may be limited to achieving a required system availability while meeting the specified reliability and maintainability requirements. Comparatively simple trade-off techniques can then be used as shown in the following paragraphs.

10.6.2 Reliability - Availability - Maintainability Trade-offs

The reliability-maintainability-availability relationship provides a measure of system effectiveness within which considerable trade-off potential usually exists, e.g., between reliability, maintainability, and logistic support factors. This potential should be re-evaluated at each successive stage of system development to optimize the balance between reliability, maintainability, and other system effectiveness parameters with respect to technical risks, life cycle cost, acquisition schedule, and operating and maintenance requirements. The latter become increasingly more important as complexity of system design increases, dictating the need for integration of system monitoring and checkout provisions in the basic design.

As stated earlier in this section and in Section 2, reliability and maintainability jointly determine the inherent availability of a system. Thus, when an availability requirement is specified, there is a distinct possibility of trading-off between reliability and maintainability, since in the steady state availability depends only on the ratio or ratios of MTTR/MTBF which was previously referred to as maintenance time ratio (MTR) α, i.e.,

$$\alpha = \text{MTTR/MTBF} = \lambda/\mu \qquad (10.88)$$

so that the inherent availability equation assumed the form

$$A_i = 1/(1+\alpha) \qquad (10.89)$$

As an example, consider systems I and II with

$$\text{MTTR}_I = 0.1 \text{ hr.}$$

$$\text{MTBF}_I = 2 \text{ hr.}$$

$MTTR_{II}$ = 10 hr.

$MTBF_{II}$ = 200 hr.

Then the steady state availability is

$$A_I = 1/\left[1 + (0.1/2)\right] = 0.952$$

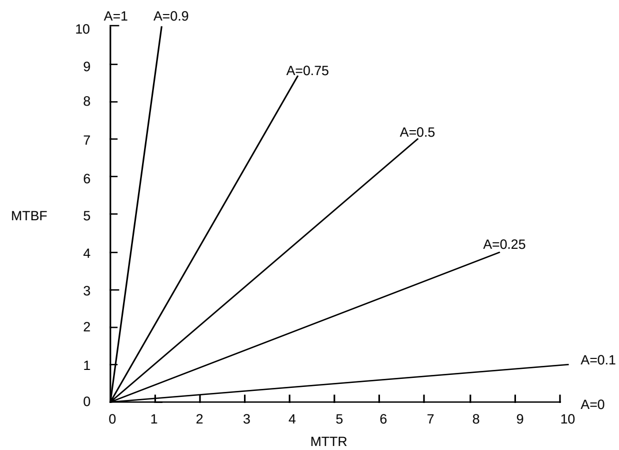
$$A_{II} = 1/\left[1 + (10/200)\right] = 0.952$$

Both systems have the same availability, but they are not equally desirable. A 10-hr MTTR might be too long for some systems, whereas a 2-hr MTBF might be too short for some systems.

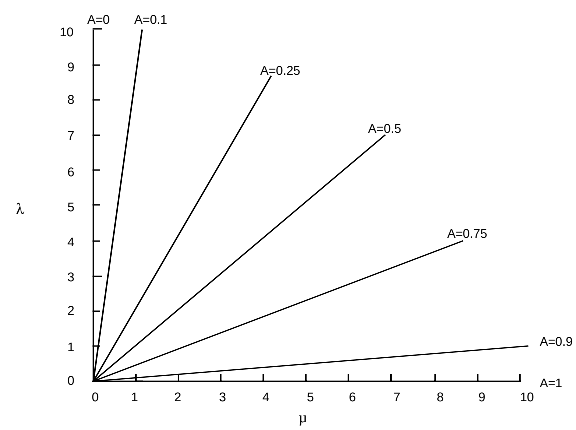
Even though reliability and maintainability individually can be increased or decreased in combinations giving the same system availability, care must be taken to insure that reliability does not fall below its specified minimum or that individually acceptable values of reliability and maintainability are not combined to produce an unacceptable level of system availability.

A generalized plot of Eq. (10.88) is given in Figure 10.6-1. A plot of A vs. λ/μ, is given in Figure 10.6-2. These equations and graphs show that in order to optimize availability it is desirable to make the ratio of MTBF/MTTR as high as possible.

Since increasing MTBF and decreasing MTTR is desirable, the equation for availability can be plotted in terms of MTBF and 1/MTTR (or μ) as shown in Figure 10.6-3. Each of the curves representing the same availability in Figure 10.6-3 just as each of the lines in Figure 10.6-1, is called isoavailability contours; corresponding values of MTBF and MTTR give the same value of A, all other things being equal. Availability nomographs useful for reliability and maintainability trade-offs are given in Reference [13]. Figure 10.6-4 is an example of an availability nomograph.



(A) AVAILABILITY AS A FUNCTION OF MTBF AND MTTR



(B) AVAILABILITY AS A FUNCTION OF λ AND μ

FIGURE 10.6-1: RELIABILITY - MAINTAINABILITY - AVAILABILITY RELATIONSHIPS

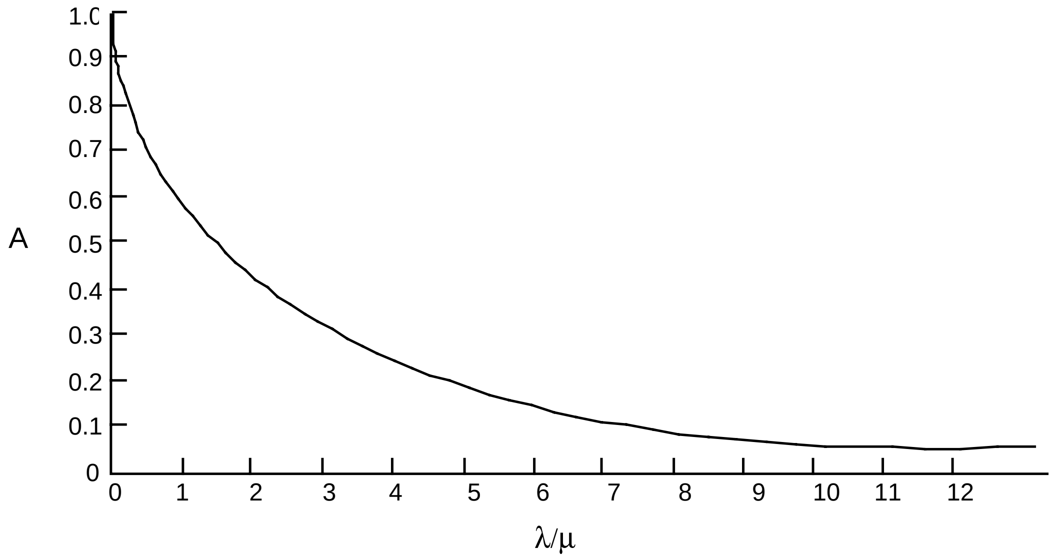


FIGURE 10.6-2: AVAILABILITY AS A FUNCTION OF λ/μ

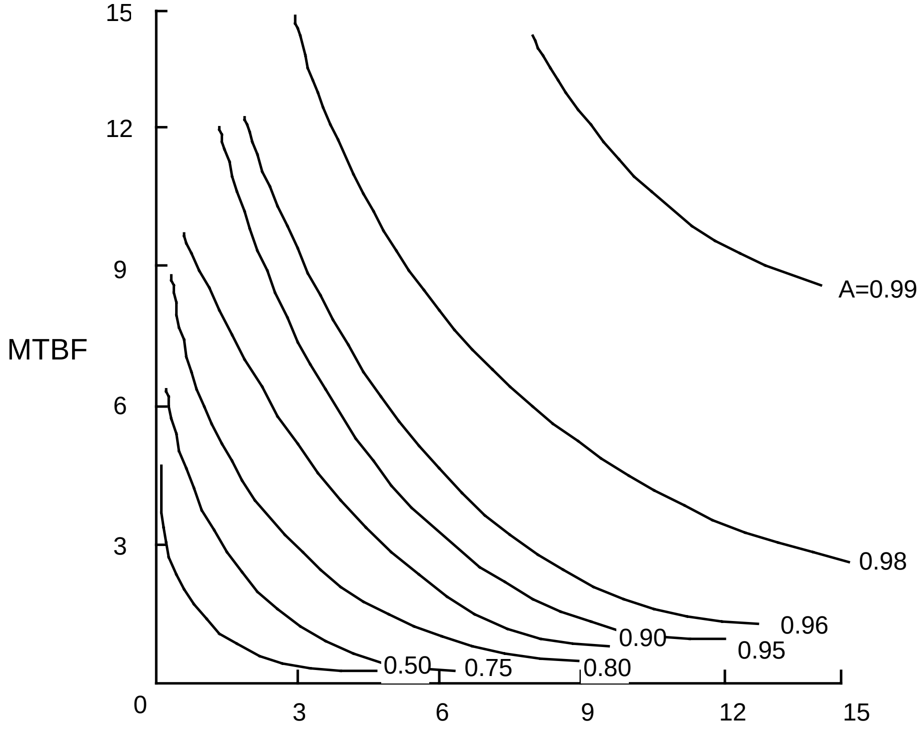


FIGURE 10.6-3: AVAILABILITY AS A FUNCTION OF MTBF AND 1/MTTR

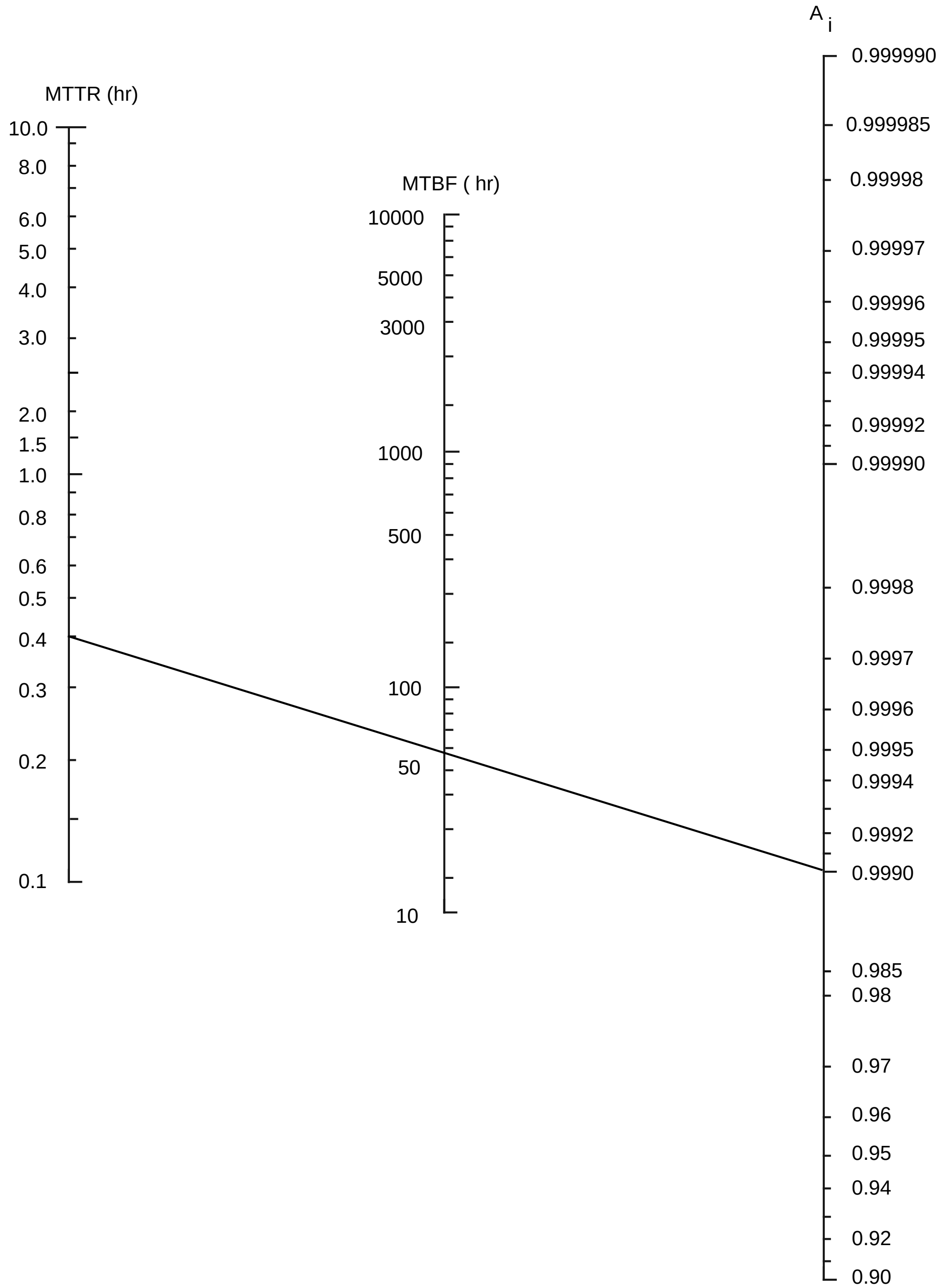


FIGURE 10.6-4: AVAILABILITY NOMOGRAPH

There are obvious practical limits which must be considered in trade-off optimization. These are called constraints, and all purposeful optimization must be bounded by constraints into feasible regions. For example, there are practical limits as to how high a value for MTBF can be achieved or how low MTTR can be made. In the one case, the reliability of system components or the required redundancy might be so high that the desired reliability could not be realistically achieved within the state-of-the-art or would be so expensive as to violate cost constraints. Similarly, MTTRs close to zero would require extreme maintainability design features, such as completely built-in test features or automatic test and checkout to allow fault isolation to each individual replaceable module, with perhaps automatic switchover from a failed item to a standby item. This also could easily violate state-of-the-art or cost constraints.

It follows, then, that trade-offs not only involve relationships among system parameters and variables but also that they are bounded by both technical and economic constraints. In a sense, all trade-offs are economic ones, requiring cost-benefit analysis (not necessarily in terms of dollar costs but rather in terms of the availability and consumption of resources, of which dollars are often the most convenient measure). Resource constraints may also include manpower and skill levels, schedule or time availability, and the technical state-of-the-art capability. Later sections of this chapter deal with the cost problem.

There are two general classes of trade-offs. In the first, the contributing system variables are traded-off against one another without increasing the value of the higher level system parameter; for example, trading-off reliability and maintainability along an isoavailability contour (no change in availability). This might be done for reasons of standardization or safety or for operational reasons such as the level at which the system and its equipments will be maintained. The other class of trade-off is one in which the system variables are varied in order to obtain the highest value of the related system parameters within cost or other constraints. For example, reliability and maintainability might be traded-off in order to achieve a higher availability. This could result in moving from one isoavailability curve to another in Figure 10.6-3, perhaps along an isocline (a line connecting equal slopes).

An example of a reliability - availability - maintainability (RAM) trade-off is given in the following paragraphs. The design problem is as follows: A requirement exists to design a radar receiver which will meet an inherent availability of 0.99, a minimum MTBF of 200 hours, and an MTTR not to exceed 4 hours. The current design is predicted to have an availability of 0.97, an MTBF of 150 hours, and an MTTR of 4.64 hours.

Using Eq. (10.88) the area within which the allowable trade-off may be made is shown by the cross-hatched portion of Figure 10.6-5. As indicated in the previous paragraph, there are two approaches which can be used for trade-off. One is to fix the availability at 0.990. This means that any combination of MTBF and MTTR between the two allowable end points on the 0.990 isoavailability line may be chosen. These lie between an MTBF of 200 hours with an MTTR of 2 hours and an MTBF of 400 hours with an MTTR of 4 hours. The other approach is to allow availability to be larger than 0.990 and thus allow any combination of MTBF and MTTR within the feasible region.

It is clearly seen that without any additional constraints the designer has a limitless number of combinations from which to choose. Assume that the following four alternative design configurations have been selected for trade-off as shown in Table 10.6-1.

Design Configuration Nos. 1, 2, and 3 all have the required availability of 0.990. Design Configuration No. 1 emphasizes the maintainability aspects in design, while Design Configuration No. 3 stresses reliability improvement. Design Configuration No. 2 is between Nos. 1 and 3 for the same availability. Design Configuration No. 4 is a combination of Nos. 1 and 2 and yields a higher availability.

Since all of these alternatives are within the feasible region shown in Figure 10.6-5 some other criterion must be used for selection of the desired configuration. In this case, we will use the least cost alternative or the one showing the greatest life cycle cost savings over the present configuration as the basis for trade-off decision. An example cost comparison of the different alternatives is shown in Table 10.6-2 (such costs would be estimated using various cost and economic models).

The cost table shows that Configuration No. 2 is the lowest cost alternative among those with equal availabilities. It also shows that Configuration No. 4, with a higher acquisition cost, has a significantly better 10-year life cycle support cost and lowest overall cost, as well as a higher availability. Thus Configuration No. 4 is the optimum trade-off, containing both improved reliability and maintainability features.

The trade-off example previously shown was a relatively simple example for the case of a single equipment. Let us now look at a more complex example. Figure 10.6-6 (a repeat of Figure 10.4-4) represents a serial system consisting of five statistically independent subsystems, each with the indicated $MTBF_i$ and $MTTR_i$. We found by the use of Eq. (10.27) that the steady state availability was:

$$A = \prod_{i=1}^{5} A_i = 0.73534$$

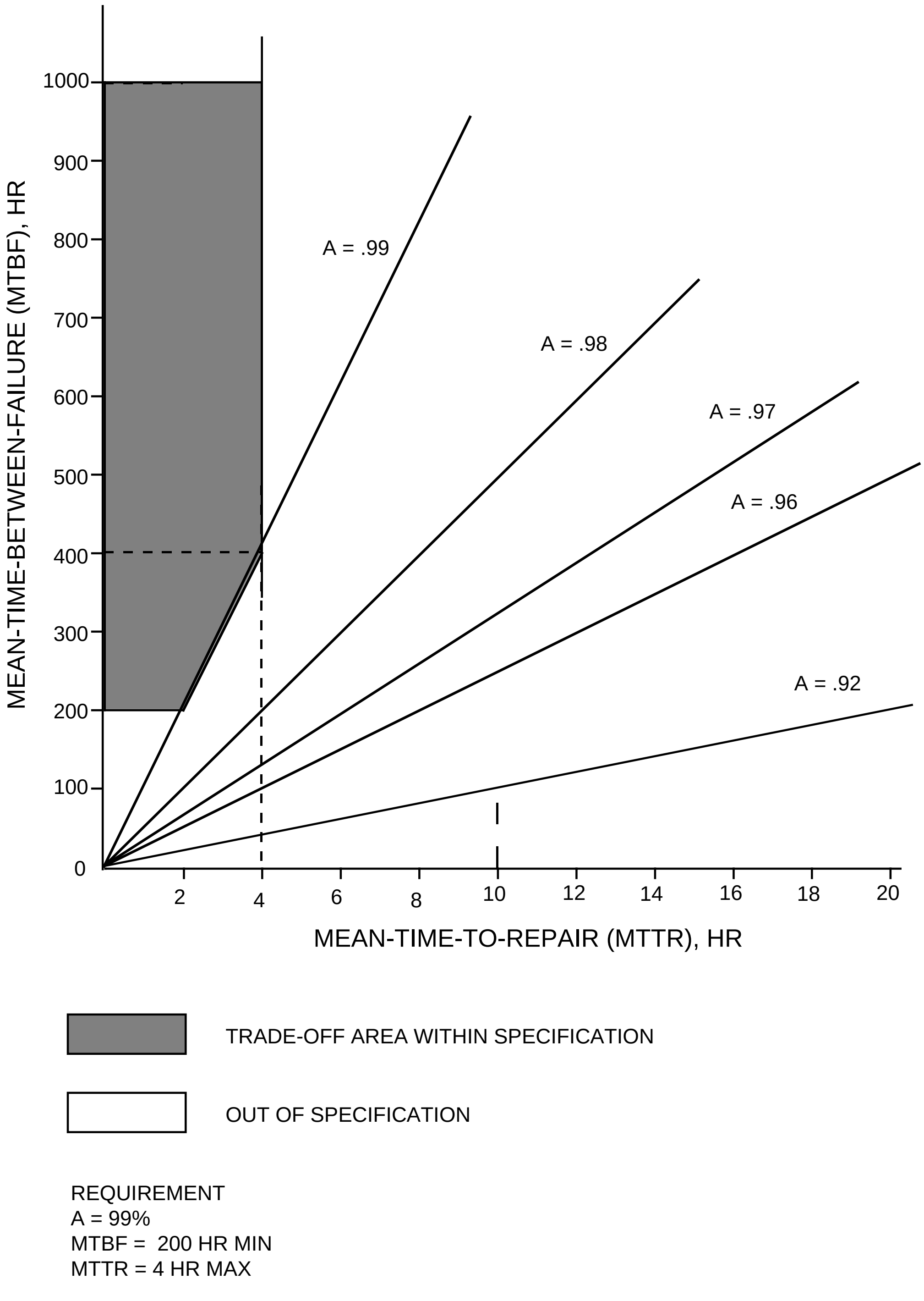


FIGURE 10.6-5: RELIABILITY-MAINTAINABILITY TRADE-OFFS

TABLE 10.6-1: ALTERNATIVE DESIGN TRADE-OFF CONFIGURATIONS

Design Configuration	A	MTBF, hr	MTTR, hr
1. R - derating of military standard parts M - modularization and automatic testing	0.990	200	2.0
2. R - design includes high reliability parts/components M - limited modularization and semi-automatic testing	0.990	300	3.0
3. R - design includes partial redundancy M - manual testing and limited modularization	0.990	350	3.5
4. R - design includes high reliability parts/components M - modularization and automatic testing	0.993	300	2.0

TABLE 10.6-2: COST COMPARISON OF ALTERNATIVE DESIGN CONFIGURATIONS

ITEM	EXISTING	1	2	3	4
Acquisition (Thousands of Dollars)					
RDT&E	300	325	319	322	330
Production	4,500	4,534	4,525	4,530	4,542
Total	4,800	4,859	4,844	4,852	4,872
10-Year Support Costs (Thousands of Dollars)					
Spares	210	151	105	90	105
Repair	1,297	346	382	405	346
Training and Manuals	20	14	16	8	14
Provisioning & Handling	475	525	503	505	503
Total	2,002	1,036	1,006	1,018	968
LIFE CYCLE COST	6,802	5,895	5,850	5,870	5,840

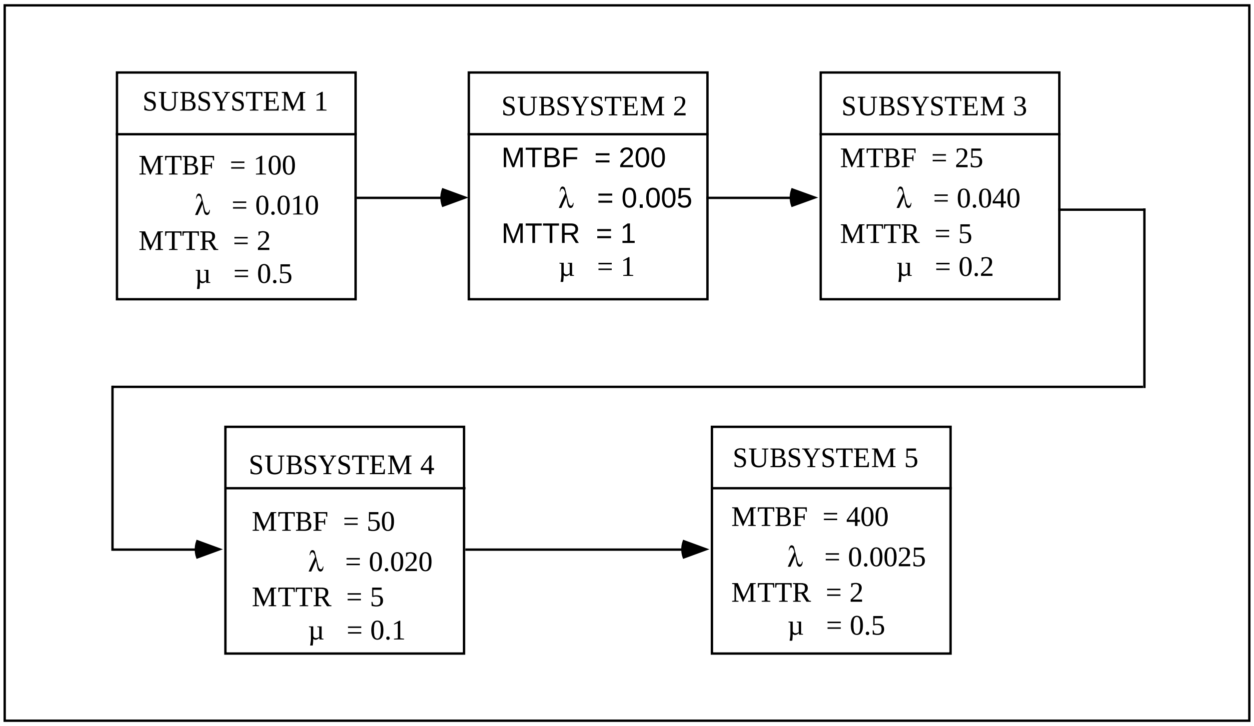


FIGURE 10.6-6: BLOCK DIAGRAM OF A SERIES SYSTEM

By inspection of the maintenance time ratios (MTRs) of each of the subsystems we note that Subsystems 3 and 4 have the lowest MTRs, given by:

$$\frac{MTTR_i}{MTBF_i} = \frac{5}{25} = 0.2$$

for Subsystem 3 and 5/50 = 0.1 for Subsystem 4. These are, therefore, the "culprits" in limiting system availability to 0.73534, which may be unacceptable for the mission at hand. If because of the state-of-the-art limitations we are unable to apply any of the design techniques detailed in Section 7 to reduce MTBF, then our first recourse is to add a parallel redundant subsystem to Subsystem 3, the weakest link in the series chain.

We shall consider two cases: (1) the case where no repair of a failed redundant unit is possible until both redundant subsystems fail and the system stops operating; or (2) repair is possible by a single repair crew while the system is operating.

For case (1) where both units must fail before repair is initiated and a single crew repairs both failed units in sequence:

$$A(1/2) = \frac{\sum_{n=1}^{2} \frac{1}{n\lambda}}{\sum_{n=1}^{2} \frac{1}{n\lambda} + \frac{n}{\mu}} = \frac{\frac{1}{\lambda} + \frac{1}{2\lambda}}{\frac{1}{\lambda} + \frac{1}{2\lambda} + \frac{2}{\mu}} \quad \text{(from Equation 10.56)}$$

$$= \frac{\frac{1}{0.04} + \frac{1}{2(0.04)}}{\frac{1}{0.04} + \frac{1}{2(0.04)} + \frac{2}{0.2}} = \frac{37.5}{47.5} = 0.789$$

This is a lower availability than the nonredundant case!

$$A_{System} = \frac{1}{1 + \frac{MTTR_{Series}}{MTBF_{Series}}} = \frac{1}{1 + .02} = 0.833 \text{ (based on Equation 10.18)}$$

For case (1), where both units must fail before repair is initiated and two repair crews simultaneously repair both failed units:

$$A(1/2) = \frac{\sum_{n=1}^{2} \frac{1}{n\lambda}}{\sum_{n=1}^{2} \frac{1}{n\lambda} + \frac{1}{\mu}} = \frac{\frac{1}{0.04} + \frac{1}{2(0.04)}}{\frac{1}{0.04} + \frac{1}{2(0.04)} + \frac{1}{0.2}} = \frac{37.5}{42.5} = 0.882$$

which is a slight improvement over the nonredundant case.

For case (2), where a single repair crew initiates repair action on a redundant subsystem as soon as it fails

$$A = \frac{\mu^2 + 2\mu\lambda}{\mu^2 + 2\mu\lambda + \lambda^2} \quad \text{(from Table 10.4-1)}$$

$$= \frac{(0.2)^2 + 2(0.2)(0.04)}{(0.2)^2 + 2(0.2)(0.04) + (0.04)^2}$$

$$= \frac{0.04 + 0.016}{0.04 + 0.016 + 0.0016} = \frac{0.056}{0.0576} = 0.972$$

as compared to 0.833 where no redundancy was used.

This corresponds to an increased system availability of

$$A = 0.73534\left(\frac{0.972}{0.833}\right) = 0.86$$

If this new value is still not acceptable redundancy might have to be applied to Subsystem 4. For example, let us use a 2-unit standby configuration for Subsystem 4 with multiple repairs; then (from Table 10.4-1), the steady state availability would be:

$$A = \frac{2\mu^2 + 2\mu\lambda}{2\mu^2 + 2\mu\lambda + \lambda^2} = \frac{2(0.2)^2 + 2(0.2)(0.02)}{2(0.2)^2 + 2(0.2)(0.02) + (0.02)^2}$$

$$= \frac{0.08 + 0.008}{0.08 + 0.008 + 0.0004} = \frac{0.088}{0.0884} = 0.995$$

Thus, the new system availability would be:

$$A = (0.86)\left(\frac{0.995}{0.909}\right) = 0.94$$

where 0.909 was the original availability of Subsystem 4.

Note, however, that to achieve these gains in availability, repair of failed redundant units must be possible while the system is operating.

Before leaving the subject of trade-offs at the system availability level, it should be pointed out that design trade-off methodologies can also be used at lower levels of the system hierarchy to increase MTBF and reduce MTTR. These are discussed in Section 7.

10.7 Allocation of Availability, Failure and Repair Rates

The previous sections discussed how availability can be calculated for various system configurations, e.g., series, parallel, etc., and how R&M can be traded off to achieve a given availability. This section discusses methods for allocating availability (and, where appropriate, failure and repair rates) among the system units to achieve a given system availability.

The reader should keep in mind that we are concerned with systems that are maintained upon failure. For the case of non-maintained systems, e.g., missiles, satellites, etc., the methods presented in Chapter 3 are appropriate for system reliability design, prediction, and allocation.

During the initial design phase of a program, detailed information is not usually available regarding the particular equipments to be employed with the system. For example, we may know that a transmitter with certain power requirements may be designed, but we do not usually know if it is less expensive to design for a low failure rate or a high repair rate. Unless the experience of similar, previously designed transmitters can guide our decisions, estimation of the best set of alternatives is necessary. Having developed a system configuration, a range of values of equipment failure rates and repair rates that would satisfy the system availability requirement can be initially specified. The state-of-the-art limits for these equipments may not be known or the expenditures required for improvement, but we can specify their ratio, which would allow considerable freedom through the design process.

10.7.1 Availability Failure Rate and Repair Rate Allocation for Series Systems

Several cases can be considered:

(1) A single repairman must repair any one of n identical, statistically independent subsystems in series. The ratio of failure rate to repair rate is such that there is a strong possibility that a second subsystem will fail while the first one is being repaired.

(2) Same as (1) except a repairman is assigned to each subsystem and can only work on that particular subsystem.

(3) Same as (1) except some intermediate number of repairmen, r, less than the number of subsystems is assigned. Any repairman can work on any system.

(4) Repeat cases (1)-(3) with nonidentical subsystems.

10.7.1.1 Case (1)

The steady state availability in Case (1) is from Reference [25]:

$$A_S = \frac{(\mu/\lambda)^n}{n!\sum_{i=0}^{n} \frac{(\mu/\lambda)^i}{i!}} \tag{10.90}$$

where:

μ = subsystem repair rate

λ = subsystem failure rate

n = number of subsystems in series

μ/λ = "operability ratio" as opposed to λ/μ (the utilization factor)

SECTION 10: SYSTEMS RELIABILITY ENGINEERING

For example, if n = 4 and A_s = 0.90, the allocation equation becomes

$$0.90 = \frac{(\mu/\lambda)^4}{24\left[1+\frac{\mu}{\lambda}+\frac{1}{2}\left(\frac{\mu}{\lambda}\right)^2+\frac{1}{6}\left(\frac{\mu}{\lambda}\right)^3+\frac{1}{24}\left(\frac{\mu}{\lambda}\right)^4\right]}$$

or $\mu/\lambda = 38.9$

The complexities of allocating failure and repair rates for even simple examples are apparent. If the subsystems are not identical, the allocation must be solved using the state matrix approach to compute availability.

10.7.1.2 Case (2)

This represents the situation in which a repairman is assigned to each subsystem. It is equivalent to the condition in which $\mu/\lambda >> 1$, i.e., failure rate is much smaller than repair rate. Since this is true of many practical systems, a wide variety of practical problems can be solved.

It was previously shown that for this case,

$$A_s = (A_i)^n = \left[\frac{1}{1+(\lambda/\mu)}\right]^n \qquad (10.91)$$

where:

A_i = subsystem availability

n = number of subsystems

From Eq. (10.91)

$$A_i = (A_s)^{1/n} \qquad (10.92)$$

Example 13:

Three identical series subsystems must operate so as to give a total system availability of 0.9. What requirement should be specified for the availability of each subsystem? For the ratio of μ/λ for each subsystem?

$$A_i = (0.9)^{1/3} = 0.965$$

$$0.965 = \frac{1}{1 + \lambda/\mu}$$

$$\lambda/\mu = \frac{1}{0.965} - 1 = 0.036$$

Example 14:

A system consists of three identical, independent subsystems connected in series. The availability requirement is 0.99, and the repair rate is limited to 0.3 per hr. What is the minimum failure rate which must be allocated to each subsystem to satisfy the system requirement? A repairman is assigned exclusively to each subsystem.

If for Case (2) the series equipments are not identical the relationship

$$A_s = \prod_{i=1}^{n} A_i \tag{10.93}$$

can be used to derive the individual subsystem availabilities.

Procedure	Example
(1) State the system availability requirement.	$A_s = 0.99$
(2) Compute the availability of each subsystem by $A_i = (A_s)^{1/n}$	$A_i = (0.99)^{\frac{1}{3}}$ $= 0.99666$
(3) For each subsystem compute the ratio λ/μ by: $\frac{\lambda}{\mu} = \frac{1}{A_i} - 1$	$\lambda/\mu = \frac{1}{0.99666} - 1$ $= 0.00336$
(4) Compute λ from the previous equation with μ = 0.3 per hr. The final answer is rounded off to 2 significant figures to avoid implying too much accuracy.	$\lambda = 0.00336 \times (0.3 \text{ per hr})$ = 1.0 per 1000 hr

SECTION 10: SYSTEMS RELIABILITY ENGINEERING

Example 15: (using Eq. (10.93))

Three subsystems must operate to give a total system availability of 0.9. Subsystem 1 has an availability of 0.99. What should be specified for the availability of each of the other two subsystems if: (1) they are equally important, or (2) Subsystem 3 should have twice the availability of Subsystem 2 (this is interpreted as Subsystem 3 having one-half the unavailability of Subsystem 2).

(1) $A_S = 0.99\, A_2 A_3$

$A_2 = A_3$

$0.9 = 0.99(2)A_2$

$A_2 = \sqrt{0.91}$

$A_2 = A_3 = 0.954$

(2) $(1 - A_2) = 2\,(1 - A_3)$

$1 - A_2 = 2 - 2A_3$

$$A_3 = \frac{A_2 + 1}{2}$$

$$0.9 = 0.99\, A_2 A_3 = 0.99\, A_2 \left(\frac{A_2 + 1}{2}\right) = 0.99\, A_2 \left(\frac{A_2^2}{2} + \frac{A_2}{2}\right)$$

$$2\left(\frac{0.9}{0.99}\right) = A_2^2 + A_2$$

$A_2^2 + A_2 - 1.82 = 0$

$A_2 = 0.94$

$$A_3 = \frac{0.94 + 1}{2} = 0.97$$

The failure and repair rate allocations for A_2 and A_3 would be

$$\lambda_2/\mu_2 = \frac{1}{A_2} - 1 = \frac{1}{0.94} - 1 = 0.064$$

$$\lambda_3/\mu_3 = \frac{1}{A_2} - 1 = \frac{1}{0.97} - 1 = 0.03$$

The previous example can be expanded to use weighting factors to derive the required subsystem availabilities. The choice of weighting factor would depend upon the system being analyzed and the significant parameters affecting availability. Some examples of weighting factors might be relative cost or equivalent complexity of the subsystem. The latter, for example, should correlate somewhat with increasing failure and repair rates. Let us examine an example of an allocation using equivalent complexity.

Example 16:

A ground surveillance series system consists of a radar, a data processor, and display subsystem. A system availability of 0.995 is required. Based upon past experience and engineering analysis, it is estimated that the complexity of each subsystem is as follows:

Display Subsystem	≈	1000 component parts
Radar Subsystem	≈	2500 component parts
Data Processor Subsystem	≈	5500 component parts

What availability requirement should be specified for each of the subsystems to meet the system requirement?

The weight assigned to each subsystem is given by:

$$W_i = \frac{\text{Number of parts for subsystem i}}{\text{Total number of parts in system}}$$

$$W_1 \text{(Display)} = \frac{1000}{1000 + 2500 + 5500} = 0.11$$

$$W_2 \text{ (Radar)} = \frac{2500}{1000 + 2500 + 5500} = 0.28$$

$$W_3 \text{ (Data Processor)} = \frac{5500}{1000 + 2500 + 5500} = 0.61$$

If the system availability requirement is 0.995, then 1 - 0.995 = 0.005 is the unavailability of the system. Using the weights previously derived to apportion the system unavailability to each of the subsystems, we get:

Display	=	(0.11)(0.005)	=	0.00055
Radar	=	(0.28)(0.005)	=	0.00140
Data Processor	=	(0.61)(0.005)	=	0.00305
SYSTEM UNAVAILABILITY			=	0.005

Thus, the required availabilities for each subsystem would be

$$A_1 \text{ (Display)} = 1 - 0.00055 = 0.99945$$

$$A_2 \text{ (Radar)} = 1 - 0.0014 = 0.9986$$

$$A_3 \text{ (Data Processor)} = 1 - 0.00305 = 0.99695$$

Verifying that the system requirement will be met

$$A_s = (0.99945)(0.9986)(0.99695) = 0.995$$

Also, as was previously shown, failure and repair rate allocation can be derived:

$$\lambda_1/\mu_1 = \frac{1}{A_1} - 1 = \frac{1}{0.99945} - 1 = 5.5 \cdot 10^{-4}$$

$$\lambda_2/\mu_2 = \frac{1}{A_2} - 1 = \frac{1}{0.9986} - 1 = 1.4 \cdot 10^{-3}$$

$$\lambda_3/\mu_3 = \frac{1}{A_3} - 1 = \frac{1}{0.99695} - 1 = 3.0 \cdot 10^{-3}$$

Another slight variation of Case (2) (Section 10.7.1.2) is a series system with nonidentical subsystems, in which each subsystem's

$$\lambda_i/\mu_i < 0.1$$

The availability of such a system with subsystems whose failures and repair are statistically independent is:

$$A_s = \frac{1}{1 + \sum_{i=1}^{n} \alpha_i} \tag{10.94}$$

where:

α_i = λ_i/μ_i with all $\alpha_i <$ 0.1

n = number of subsystems in series

$$\alpha_{(system)} = \alpha_1 + \alpha_2 + ... + \alpha_n \qquad (10.95)$$

To design such a system, one merely allocates the subsystem α_i according to some weighting scheme. For example, there may be a requirement to design a new system with higher availability which is similar in design to the old system, where the relative weighting factors are the same for each new subsystem.

$$W_i = \frac{\alpha_i \text{ (new)}}{\alpha_i \text{ (old)}} \qquad (10.96)$$

Example 17:

A system consisting of two statistically independent subsystems has an availability of 0.90. Subsystem 1 has an availability of 0.97, and subsystem 2 has an availability of 0.93. A new system, similar in design to this one, must meet a required 0.95 availability. What are the new subsystem availabilities and ratios of failure-to-repair rate?

Since the allocated ratios are known, additional trade-off studies can be performed to optimize λ_i and μ_i for each subsystem.

10.7.2 Failure and Repair Rate Allocations For Parallel Redundant Systems

A system comprising several stages of redundant subsystems whose λ/μ ratio is less than 0.1 can be treated as if the stages were statistically independent. The system steady-state availability A_s is:

$$A_s = A_1 \bullet A_2 \bullet A_3 \bullet ... \bullet A_n$$

where:

A_i = the availability of State I

The procedure for allocating the failure and repair rates and the availability is as follows.

Procedure	Example
(1) State the availability requirement A_S of the new system	$A_S = 0.95$
(2) Compute the sum α_S of the α = ratios for the old system $\alpha_S(\text{old}) = \alpha_1 + \alpha_2$	(Remember $\alpha_i = \lambda_i/\mu_i = \frac{1}{A_i} - 1$) $\alpha_S(\text{old}) = 0.0309 + 0.0753 = 0.01062$
(3) Compute the relative weights W_i by Eq. (10.96)	$W_1 = \frac{0.0309}{0.1062} = 0.291$ $W_2 = \frac{0.0753}{0.1062} = 0.709$
(4) Compute an overall A_S for the new system by: $\alpha_S'(\text{new}) = \frac{1}{A_S'} - 1$	$\alpha_S' = \frac{1}{0.95} - 1 = 0.0526$
(5) Compute the allocated α_i' for each subsystem of the new design by: $\alpha_i' = W_i\, \alpha_S'$	$\alpha_1 = (0.291)(0.0526) = 0.0153$ $\alpha_2' = (0.709)(0.0526) = 0.0373$
(6) Compute the availabilities A_i' allocated to each subsystem by: $A_i' = \frac{1}{1 + \alpha_i'}$	$A_1' = \frac{1}{1 + 0.0153} = 0.985$ $A_2' = \frac{1}{1 + 0.0373} = 0.964$
(7) Check the allocated availability A_S of the new system by: $A_S' = A_1' \cdot A_2'$	$A_S = (0.985)(0.964) = 0.95$

This is equivalent to treating each stage as if it had a repairman assigned to it. It is also equivalent to saying that a single repairman is assigned to the system but that the probability of a second failure occurring while the first is being repaired is very small. If the stages are not statistically independent, the system availability must be computed by the state matrix approach. In either case, the system requirement can be obtained with a range of failure and repair rates. Trade-off procedures must be used to determine the best set of these parameters.

It will be recalled (from Eq. (10.52)) that the steady-state measure of availability for a stage where at least m out of n equipments must be available for the stage to be available can be expressed by the binomial expansion

$$A_S = \sum_{i=m}^{n} \binom{n}{i} A^i (1 - A)^{n-i} \tag{10.97}$$

and, where m = 1, i.e., only one equipment of n need be available at any one time, Eq. (10.97) simplifies to:

$$A_S = 1 - (1 - A)^{n-1} \tag{10.98}$$

If Eq. (10.97) can be expressed in terms of the operability ratio μ/λ, the initial allocation may be made. Eq. (10.97) can be expressed in terms of the operability ratio as:

$$A_S = \sum_{i=m}^{n} \frac{\binom{n}{m} (\mu/\lambda)^i}{(1+\mu/\lambda)^n} \tag{10.99}$$

Now if a value of A_S is specified and we know the system configuration (at least how many equipments out of n-equipments must be available within each stage), we can solve for the operability ratios μ/λ.

For example, consider Table 10.7-1, in which the system availability requirement of 0.992 has been allocated to each of 4 series subsystems (stages) as indicated in column (2). In turn, in order to achieve the given stage availability, it has been determined that parallel redundant subsystems are required for each stage (column (3)) in which at least one of the redundant subsystems per stage must be available for the system availability requirement to be met.

TABLE 10.7-1: PRELIMINARY SYSTEM AND SUBSYSTEM RELIABILITY SPECIFICATIONS

(1)	(2)	(3)	(4)	(5)
Stage	Stage Availability	Number of Subsystems (n)	Number of Subsystems Required (m)	Operability Ratio
1	0.9984	4	1	4.0
2	0.9976	5	1	2.5
3	0.9984	4	1	4.0
4	0.9976	5	1	2.5

The final column (5) indicates the calculated μ/λ (operability ratio) required of each subsystem in the redundant configuration of each stage in order to achieve the allocated stage availability. Column (5) results are obtained by the use of Eqs. (10.98) or (10.99). For example, for Stage 1, m = 1, n = 4. Therefore, since m = 1, we may use Eq. (10.98).

$$A_S = 1 - (1 - A)^n$$

$$0.9984 = 1 - \left(1 - \frac{\mu}{\lambda + \mu}\right)^4$$

$$0.9984 = 1 - \left(\frac{\lambda}{\lambda + \mu}\right)^4$$

$$\frac{1}{1 + \mu/\lambda} = (1 - 0.9984)^{1/4} = 0.2$$

$$0.2\mu/\lambda = 1 - 0.2$$

$$\frac{\lambda}{\mu} = .25$$

This represents an upper bound of the ratio. All solutions for which the ratio $\leq$ .25 are acceptable.

Obviously, there are a multitude of combinations that would satisfy this equation as shown in Figure 10.7-1. Until more information becomes available concerning the cost of various failure rates and repair rates of the particular equipments involved, this initial specification allows preliminary equipment design to start with an availability goal that is consistent with the system's requirements. To facilitate calculations of operability ratio, solutions to Eq. (10.99) for n from two through five (Ref. [25]) are given in Figures 10.7-2a through 10.7-2d. The abscissa of the graphs is expressed in terms of unavailability since the plot allows for greater linearity, and, thus, ease of reading. Let us solve an example problem utilizing the graphs.

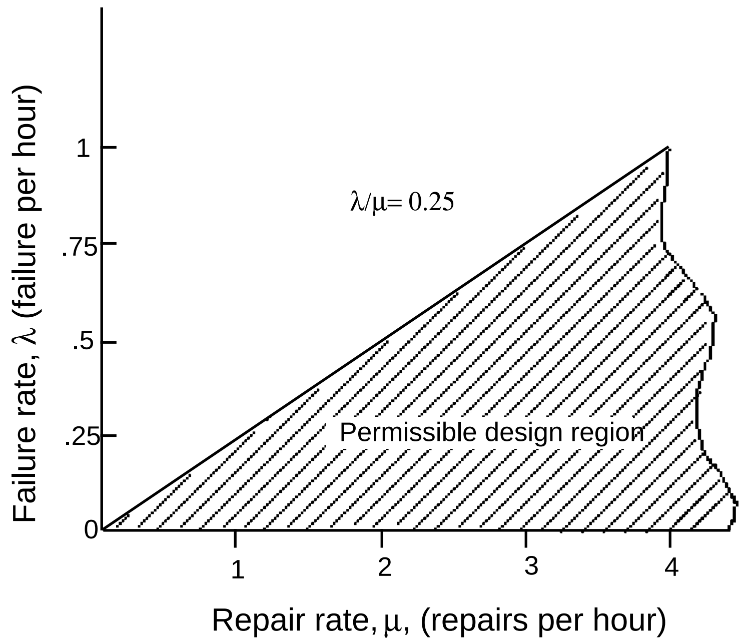


FIGURE 10.7-1: PERMISSIBLE EQUIPMENT FAILURE AND REPAIR RATES FOR λ/μ = .25

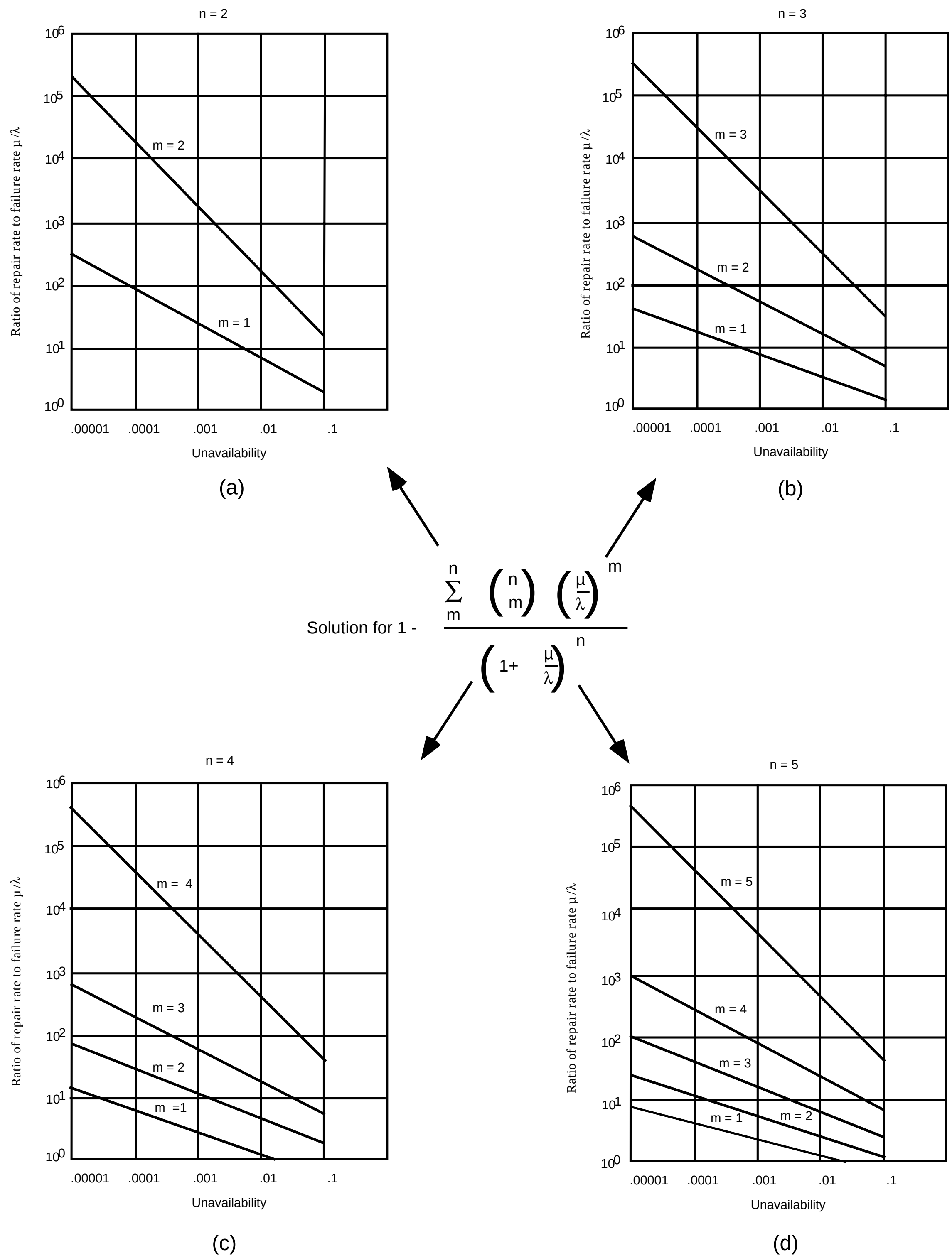


FIGURE 10.7-2: UNAVAILABILITY CURVES

Example 18:

A system consists of five identical, statistically independent subsystems connected in a parallel redundant configuration. A system availability of 0.999 is required. Four out of five subsystems must be operating for the system availability requirement to be met. What is the required λ/μ ratio? The procedure for finding this ratio is as follows.

Procedure
(1) State the system availability requirement, A_s (e.g., A_s = 0.999)
(2) Compute the system unavailability, U_s, by subtracting As from 1 (e.g., U_s = 1-0.999 = 0.0010)
(3) Enter Figure 10.7.2-2d using m = 2 and U_s = 0.0010, and read the required ratio (e.g., λ/μ = .01)

10.7.3 Allocation Under State-of-the-Art Constraints

Following through the example of the previous section, we note that the allocation of an operability ratio λ/μ to each equipment does not recognize limitations on the range of either of these parameters. If R&M predictions indicate what these constraints are and they turn out to be in conflict with the preliminary allocation, revised allocations are warranted. During the reallocation, the cost of reducing the equipment failure rates and repair rates should also be considered to provide for a best balance of operability objectives. For example, in the previous section (see Table 10.7-1) the operability ratio allocated to the subsystems within the first stage was $\lambda/\mu \leq .25$. If reliability predictions indicate that a failure rate of 0.7 failures/hour can be achieved without much difficulty, this would indicate that a repair rate of at least 2.8 repairs/hour is required to meet the specifications. If, however, it is expected that repairs cannot be made at a rate greater than 2.0/hour, the specification will not be met.

As an example, let it be assumed that it is possible to design the equipment so that it can achieve a failure rate of 0.1 failures/hour - however, only at a considerable expenditure over and above that which would be required to design for a failure rate of 0.7 failures/hour. Now, it may be possible that the predicted failure rates and repair rates of the subsystems within the remaining stages are well within the operability ratio. Thus, it may be feasible to tighten the specifications of the subsystems within the other stages while relaxing the specification of the subsystems within the first stage and still achieve the required level of system availability. Again, there may be many ways of balancing the specifications. It is desirable, therefore, to choose that balance which minimizes any additional expenditure involved over that allocated for the system configuration.

Dynamic programming (Ref. [25]) is a powerful tool for balancing operability ratios in determining a system configuration at least cost.

Before leaving this subsection on allocation with redundancy, it should be pointed out that if the redundant subsystems in each stage are not identical, state matrix techniques must be used to compute availability.

10.8 System Reliability Specification, Prediction and Demonstration

Sections 6, 7 and 8 presented in great detail methods for specifying, predicting, and demonstrating system reliability.

The methods and design procedures presented in Section 7 are directly applicable to system reliability parameters for the case of non-maintained systems, e.g., missiles, satellites, “one-shot” devices, etc.

For maintained systems, the methods and procedures presented in References [26] and [50] are directly applicable to system maintainability parameters. When these are combined with the methods of Section 7 and the appropriate sections of this section, they provide a complete capability for specifying, predicting, and demonstrating most system R&M parameters, as well as trading them off to maximize system availability or some other appropriate effectiveness parameter at minimum cost.

Perhaps the only area that may need some further discussion is availability demonstration methods. At the present time no accepted test plans exist for steady state availability; however, MIL-HDBK-781 describes two availability demonstration tests; one for fixed sample size, the other a fixed time test. The tests are based upon a paper presented at the 1979 Annual Reliability and Maintainability Symposium (Ref. [26]). The paper also provides a theoretical discussion of sequential test plans, but no standardized plans are presented. Program managers or R&M engineers who wish to consider using sequential availability tests should consult the referenced paper. The proposed demonstration plans are described in the following subsection.

10.8.1 Availability Demonstration Plans

The availability tests are based on the assumption that a system can be treated as being in one (and only one) of two states, “up” or “down.” At t = 0 the system is up (state X) and operates until the first failure at $T = X_1$; it is down for repairs during the restore cycle Y_1. An up/down cycle is complete by time $X_1 + Y_1$. The random variables, X_i and Y_i are each assumed to be independent and identically distributed with means E(X) and E(Y). The sequence of pairs (X_i, Y_i) forms a two dimensional renewal process.

For this system, the availability, A(t), equals the fraction of time the system is up during (0, t).

The steady state availability is

$$A = \lim_{t \to \infty} A(t) = \frac{E(X)}{E(X) + E(Y)} \tag{10.100}$$

Assume that E(X) and E(Y) and, therefore, A are unknown. Hypothesize two values of A.

$$H_o: \quad A = A_o \tag{10.101}$$
$$H_1: A = A_1 \quad \text{where } A_1 < A_o$$

On the basis of test or field data, accept or reject the hypothesis H_o by comparing the computed A to a critical value appropriate to the test type and parameters.

It is assumed that both the up and down times are gamma distributed in order to derive the relationships of each test type. However, extremely useful results can be derived assuming the exponential distribution in both cases; the exponential distribution is used in the following examples.

10.8.1.1 Fixed Sample Size Plans

This test plan is based on having the system perform a fixed number of cycles R. The result is R pairs of times-to-failure and down times (X_1, Y_1), ..., (X_R, Y_R).

Let A^R = the observed availability of the test

$$A^R = \frac{\sum_{i=1}^{R} X_i}{\sum_{i=1}^{R} X_i + \sum_{i=1}^{R} Y_i} = \frac{1}{1 + Z_R} \tag{10.102}$$

where:

$$Z_R = \frac{\sum_{i=1}^{R} Y_i}{\sum_{i=1}^{R} X_i} \tag{10.103}$$

and

A^R = the maximum likelihood estimate of A

Let,

$$\rho_o = \frac{A_o}{1 - A_o} \quad \text{under the hypothesis } H_o \tag{10.104}$$

and

$$\rho_1 = \frac{A_1}{1 - A_1} \quad \text{under the hypothesis } H_1 \tag{10.105}$$

The procedure to be followed is:

$$\text{If} \quad \rho_o Z_R > C \quad \text{reject } H_o \tag{10.106}$$

$$\rho_o Z_R \le C \quad \text{accept } H_o$$

where C will be derived in the following paragraphs.

Assume that the up-times, X_i, are gamma distributed with parameters (m, θ) and the down times, Y_i, are gamma distributed with parameters (n, ϕ) with $n\phi = 1$.

Then it can be shown that ρZ_R is F-distributed with parameters (2nR, 2mR)

The critical value, C, and number of up/down cycles, R, are determined so that the significance test satisfies the consumer and producer risk requirements, α and β, i.e.,

$$P(\rho_o Z_R > C | A_o, R) \le \alpha \tag{10.107}$$

$$P(\rho_o Z_R \le C | A_1, R) \le \beta \tag{10.108}$$

which is equivalent to:

$$C \ge F_{\alpha}(2nR, 2mR) \tag{10.109}$$

$$\frac{\rho_1}{\rho_o} C \le F_{1-\beta}(2nR, 2mR) \tag{10.110}$$

Here $F_\alpha(\upsilon_1, \upsilon_2)$ denotes the upper α percentile of the F-distribution with parameters υ_1 and υ_2.

This system of inequalities has two unknowns and is solved numerically by finding the smallest integer R satisfying

$$F_\alpha(2nR, 2mR) \bullet F_\beta(2mR, 2nR) \leq D$$

where D is the discrimination ratio,

$$D = \frac{A_o(1 - A_1)}{A_1(1 - A_o)} = \frac{\rho_o}{\rho_1} \tag{10.112}$$

The value of R obtained in this way is used to calculate the critical value, C:

$$C = F_\alpha(2nR, 2mR) \tag{10.113}$$

The OC function is

$$OC(A) = P_r\left(\rho_o Z_R \leq C \middle| A\right) = F\left(2nR, 2mR; \frac{A}{1\text{-}A} \bullet \frac{C}{\rho_o}\right) \tag{10.114}$$

where $F(\upsilon_1, \upsilon_2; x)$ is the c.d.f. of the F-distribution with parameters υ_1 and υ_2.

The expected test duration is:

$$E(T) = \frac{R}{1 - A} \tag{10.115}$$

The variance of the total test duration is:

$$Var(T) = R \bullet \left\{\frac{1}{n} + \frac{1}{m} \bullet \left(\frac{A}{1 - A}\right)^2\right\} \tag{10.116}$$

For large sample size, R > 20, the distribution of T is approximately normal.

Example 19: Exponential Distribution

Let the time-to-failure and downtime distributions be exponentially distributed. Therefore, n = m = 1. Let $A_o = 0.9$ and $A_1 = 0.8$ and $\alpha = \beta = 0.2$. Calculate the parameters of the test.

Therefore,

$$\rho_o = \frac{0.9}{1 - 0.9} = 9$$

$$D = \frac{0.9(1 - 0.8)}{0.8(1 - 0.9)} = 2.25$$

Find the smallest integer R satisfying

$$F_{0.2}(2R, 2R) \leq \sqrt{2.25} = 1.5 \text{ where } F_{\alpha}(2R, 2R) = F_{\beta}(2R, 2R)$$

From a Table of Percentiles of the F-distribution we find

$$F_{0.2}(16,16) = 1.536 \text{ and } F_{0.2}(18,18) = 1.497$$

Therefore,
R = 9 satisfies the inequality

Therefore,
C = 1.497

The OC function is

$$OC(A) = F\left[18, 18;\ 0.166 \frac{A}{(1 - A)}\right]$$

10.8.1.2 Fixed-Time Sample Plans

In fixed-time sample plans, the system performs consecutive up/down cycles until a fixed-time T has elapsed. At this point, the test is terminated and the system may be either up or down. In this case the test time is fixed and the number of cycles is random.

Let A(T) = the observed availability at the end of the test.

The test procedure is

$$A(T) < A_c, \text{ then reject } H_o \tag{10.117}$$

$$A(T) \geq A_c, \text{ then accept } H_o \tag{10.118}$$

where the critical value A_c and test time T are chosen so that the significance test satisfies the following requirements on α and β.

$$P\left[A(T) < A_c \middle| A_o, T\right] \leq \alpha \tag{10.119}$$

$$P\left[A(T) \geq A_c \middle| A_1,\ T\right] \leq \beta \tag{10.120}$$

If λ_P is the upper P percentile of the standardized normal distribution and time is in mean down time units, the test time to be used is:

$$T = \left(\frac{1}{m} + \frac{1}{n}\right)\left\{\frac{\lambda_\alpha \bullet A_o\sqrt{1 - A_o} + \lambda_\beta \bullet A_1\sqrt{1 - A_1}}{A_0 - A_1}\right\}^2 \tag{10.121}$$

The critical value A_c is

$$A_c = \frac{A_o A_1\left[\lambda_\alpha\sqrt{1 - A_o} + \lambda_\beta\sqrt{1 - A_1}\right]}{\lambda_\alpha A_o\sqrt{1 - A_o} + \lambda_\beta A_1\sqrt{1 - A_1}} \tag{10.122}$$

The operating characteristic function is given by

$$OC(A) = 1 - \phi\left[\frac{A_c - A}{A \bullet \sqrt{\dfrac{\left(\dfrac{1}{m} + \dfrac{1}{n}\right)(1 - A)}{T}}}\right] \tag{10.123}$$

where ϕ is the standardized normal c.d.f.

Example 20: Exponential Distribution

In this example use the same data as in the previous example: A_o = 0.9, A_1 = 0.8, m = n = 1 by the exponential assumption, α = β - 0.2.

Using Eq. (10.121),

T = 58.5 (Mean Down Time Units)

Using Eq. (10.122),

$$A_c = 0.856$$

The OC function is

$$OC(A) = 1 - \phi \frac{0.856 - A}{A \cdot \sqrt{\frac{(1 - A) \cdot 2}{58.5}}}$$

10.9 System Design Considerations

Many of the design techniques and procedures detailed in Section 7 are directly appropriate to system design considerations.

As distinct from equipment design, system design is concerned with the broader aspects of organization and communication as they relate to the design of the individual equipment/systems. In the design of large scale systems, the need to think in terms of the whole in addition to the operation of individual equipment has become apparent. Complexity which characterizes large scale systems is at the root of the need for this broad perspective. Complex systems may perform many functions, process many inputs, translate and display many outputs, and cost a great deal of money. Therefore, only a broad perspective will permit a search for the optimum means of performing the required operations reliably.

A system R&M goal which is determined by some pertinent measure of system effectiveness stems from the system concept. Preliminary system design determines the types and minimum numbers of equipments in the network. The configuration of these equipments to achieve the system reliability goal is then determined. After a configuration is determined, an allocation of failure and repair rates is made to each equipment consistent with the system R&M goal. During the system development process, continual adjustments and re-evaluations of the means of achieving the R&M goal at least cost, are made.

The overall system design activity begins with the system concept and culminates with a set of equipment specifications that are meaningful enough to permit sound planning and comprehensive enough to present a broad perspective of the system as a single design entity. A basic philosophy of the system design is sought which allows for the determination of all important parameters in such a way that detailed design will not warrant serious redesign and the system will be optimized in its total aspect.

Equipment R&M predictions are most valuable in the early stage of a system's development. Once equipment R&M predictions are available to compare with the allocated operability ratios, discrepancies (if they exist) can be analyzed. It is also desirable to determine the expected state-of-the-art limits of failure rate and repair rate for each equipment in the system. Thus, if predictions indicate that the operability ratio allocated to certain equipments cannot be met without additional expenditures, it may be necessary to reallocate equipment failure and repair rates such that any additional expenditures may be minimized.

Basic to the system design process is the use of comprehensive mathematical models (usually computerized) in order to optimize the system parameters to be achieved at minimum cost. There is a logical sequence to system design, an example of which is presented here for guidance:

(1) Define system R&M parameters in terms of the operational requirements of the system.

(2) Develop an index of system R&M effectiveness.

(3) Rearrange the system into convenient non-interacting stages and equipments within each stage.

(4) Apply mathematical (and statistical) techniques to evaluate alternate system configurations in terms of reliability and cost.

(5) If necessary, evaluate the consequences in terms of cost and intangible factors of each alternate configuration.

(6) Specify a system configuration, a maintenance philosophy, and the relationship with other factors (interfaces).

(7) Allocate specifications in terms of failure rate (λ) (and/or repair rate (μ)) to the equipment within the system as design criteria.

(8) Predict the reliability and maintainability of each equipment and the system using available data either for similar equipments or, if this is not available, from published part failure rates and estimated repair rates.

(9) Compare allocated (goal) and predicted values to determine the next best course of action.

(10) Update R&M predictions and compare with goals to allow for continuous information feedback to choose the best course of action on the system level.

The procedure is by no means rigid and should vary from system to system. However, what is important is that the systematization of objectives and the use of analytic techniques.

Since availability is a system R&M parameter which is a combined measure of reliability and maintainability, it should be maximized in the most cost effective manner. Following are some design guidelines to maximize system availability:

(1) The designed-in failure rate should be minimized, and the MTBF should be maximized.

(2) The designed-in repair rate should be maximized, and the MTTR should be minimized.

(3) Whenever possible, maintenance actions should be carried out while the equipment is running normally, thus minimizing equipment downtime.

(4) If certain functions must be shut down for maintenance, the time required for shutting down the equipment should be minimized.

(5) Should certain components require shutdowns for maintenance actions, these maintenance actions should be required as rarely as possible.

(6) Should certain maintenance actions require shutdown, the time needed for these actions should be minimized.

(7) If certain components or subsystems require shutdowns for maintenance actions, as few components as possible should be shut down.

(8) The time required for logistics should be minimized.

(9) The time required for administrative actions should be minimized.

(10) Very well written and explicitly illustrated startup and operating manuals should be prepared and made available to the users of the equipment and to the maintenance personnel.

(11) Frequent and time-consuming, prescribed, preventive maintenance actions should be minimized.

(12) Special effort should be expended to use qualified and well trained maintenance personnel; their training should be updated as required and as design changes and more modern equipment are introduced.

(13) The Reliability Design Criteria (Section 7) and the Maintainability Design Criteria given in MIL-HDBK-470.

(14) Maintenance actions which require the dismantling, moving and assembling of heavy components and equipment should be facilitated by the provisioning of special lift-off lugs and accessories.

(15) Frequently inspected, serviced, maintained, and/or replaced components should be so located in the equipment that they are more accessible and easily visible.

(16) Servicing media like lubricants, impregnates, detergents, fuels, and other consumables should preferably be supplied automatically, and waste media should be removed automatically.

(17) Whenever possible, automatic diagnostics for fault identification should be provided via failure-indicating hardware and/or special minicomputers with the associated software.

(18) There should be maximum utilization of designed and built-in automatic test and checkout equipment.

(19) The distributions of all equipment downtime categories should be determined and studied, and those maintenance actions which contribute excessively to the overall equipment downtime should be singled out and their downtimes minimized.

(20) The distributions of the equipment downtimes resulting from the failure of key components should be studied; those components contributing significantly to the overall equipment downtime should be singled out and redesigned with lower failure rates and higher repair rates.

(21) The design should be such as to achieve maximum availability at budgeted cost or acceptable availability at minimum life cycle cost.

The last item in the previous list is what it's all about - designing for maximum availability at budgeted cost or acceptable availability at minimum cost. The rest of this section is devoted to that aspect of system R&M engineering.

10.10 Cost Considerations

The most important constraint that a system designer of today must consider is cost. All of the military services face the problem of designing and fielding systems that they can "afford," i.e., which have reasonable life cycle costs (LCC). R&M have a significant impact on life cycle costs (LCC) because they determine how frequently a system fails and how rapidly it is repaired when it fails.

Thus, a vital system design consideration is how to minimize LCC by maximizing R&M within given design cost constraints.

10.10.1 Life Cycle Cost (LCC) Concepts

Life cycle cost is the total cost of acquiring and utilizing a system over its entire life span. LCC includes all costs incurred from the point at which the decision is made to acquire a system, through operational life, to eventual disposal of the system. A variety of approaches can be used to estimate the cost elements and provide inputs to the establishment of a life cycle cost model. The total life cycle cost model is thus composed of subsets of cost models which are then exercised during trade-off studies. These cost models range from simple informal engineering/cost relationships to complex mathematical statements derived from empirical data.

Total LCC can be considered as generated from two major areas:

(1) system acquisition cost

(2) system utilization cost

In simple mathematical terms, the above can be stated by:

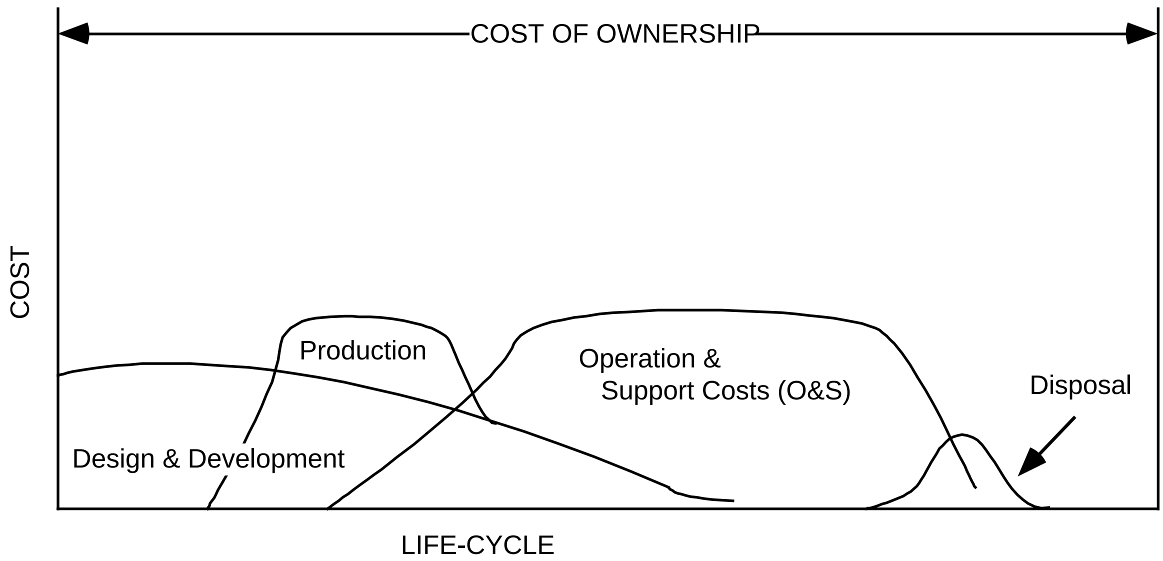
$$LCC = AC + SUC \tag{10.124}$$

where:

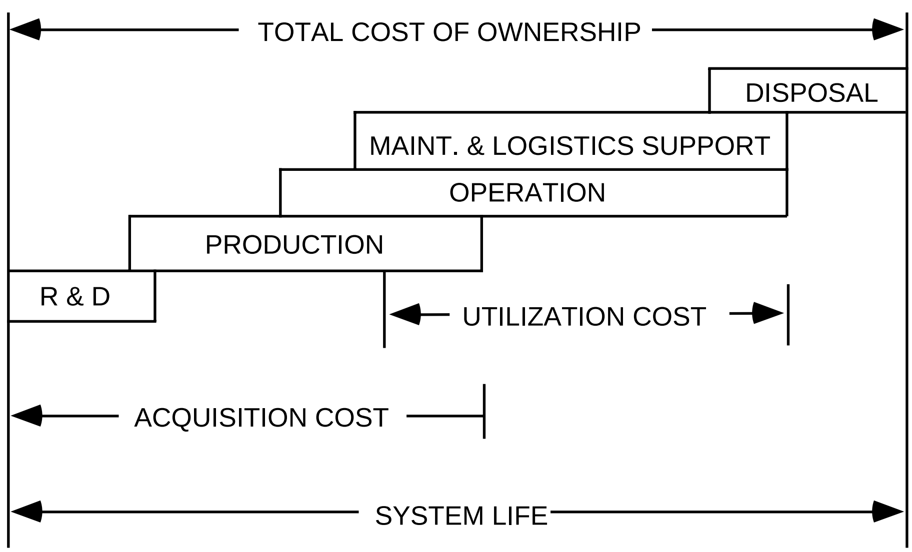
LCC = life cycle cost
AC = acquisition cost
SUC = system utilization cost

Figure 10.10-1 identifies the more significant cost categories and shows (conceptually) how LCC may be distributed in terms of the major cost categories over a system life cycle.

In general, design and development costs include basic engineering, test and system management; production costs include materials, labor, General and Administrative, overhead, profit, capitalization, handling, and transportation; operational and support (O&S) cost includes a sizable number of factors including initial pipeline spares and replacement, equipment maintenance (on/off), inventory entry and supply management, support equipment, personnel training, technical data/ documentation, and logistics management. Disposal costs include all costs associated with deactivating and preparing the system for disposal through scrap or salvage programs. Disposal cost may be adjusted by the amount of value received when the disposal process is through salvage.



(a)



(b)

FIGURE 10.10-1: LCC CATEGORIES VS. LIFE CYCLE

The application of R&M disciplines plays a key role in minimizing LCC, since one, (R), determines the frequency of failure and the other, (M), determines the time to fix a failure. System designers must balance performance, reliability, maintain- ability, and production goals in order to minimize LCC.

To meet this need, attention is focused on structuring a balanced design approach derived from a life cycle cost model that is comprised of and governed by submodels, which calculate R&M and cost variables. Figure 10.10-2 presents an overview of the R&M and cost methodology within this framework. This figure shows the life cycle cost model as the vehicle for which estimates for operation, performance, reliability, maintainability, and cost are traded off to obtain "design to" target goals which collectively represent a balanced design. This life cycle cost model includes submodels which are representative of acquisition costs and maintenance and logistics support costs, subject to the constraints of functional objectives and minimal performance requirements.

Some of the major controllable factors contributing to system life cycle cost related to these cost categories are shown in Table 10.10-1. In practice, however, all of these cost factors will not appear in each LCC analysis. Only those factors relative to the objective and life cycle phase of the analysis are included. For example, a comparison of standard commercial equipment would not include design and development costs but would include procurement and support costs. Similarly, a throwaway part or assembly would result in a simpler decision model than an item requiring on-site and off-site maintenance and repair. Thus, a system LCC decision model

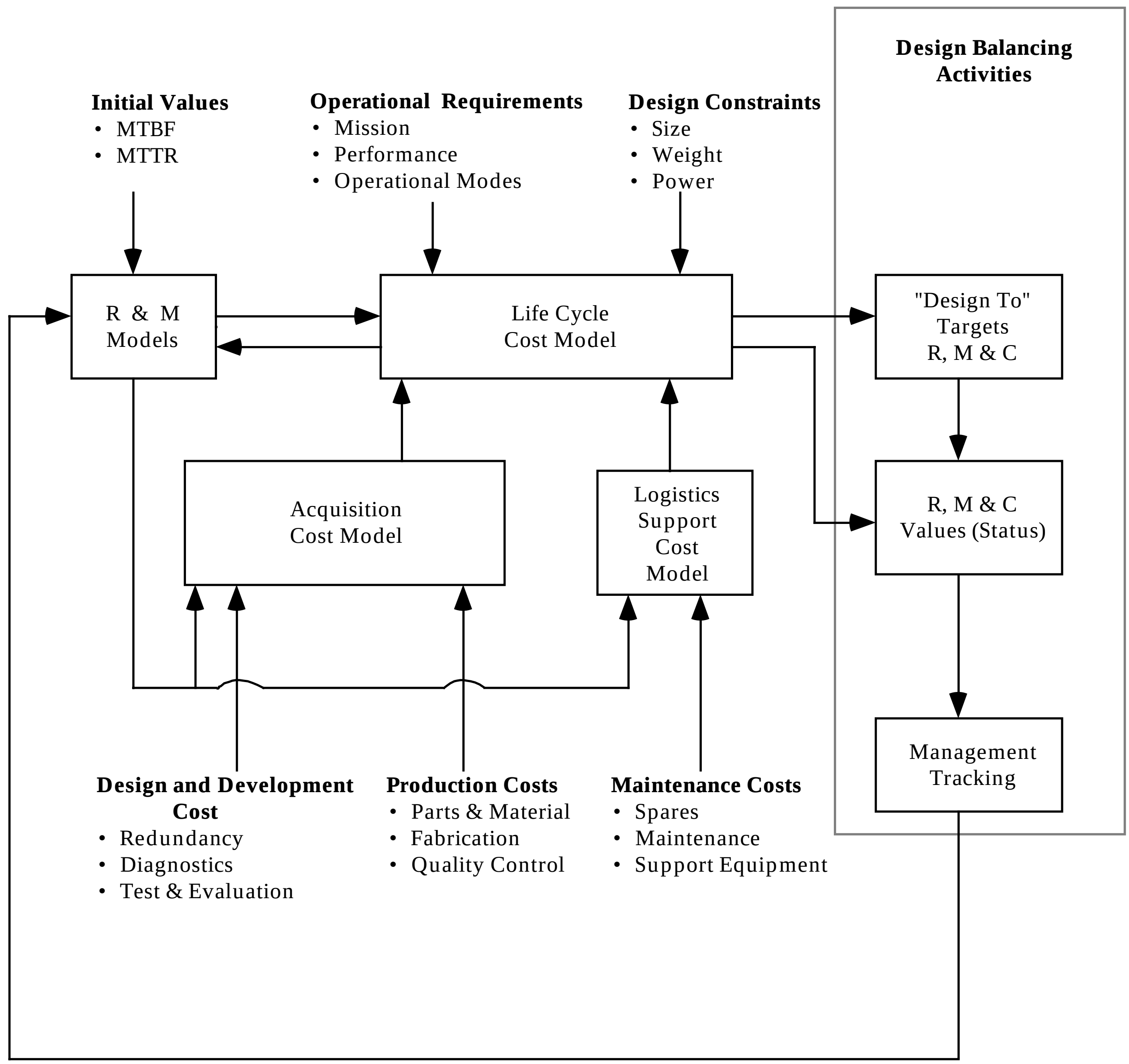


FIGURE 10.10-2: R&M AND COST METHODS

TABLE 10.10-1: LIFE CYCLE COST BREAKDOWN

Total Life Cycle Cost			
Acquisition		Operation & Support	
Basic Engineering - Design (Electrical, Mechanical) - Reliability, Maintainability - Human Factors Producibility - Component - Software Test & Evaluation - Development - $\bar{R}$ Growth - R&M Demonstration - $\bar{R}$ Screening - $\bar{R}$ Acceptance Experimental Tooling - System - $\bar{R}$ Program - $\bar{M}$ Program - Cost Manufacturing & Quality Engineering - Process Planning - Engineering Change Control - Q.A. Planning, Audits, Liaison, etc.	Recurring Production Costs - Parts & Materials - Fabrication - Assembly - Manufacturing Support - Inspection & Test - Receiving - In-process - Screening - Burn-In - Acceptance - Material Review - Scrap Rate - Rework Nonrecurring Production Costs - First Article Tests - Test Equipment - Tooling - Facilities - System Integration - Documentation (including maintenance instructions & operating manuals) - Initial spares (organizational, intermediate and depot) (pipeline)	Logistics & Maintenance Support - Pipeline Spares - Replacement Spares (organization, intermediate, depot) - On-Equipment Maintenance - Off-Equipment Maintenance - Inventory Entry & Supply Management - Support Equipment (including maintenance) - Personnel Training & Training Equipment - Technical Data & Documentation - Logistics Management - Maintenance Facilities & Power - Transportation (of failed items to and from depot)	Operation - Supply Management - Technical Data - Personnel - Operational Facilities - Power - Communications - Transportation - Materials (excluding maintenance) - General Management - Modifications - Disposal

Figure 10.10-3 illustrates (conceptually) the relationships between reliability and cost. The top curve is the total life cycle cost and is the sum of the acquisition (or investment) and O&S costs. The figure shows that as a system is made more reliable (all other factors held constant) the support cost will decrease since there are fewer failures. At the same time, acquisition cost (both development and production) is increased to attain the improved reliability. At a given point, the amount of money (investment) spent on increasing reliability will result in exactly that same amount saved in support cost. This point represents the reliability for which total cost is minimum. Consequently, reliability can be viewed as an investment during acquisition for which the return on investment (ROI) is a substantial reduction of maintenance support (the operational costs tend to remain constant regardless of reliability investment). An analogous relationship exists between maintainability and cost.

The implementation of an effective program based on proven LCC principles complete with analytical models and supporting input cost data will provide early cost visibility and control, i.e., indicate the logistics and support cost consequences of early research, development, and other subsequent acquisition decisions, such that timely adjustments can be made as the program progresses.

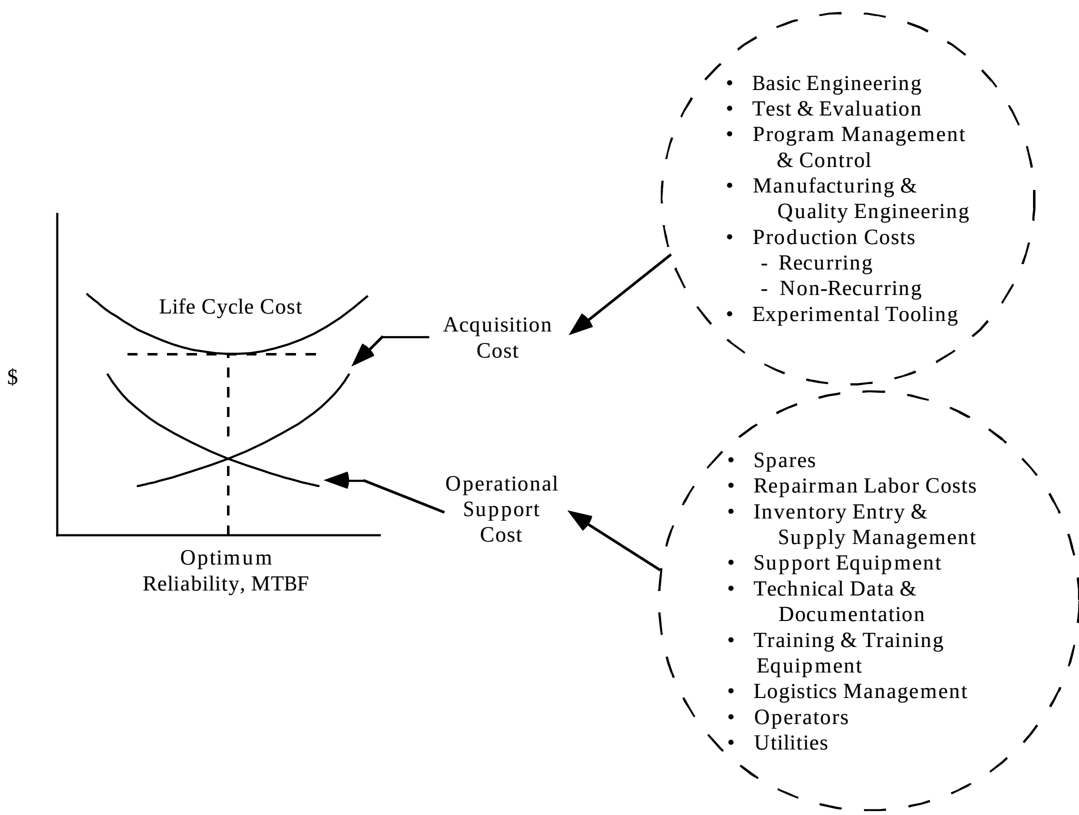


FIGURE 10.10-3: LIFE CYCLE COSTS VS. RELIABILITY

10.11 References for Section 10

1. Von Alven, W.H., Ed., Reliability Engineering. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1964.

2. AFSC-TR-65-6, Chairman's Final Report. Weapon System Effectiveness Industry Advisory Committee (WSEIAC), Air Force Systems Command, January 1965, (AD-467816), also

AFSC TR-65-1	"Requirements Methodology," Final Report of Task Group I
AFSC TR-65-2	"Prediction Measurement (Concepts, Task Analysis, Principles of Model Construction)," Final Report of Task Group II
AFSC TR-65-3	"Data Collection and Management Reports," Final Report of Task Group III
AFSC TR-65-4	"Cost Effectiveness Optimization," Final Report of Task Group IV
AFSC TR-65-5	"Management Systems," Final Report of Task Group V

3. Elements of Reliability and Maintainability. DoD Joint Course Book, U.S. Army Management Engineering Training Agency, Rock Island, IL, 1967.

4. Systems Effectiveness. System Effectiveness Branch, Office of Naval Material, Washington, DC, 1965, (AD-659520).

5. Navy Systems Performance Effectiveness Manual. NAVMAT P3941, Headquarters Naval Material Command, Washington, DC, 1 July 1960.

7. Blanchard, B.S., "Cost Effectiveness, System Effectiveness, Integrated Logistic Support, and Maintainability," IEEE Transactions in Reliability, R-16, No. 3, December 1967.

8. Barlow, R.E., and F. Proschan, Mathematical Theory of Reliability. New York, NY: John Wiley & Sons, Inc., 1965.

9. Kozlov, B.A., and I.A. Ushakov, Reliability Handbook. Holt, Rinehart and Winston, Inc., NY, 1970.

10. Myers, R.H., K.L. Wong and H.M. Gordy, Reliability Engineering for Electronic Systems. New York, NY: John Wiley and Sons, Inc., 1964.

11. Mathematical Models for the Availability of Machine Gun Systems. Technical Report No. 3, prepared by Igor Bazovzky and Associates, Inc., for the Small Arms System Laboratory, U.S. Army Weapons Command, February 1970.

12. Availability. PAM 69-8, U.S. Army Combat Development Command Maintenance Agency, Aberdeen Proving Ground, MD, November 1970.

13. Maintainability Design Criteria Handbook for Designers of Shipboard Electronic Equipment. NAVSHIPS 94324, Department of the Navy, Change 2, 1965.

14. Orbach, S., The Generalized Effectiveness Methodology (GEM) Analysis Program. U.S. Naval Applied Science Laboratory, Brooklyn, NY, May 1968.

15. Evaluation of Computer Programs for System Performance Effectiveness, Volume II. RTI Project SU-285, Research Triangle Institute, Research Triangle Park. NC, August 1967.

16. "Computer Tells Launch Vehicle Readiness," Technology Week, April 1967.

17. Dresner, J., and K.H. Borchers, "Maintenance, Maintainability and System Requirements Engineering," Proceedings of the Third Annual Aerospace Reliability and Maintainability Conference, 1964.

18. Economos, A.M., "A Monte Carlo Simulation for Maintenance and Reliability," Proceedings of the Third Annual Aerospace Reliability and Maintainability Conference, 1964.

19. Faragher, W.E., and H.S. Watson, "Availability Analyses - A Realistic Methodology," Proceedings of the Tenth National Symposium on Reliability and Quality Control, 1964, pp. 365-378.

20. Horrigan, T.J., Development of Techniques for Prediction of System Effectiveness, RADC-TDR-63-407, Cook Electric Company, February 1964, AD-432844.

21. Maintainability Bulletin No. 8, "Maintainability Trade-Off Techniques," Electronic Industries Association, July 1966.

22. Ruhe, R.K., "Logic Simulation for System Integration and Design Assurance," Proceedings of the Third Annual Aerospace Reliability and Maintainability Conference, 1964.

23. Smith, T.C., "The Support Availability Multi-System Operations Model," Proceedings of the Third Annual Aerospace Reliability and Maintainability Conference, 1964.

24. Survey of Studies and Computer Programming Efforts for Reliability, Maintainability, and System Effectiveness. Report OEM-1, Office of the Director of Defense Research and Engineering, September 1965, AD-622676.

25. Sandler, G.H., System Reliability Engineering. Englewood Cliffs, NJ: Prentice-Hall, 1963.

26. Rise, J.L., "Compliance Test Plans for Availability," Proceedings of the 1979 Annual Reliability and Maintainability Symposium, Washington, DC, January 1979.

27. Arsenault, J.E., and J.A. Roberts, "Reliability and Maintainability of Electronic Systems," Computer Science Press, 9125 Fall River Lane, Potomac, MD 20854, 1980.

28. SD-2, "Buying Commercial Nondevelopmental Items: A Handbook," Office of the Assistant Secretary of Defense for Production and Logistics, April 1996.

29. SHARP Notes, SHARP (Standard Hardware Acquisition and Reliability Program) Program Manager, Naval Warfare Surface Center, Crane Division, Crane, IN, Code PMI.

30. "SHARP Handbook on COTS/MIL Systems," SHARP (Standard Hardware Acquisition and Reliability Program) Program Manager, Naval Warfare Surface Center, Crane Division, Crane, IN, Code PMI, 1994.

31. NAVSO P-3656, Department of the Navy Handbook for the Implementation of Nondevelopmental Acquisition.

32. MAN PRIME Handbook for Nondevelopmental Item (NDI) Acquisition.

33. ARMP-8, "Reliability and Maintainability: Procurement of Off-the-Shelf Equipment," Ministry of Defense Standard 00-40 (Part 8).

34. RADC-TR-85-91, "Impact of Nonoperating Periods on Equipment Reliability," Rome Laboratory, 1985.

35. RADC-TR-89-299, "Reliability and Maintainability Operational Parameter Translation II" Rome Laboratory, 1989.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

11.0 PRODUCTION AND USE (DEPLOYMENT) R&M

11.1 Introduction

An effective system reliability engineering program begins with the recognition that the achievement of a high level of R&M in actual use is a function of design as well as all life cycle activities. Design establishes the inherent R&M potential of a system or equipment item. The transition from the computer-aided-design or paper design to actual hardware, and ultimately to operation, many times results in an actual R&M that is far below the inherent level. The degree of degradation from the inherent level is directly related to the inspectability and maintainability features designed and built into the system, as well as the effectiveness of the measures that are applied during production and storage prior to deployment to eliminate potential failures, manufacturing flaws and deterioration factors.

The impact of production, shipment, storage, operation and maintenance degradation factors on the reliability of a typical system or equipment item and the life cycle growth that can be achieved is conceptually illustrated in Figure 11.1-1. The figure depicts the development of a hardware item as it progresses through its life cycle stages. The figure shows that an upper limit of reliability is established by design, and that, as the item is released to manufacturing, its reliability will be degraded and as production progresses, with resultant process improvements and manufacturing learning factors, reliability will "grow." The figure further shows that when the item is released to the field, its reliability will again be degraded. As field operations continue and as operational personnel become more familiar with the equipment and acquire maintenance experience reliability will again "grow."

As was discussed in Section 7, reliability design efforts include: selecting, specifying and applying proven high quality, well-derated, long life parts; incorporating adequate design margins; using carefully considered, cost effective redundancy; and applying tests designed to identify potential problems. Emphasis is placed on incorporating ease of inspection and maintenance features, including use of easily replaceable and diagnosable modules (or components) with built-in test, on-line monitoring and fault isolation capabilities. During development, reliability efforts include the application of systematic and highly-disciplined engineering analyses and tests to stimulate reliability growth and to demonstrate the level of reliability that has been achieved and the establishment of an effective, formal program for accurately reporting, analyzing, and correcting failures which would otherwise occur during operation.

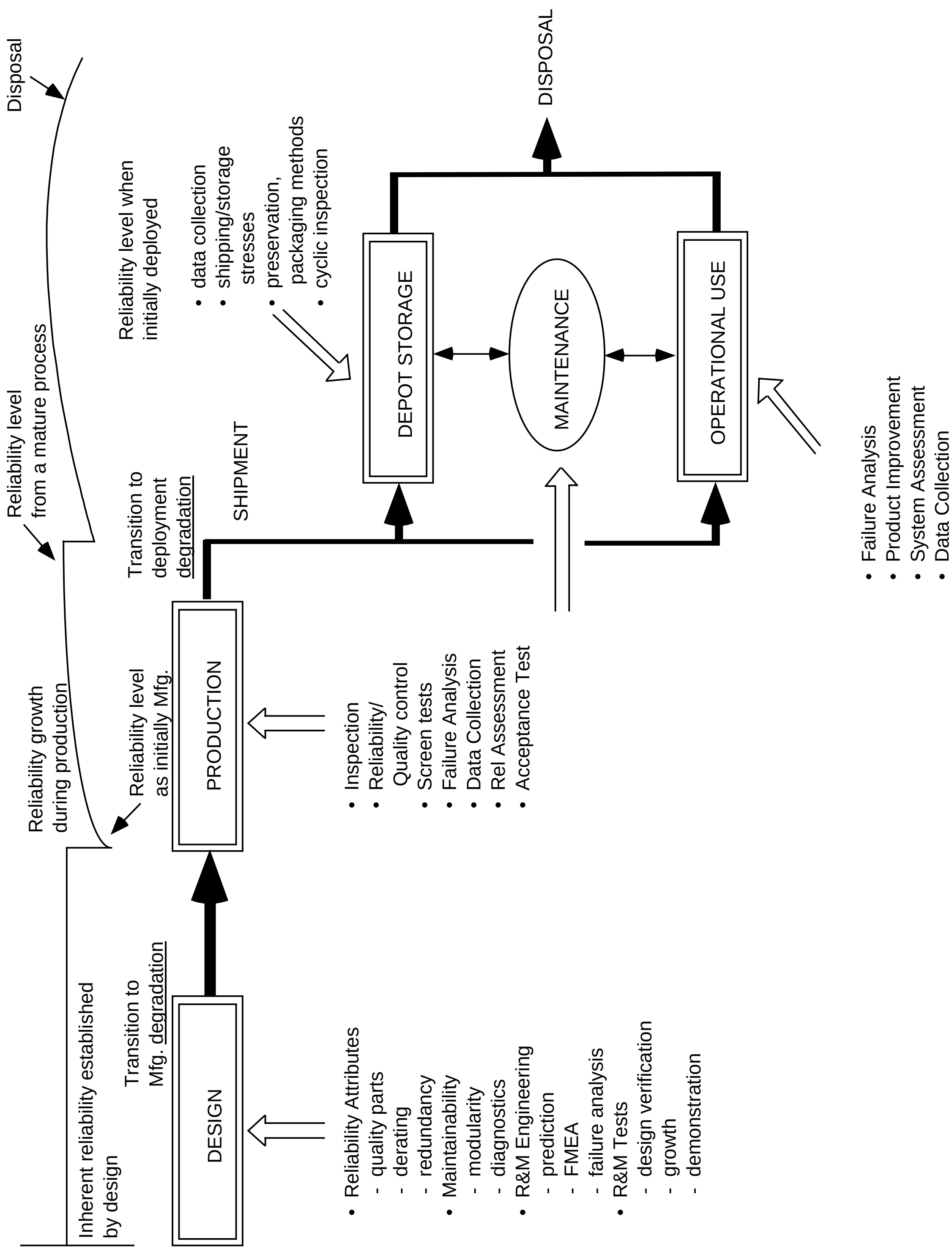


FIGURE 11.1-1: RELIABILITY LIFE CYCLE DEGRADATION & GROWTH CONTROL

SECTION 11: PRODUCTION AND USE (DEPLOYMENT) R&M

Once the inherent or designed-in R&M is established, engineering efforts focus on the prevention or reduction of degradation. Well-planned and carefully-executed inspections, tests, and reliability/quality control methods are applied during production (as well as during storage and operation), to eliminate defects and minimize degradation. Manufacturing, transportation, and storage environmental stresses, as well as inspection methods and operational/maintenance procedures are continually assessed to determine the need for better inspection, screening, and control provisions to improve R&M.

This section discusses reliability degradation and growth during production and deployment. Basic procedures and guidelines are described that can be used to plan post-design reliability control measures, including the assessment and improvement of reliability during production, shipment, storage and use. Also discussed are maintainability control procedures during production and deployment.

11.2 Production Reliability Control

The need for a reliability program applicable to production becomes evident when considering that:

(1) Manufacturing operations introduce unreliability into hardware that is not ordinarily accounted for by reliability design engineering efforts.

(2) Inspection and test procedures normally interwoven into fabrication processes are imperfect and allow defects to escape which later result in field failure.

Therefore, if the reliability that is designed and developed into an equipment/system is to be achieved, efforts must also be applied during production to ensure that reliability is built into the hardware. To realistically assess and fully control reliability, the degradation factors resulting from production must be quantitatively measured and evaluated. This is particularly important for a newly fabricated item, where manufacturing learning is not yet complete and a high initial failure rate can be expected.

Since the effectiveness of inspection and quality control relates directly to reliability achievement, it would be useful to discuss basic quality engineering concepts prior to discussing specific aspects of production reliability degradation and improvement.

11.2.1 Quality Engineering (QE) and Quality Control (QC)

The quality of an item is the degree to which it satisfies the user, or it may be stated as a measure of the degree to which it conforms to specified requirements. It can be expressed in terms of a given set of attributes defined in measurable quantitative terms to meet operational requirements. Quality level can be measured by the number of defects in a given lot or item.

The purpose of a quality control program is to ensure that these attributes are defined and maintained throughout the production cycle (and continued during storage and operation). Included as part of the quality control program is the verification and implementation of inspection systems, statistical control methods, and cost control and acceptance criteria. Critical to the quality control function is the establishment of adequate acceptance criteria for individual items to assure appropriate quality protection.

Another reason for the importance of a quality control program has to do with continuous quality improvement. Measurement systems must be in place to be able to separate special problems from those that can be attributed to common causes such as random variation in the design, development and manufacturing process. Further, as stated in reference 16, "the collection and analysis of data is the key to identifying specific improvement targets. Data is the raw material that we turn into improvement projects. Data must be the basis for making every decision about improvement." While all data are not necessarily a result of the QC program, having such a program is a key element to ensuring that data are produced and collected that can be used to ensure quality protection and provide a baseline for quality improvement.

Quality, as with reliability, is a controllable attribute which can be planned during development, measured during production, and sustained during storage and field repair actions. The achievement of acceptable quality for a given item involves numerous engineering and control activities. Figure 11.2-1 depicts some of these activities as they apply to a system over time. These activities represent an approach to a comprehensive and well rounded Quality Control Program.

Keys to ensuring the basic quality of a hardware item as depicted in Figure 11.2-1 are: the specification of cost effective quality provisions and inspections covering the acquisition of new hardware items; the storage of hardware and material; and the repair, reconditioning or overhaul of deployed items. This means that quality requirements should be included in procurement specifications, in-storage inspection requirements, and in-maintenance work requirements, as applicable, and that responsive quality programs are to be planned and implemented to meet these requirements. This section discusses quality control during the acquisition of new systems and hardware items.

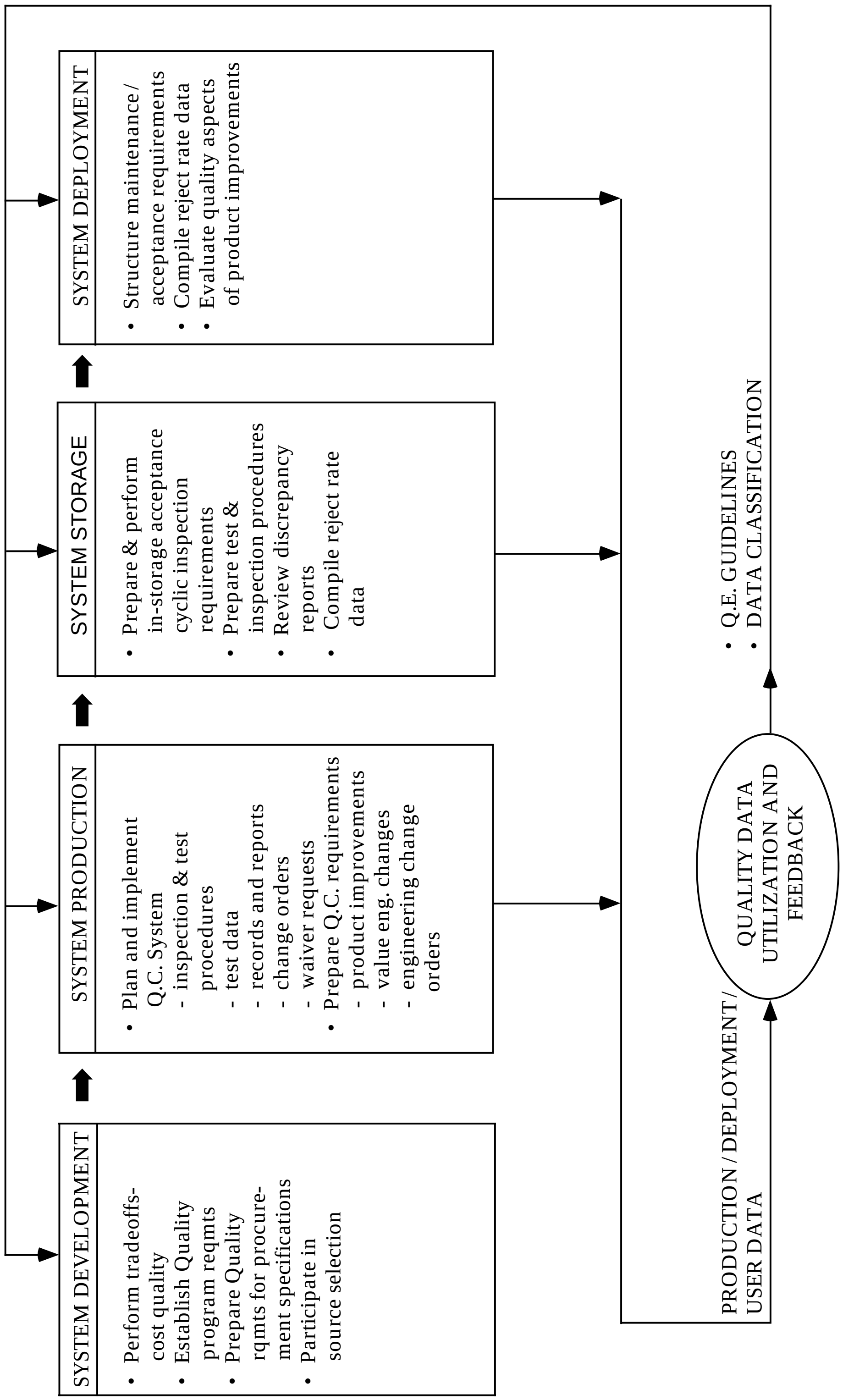


FIGURE 11.2-1: QUALITY ENGINEERING AND CONTROL OVER TIME

11.2.1.1 Quality System Requirements

Until recently, due to acquisition reform (AR) changes, quality requirements applied during acquisition generally followed Military Specification MIL-Q-9858, Quality Program Requirements. MIL-Q-9858 was the basic standard for planning quality programs for DoD development and production contracts. Under AR, MIL-Q-9858A was cancelled by Notice 2 dated October 1996. As with other canceled military documents, there is no barrier to a system developer, or contractor, using MIL-Q-9858A as a basis for a quality program, or quality system.

Prior to its cancellation, MIL-Q-9858A, Amendment 3, dated 5 September 1995 stated that for new designs, the International Organization for Standardization (ISO) 9001, ISO 9002 quality system standards, the ANSI/ASQC Q9001, ANSI/ASQC Q9002 quality system standards, or a comparable higher-level non-government quality system should be used. The ANSI/ASQC Q9000 series documents are the technical equivalent to the ISO 9000 series documents.

11.2.1.1.1 ISO 9000

Because the DoD has adopted the ANSI/ASQC Q9000 Standards Series (technical equivalent to the ISO 9000 Standards Series), it is prudent to provide some information on the ISO 9000 quality system. Adoption by the DoD means that the ANSI/ASQC Q9000 documents are listed in the DoD Index of Specifications and Standards (DODISS) and are available to DoD personnel through the DODISS publications distribution center. Note, however, that the use of the Q9000 standards has not been included within the Federal Acquisition Regulation (FAR) or the DoD FAR Supplement (DFARS). In fact, DFARS paragraph 246.102(4) states that departments and agencies shall: "Provide contractors the maximum flexibility in establishing efficient and effective quality programs to meet contractual requirements. Contractor quality programs may be modeled on military, commercial, national, or international quality standards." The last sentence allows application of MIL-Q-9858A, ISO 9000 or ANSI/ASQC Q9000 standards for quality planning.

As previously noted, ISO 9000 is a family of standards on quality. These standards have been developed by ISO Technical Committee TC 176. The family of standards is shown in Figure 11.2-2.

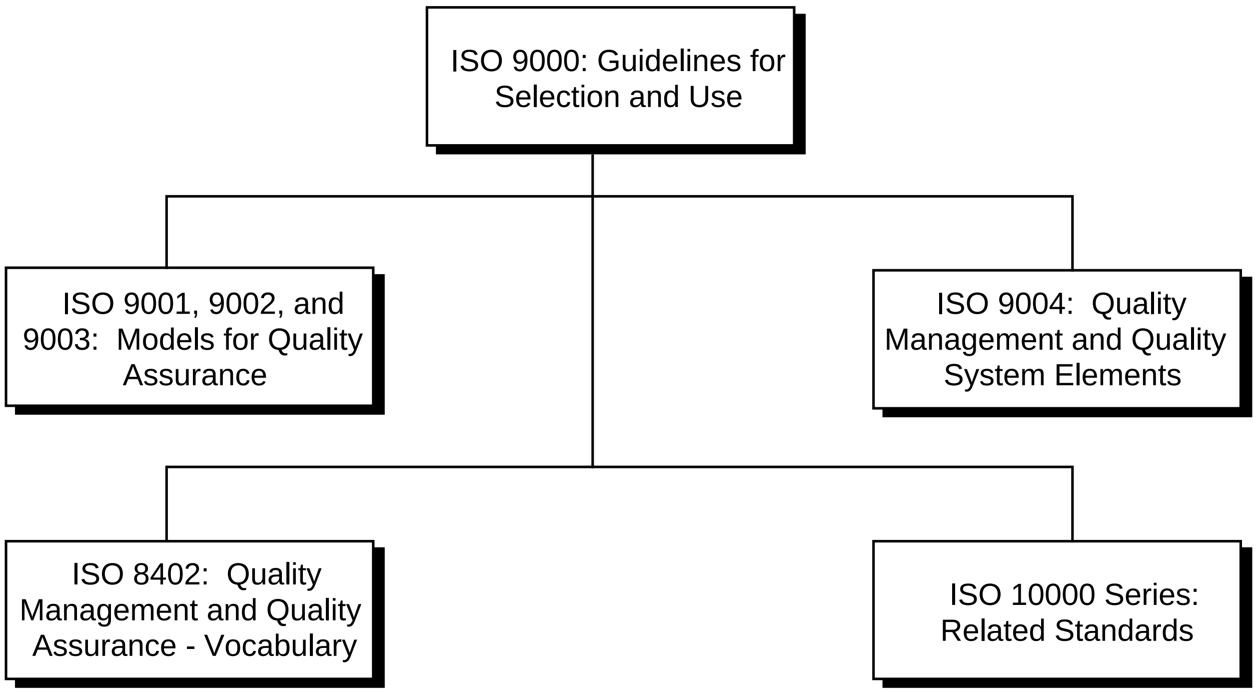


FIGURE 11.2-2: ISO 9000 FAMILY OF STANDARDS

The titles of the five standards are:

- ISO 9001: Quality Systems - Model for quality assurance in design, development, production, installation, and servicing certification system

- ISO 9002: Quality Systems - Model for quality assurance in production, installation and servicing

- ISO 9003: Quality Systems - Model for quality assurance in final inspection and test.

- ISO 9000
  - Part 1 (9000-1): Guidelines for Selection and Use
  - Part 2 (9000-2): Quality Management and Quality Assurance Standards
    - Generic guidelines for the application of ISO 9001, ISO 9002 and ISO 9003
  - Part 3 (9000-3): Quality Management and Quality Assurance Standards
    - Guidelines for the application of ISO 9001 to the development, supply and maintenance of software
  - Part 4 (9000-4): Quality Management and Quality Systems Elements - Guidelines

In reviewing the family of ISO 9000 standards, it can be seen that rather than having one, all-encompassing quality standard and relying on the user to tailor it accordingly, some tailoring has already been done. Further, there are several guidance documents (i.e., ISO 9000-1 through 9000-4) are available to assist in selecting a particular quality system standard.

11.2.1.1.1.1 Comparing ISO 9000 to MIL-Q-9858

The major thrusts behind both MIL-Q-9858 and the ISO 9000 Series Standards are essentially the same. Table 11.2-1, previously printed in MIL-HDBK-338, shows that MIL-Q-9858 covered 17 quality program elements.

As a comparison, ISO 9001 (ANSI/ASQC Q9001-1994) defines the following 20 elements of a quality system:

1. Management responsibility
2. Quality system
3. Contract review
4. Design control
5. Document and data control
6. Purchasing
7. Control of customer - supplied product
8. Product identification and traceability
9. Process control
10. Inspection and testing
11. Control of inspection, measuring and test equipment
12. Inspection and test status
13. Control of nonconforming product
14. Corrective and preventive action
15. Handling, storage, packaging, preservation and delivery
16. Control of quality records
17. Internal quality audits
18. Training
19. Servicing
20. Statistical techniques

Many of the subparagraphs within the above 20 elements cover the same subject area as did MIL-Q-9858. The MIL-Q-9858 elements are listed in Table 11.2-1.

Whereas MIL-Q-9858 recommended the use of MIL-I-45208, Inspection System Requirements, when requirements were less stringent, the ISO 9000 family of standards include ISO 9002 and ISO 9003 for use in such cases.

TABLE 11.2-1: MIL-Q-9858 QUALITY PROGRAM ELEMENTS

- *Quality Program Management*
  - Organization
  - Initial Quality Planning
  - Work Instructions
  - Records
  - Corrective Action
- *Facilities and Standards*
  - Drawings, Documentation and Changes
  - Measuring and Testing Equipment
  - Production Tooling Used as Media of Inspection
  - Use of Contractor's Inspection Equipment
  - Advanced Metrology Requirements
- *Control of Purchases*
  - Responsibility
  - Purchasing Data
- *Manufacturing Control*
  - Materials and Material Control
  - Production Processing and Fabrication
  - Completed Item Inspection and Testing
  - Handling, Storage and Delivery
- *Statistical Quality Control and Analysis*
  - Indication of Inspection Status

11.2.1.1.1.2 Why ISO 9000?

There are varied reasons for recent interest in ISO 9000, and in becoming what is called "ISO 9000 Registered," both within the US and world-wide. A detailing of the reasons and history of ISO 9000 can be found in references 17 - 19. However, a brief explanation is provided here. The development for a worldwide set of quality standards grew as each country's economy became a global one, rather than local. To meet this need, and to develop a set of standards that would be acceptable to a large number of countries worldwide, ISO, having a global membership, created the ISO 9000 series standards in 1987. The US member of ISO is ANSI.

Recently, the European Community (EC), made up primarily of the Western European powers, adopted a policy of buying "regulated products" (e.g., environmental, health, safety related) from companies that were proven to be compliant with the quality system requirements of ISO 9000.

(7) Quality of purchased material

(8) Extent of quality control audits and vendor surveillance

(9) Extent of line certification

One of the basic functions of a manufacturer's quality control organization is to make tradeoff decisions relative to these considerations and to ensure that quality is adequately planned and controlled, consistent with specified requirements and the constraints of the particular system.

Accomplishment of the quality control function, like reliability, requires a comprehensive and highly detailed program comprising of effective, systematic, and timely management activities, engineering tasks, and controlled tests. The production and acceptance of high quality hardware items requires definition and implementation of an effective quality management and control program that includes:

(1) Performance of detailed quality analysis, planning and cost tradeoff analyses during hardware development.

(2) Application of systematic and highly disciplined quality control tasks during production

SECTION 11: PRODUCTION AND USE (DEPLOYMENT) R&M

TABLE 11.2-2: QUALITY ENGINEERING TASKS

- Review engineering drawings and specifications, prototype test data, and R&M engineering data to determine impact on item quality.
- Review purchased material from a quality standpoint. This would include:
  - Evaluation of purchase requisitions and orders
  - Selection and certification of vendors
  - Approval of vendor component part/assembly samples
  - Review of part/material specifications (in particular, critical component identification and control)
  - Evaluation of purchased material through inspection planning, incoming inspection, and complete test data documentation control
  - Disposition and allocation of inspected material, discrepant material, review board provisions
- Evaluate material item manufacturing through a review of process inspection planning, workmanship and acceptance standards, instructions and procedures, production and QA inspection and testing.
- Determine adequacy (accuracy, calibration program, etc.) of inspection tests, production equipment, and instrumentation.
- Provide engineering direction and guidance for the acceptance inspection and test equipment in support of new item procurement production, reconditioning, and maintenance activities.
- Exercise control over the acquisition, maintenance, modification, rehabilitation, and stock level requirements of final acceptance inspection and test equipment.
- Provide product assurance representation to Configuration Control Boards, and serve as the control point for evaluation and initiation of all configuration change proposals.
- Advise, survey, and provide staff guidance for special materials and processes technology, as applied to quality control systems.
- Evaluate the adequacy, effect, and overall quality of process techniques, particularly those processes which historically have a significant impact on an item's quality.
- Evaluate reliability/quality data stemming from production, storage and use to:
  - Identify critical items having high failure rates, poor quality or requiring excessive maintenance
  - Identify significant failure modes, mechanisms, and causes of failure
  - Reduce and classify data and, in particular, define and classify quality defect codes
  - Formulate Q.C. guidelines to support preparation of procurement specifications
  - Prepare failure data and statistical summary reports
- Identify critical material items where cost effective reliability and quality improvement can be effectively implemented. Candidates for improvement include those items which have a history of poor quality, frequent failure, require extensive maintenance effort, and have excessive support costs.

TABLE 11.2-2: QUALITY ENGINEERING TASKS (CONTD)

• Make general reliability/quality improvement recommendations on selected equipment. • Provide product assurance engineering impact evaluations for configuration change, product improvement, and value engineering or cost improvement proposals. • Determine the effectiveness of improvements on item reliability/quality. • Develop calibration procedures and instructions, maintain and recommend changes to publications, equipment improvement recommendations and new calibration requirements, addressing calibration parameters.

An integral part of an effective quality control program is to make available to its engineers documented instructions, procedures, or guidance which fully describe the functions and tasks required to achieve its objective. Data collected during early production and testing activities, as well as historical data on similar systems from depot storage, maintenance actions, and field operations, can be compiled, reduced and applied to improve the production quality engineering and control activities. This data, for example, can be used to:

(1) Track quality

(2) Compare the benefits of various quality programs and activities:
- Production quality control techniques
- Vendor control and audits
- 100% inspection
- Sampling inspection
- Special quality assurance procedures

(3) Determine the effectiveness of quality control programs related to:
- Materials and materials control
- Inspection and testing of piece parts and subassemblies
- Production processing fabrication
- Completed item inspection and testing
- Handling, storage and delivery
- Corrective action implementation

(4) Determine the effects of depot storage, operation and maintenance factors:
- Depot level inspections
- Personnel
- Logistics
- Operational environment
- Mission profile

- Maintenance organization
- Quality classification codes
- Quality guidelines to support preparation of procurement specifications, storage inspection requirements and maintenance requirements

11.2.2 Production Reliability Degradation Assessment & Control

As was previously shown, the extent of reliability degradation during production depends on the effectiveness of the inspection and quality engineering control program. Reliability analysis methods are applied to measure and evaluate its effectiveness and to determine the need for process improvement or corrective changes. The accomplishment of the analysis task and, more important, how well subsequent corrective measures are designed and implemented will dictate the rate at which reliability degrades/grows during production. Specifically, reliability degradation is minimized during manufacturing, and reliability grows as a result of improvements or corrective changes that:

(1) Reduce process-induced defects through:
- Accelerated manufacturing learning
- Incorporation of improved processes

(2) Increase inspection efficiency through:
- Accelerated inspector learning
- Better inspection procedures
- Incorporation of controlled screening and burn-in tests

As process development and test and inspection efforts progress, problem areas become resolved. As corrective actions are instituted, the outgoing reliability approaches the inherent (design-based) value.

The approach to assessing and controlling reliability degradation involves quantifying process-induced defects and determining the effectiveness of the inspections and tests to remove the defects, i.e., estimating the number of defects induced during assembly and subtracting the number estimated to be removed by the quality/reliability inspections and tests. This includes estimating defects attributable to purchased components and materials, as well as those due to faulty workmanship during assembly.

Process-induced defects can be brought about by inadequate production capability or motivation and from fatigue. Quality control inspections and tests are performed to “weed out” these defects. No inspection process, however, can remove all defects. A certain number of defects will escape the production process, be accepted, and the item released to storage or field operation.

More important, these quality defects can be overshadowed by an unknown number of latent defects. These latent defects, which ordinarily pass factory quality inspection, are due to flaws, either inherent to the parts or induced during fabrication, that weaken the fabricated hardware such that it will fail later under the proper condition of stress during field operation. Reliability screen tests (Environmental Stress Screening) are designed to apply a stress during manufacturing, at a given magnitude and over a specified duration, to identify these latent defects. As in the case of conventional quality inspections, screen tests designed to remove latent defects are not 100% effective.

It must be emphasized that reliability prediction and analysis methods, as discussed in Sections 6, 7, and 8, are based primarily on system design characteristics and data emphasizing the attribute characteristics of the constituent parts. Resulting estimates generally reflect the reliability potential of a system during its useful life period, i.e., during the period after early design when quality defects are dominant and prior to the time when wearout becomes dominant. They represent the inherent reliability, or the reliability potential, of the system as defined by its design configuration, stress and derating factors, application environment, and gross manufacturing and quality factors. A design-based reliability estimate does not represent the expected early life performance of the system, particularly as it is initially manufactured.

11.2.2.1 Factors Contributing to Reliability Degradation During Production: Infant Mortality

In order to assess the reliability of a system or equipment item during its initial life period (as well as during wearout), it is necessary to evaluate the components of failure that comprise its overall life characteristics curve. In general, the total distribution of failure over the life span of a large population of a hardware item can be separated into quality, reliability, wearout and design failures as shown in Table 11.2-3. These failure distributions combine to form the infant mortality, useful life, and wearout periods shown in Figure 11.2-3. It should be noted that design and reliability defects normally would exhibit an initially high but decreasing failure rate and that in an immature design these defects would dominate all other defects.

TABLE 11.2-3: FOUR TYPES OF FAILURES

QUALITY	Unrelated to operating stress	Eliminated by process control and inspection
RELIABILITY	Stress dependent	Minimized by screening
WEAROUT	Time dependent	Eliminated by replacement
DESIGN	May be stress and/or time dependent	Eliminated by proper application, derating, testing and failure data analysis

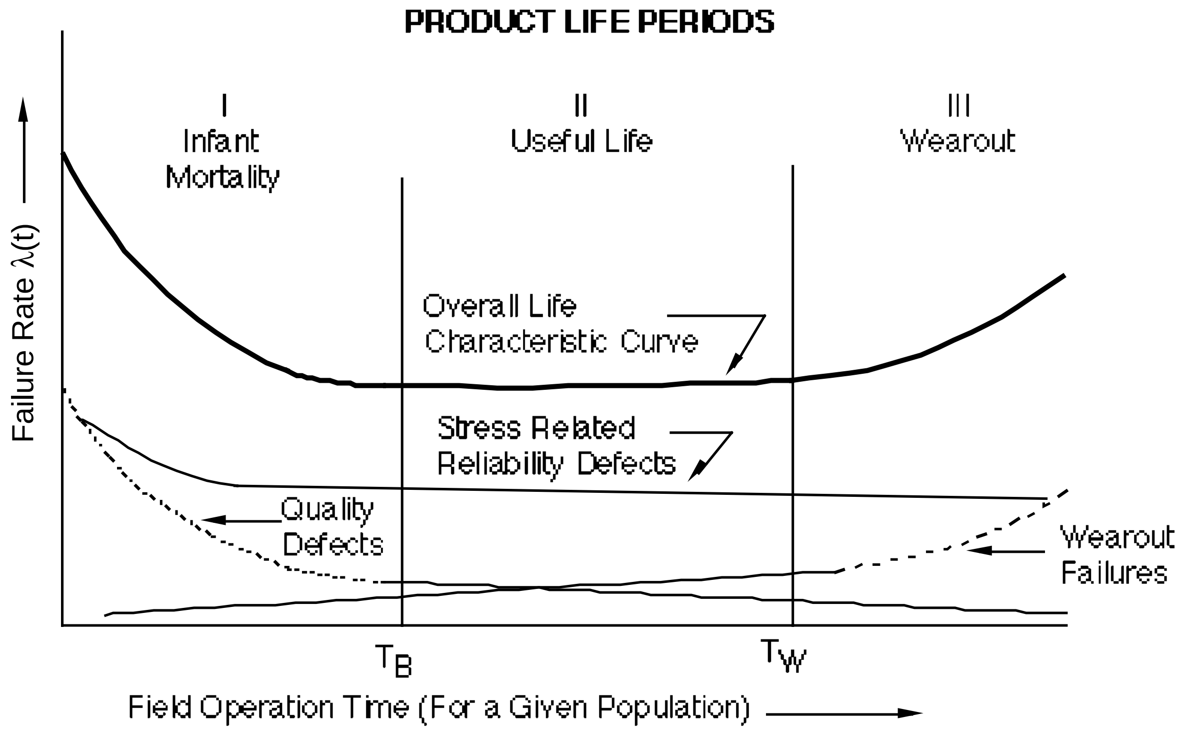


FIGURE 11.2-3: LIFE CHARACTERISTIC CURVE

As indicated in earlier sections, the general approach to reliability design for electronic equipment/systems is to address the useful life period, where failure rate is constant. Design is focused on reducing stress-related failures and generally includes efforts to select high quality, long life parts that are adequately derated.

For new items, this design-based approach in itself is not adequate to ensure reliability. Examination of Figure 11.2-3 shows that the infant mortality period consists of a high but rapidly decreasing quality-related failure distribution, a relatively high and decreasing latent stress-related (reliability) failure distribution, and a low but slightly increasing wearout-related failure distribution. Experience has shown that the infant mortality period can vary from a few hours to well over 1000 hours, although for most well designed, complex equipment it is seldom greater than 100 hours. The duration of this critical phase in reliability growth depends on the maturity of the hardware and, if not controlled, would dominate the overall mortality behavior, leaving the item without a significantly high reliability period of useful life. Positive measures must be taken, beginning with design, to achieve a stabilized low level of mortality (failure rate). This includes evaluating the impact of intrinsic part defects and manufacturing process-induced defects, as well as the efficiency of conventional inspections and the strength of reliability screening tests.

The intrinsic defects arise from the basic limitation of the parts that constitute the system or equipment and are a function of the supplier's process maturity, and inspection and test methods. Intrinsic (or inherent) reliability is calculated using design-based reliability prediction techniques (e.g., MIL-HDBK-217 methods described in Section 6).

The process-induced defects, as previously discussed, are those which enter or are built into the hardware as a result of faulty workmanship or design, process stresses, handling damage, or test efforts and lead to degradation of the inherent design-based reliability. Examples of the types of failures which may occur due to manufacturing deficiencies are poor connections, improper positioning of parts, contamination of surfaces or materials, poor soldering of parts, improper securing of component elements, and bending or deformation of materials.

These defects, as mentioned earlier, whether intrinsic to the parts or introduced during fabrication, can be further isolated into quality and reliability defects. Quality defects are not time dependent and are readily removed by conventional quality control measures (i.e., process control, inspections and tests). The better the process and the more efficient the inspection and test the more defects that are avoided or removed. However, since no test or inspection is perfect, some defects will escape to later manufacturing stages and then must be removed at a much higher cost or, more likely, pass through to field use and thus result in lower actual operational reliability with higher maintenance cost.

Stress/time dependent reliability defects cannot generally be detected (and then removed) by conventional QC inspections. These defects can only be detected by the careful and controlled application of stress screen tests. Screen tests consist of a family of techniques in which electrical, thermal, and mechanical stresses are applied to accelerate the occurrence of potential failures. By this means, latent failure-producing defects, which are not usually detected during normal quality inspection and testing, are removed from the production stream. Included among these tests are temperature burn-in, temperature cycling, vibration, on/off cycling, power cycling, and various nondestructive tests. Burn-in is a specific subclass of screens which employs stress cycling for a specified period of time. A discussion of screening and burn-in is presented in the next section.

As an example of two types of defects, consider a resistor with the leads bent close to its body. If the stress imposed during bending caused the body to chip, this is a quality defect. However, had the stress been inadequate to chip the body, the defect would go unnoticed by conventional inspection. When the body is cycled through a temperature range, small cracks can develop in the body. This would allow moisture and other gases to contaminate the resistive element, causing resistance changes. This is a reliability defect. Note that this defect can also be a design defect if the design specifications require a tight bend to fit the component properly in a board. However, if the improper bend is due to poor workmanship or process design, the defect is classified as a process-induced reliability defect. Consequently, the types of defects to which a system and its subsystems are susceptible are determined by the parts selected and their processing, while the presence of these defects in the finished item is a function of the quality controls, tests and screens that are applied.

Figure 11.2-4 pictorially shows the reliability impact of the part and process defects. As shown, an upper limit of reliability is established by design based on part derating factors, application environment, quality level, etc. The shaded area indicates that the estimated inherent reliability level may have a relatively broad range depending on the parts that comprise the system and the values for the parameters of the part failure estimating models.

The reliability of initially manufactured units will then be degraded from this upper limit; subsequent improvement and growth is achieved through quality inspections, reliability screening, failure analysis, and corrective action. The extent and rigor with which the tests, failure analysis and corrective actions are performed determine the slope of the reliability improvement curve. As such, process defects, along with the inherent part estimates, must be evaluated in order to accurately estimate reliability, particularly during initial manufacturing.

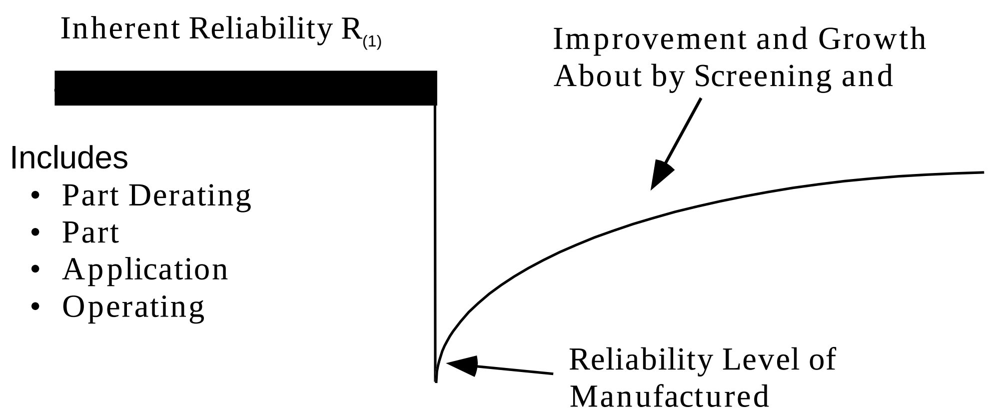


FIGURE 11.2-4: IMPACT OF DESIGN AND PRODUCTION ACTIVITIES ON EQUIPMENT RELIABILITY

11.2.2.2 Process Reliability Analysis

The infant mortality period (as was shown in Figure 11.2-3) is composed of a high but rapidly-decreasing quality component, a relatively high and decreasing stress component, and a low but slightly increasing wearout component. Because of this non-constant failure rate, this life period cannot be described simply by the single parameter exponential distribution; computation of reliability during this period is complex. It would require application of the Weibull distribution or some other multi-parameter distribution to account for the decreasing failure rate. Controlled life tests would have to be performed or extensive data compiled and statistically evaluated to determine the parameters of the distributions.

A practical approach, however, that would perhaps be useful during pre-production planning or during early production is to compute an average constant failure rate (or MTBF). This average MTBF represents a first approximation of the reliability during this early period. It can be viewed as a form of "step" MTBF, as shown in Figure 11.2-5 where the "step" MTBF includes both stress and quality failures (defects) at both the part and higher assembly levels, while the inherent MTBF (experienced during the useful life period) includes only stress related failure (defects) at the part level.

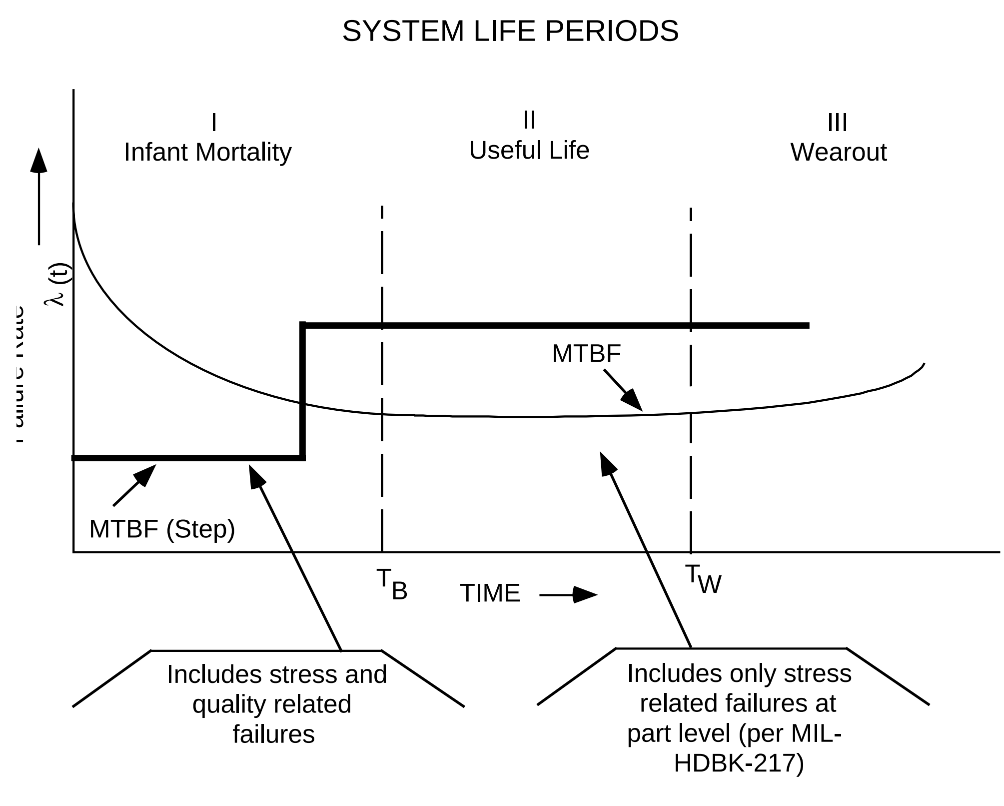


FIGURE 11.2-5: "STEP" MTBF APPROXIMATION

mechanical coupling, and mechanical resonant frequencies assure that differently configured assemblies will respond differently to identical thermal and vibrational inputs.

- Stress profiles should be defined in terms of responses rather than input, especially for vibration. A uniform level of stress may not be achieved throughout the unit, because units are not generally internally homogeneous. The response can be specified and measured at only a few points, so it will still differ locally within differently configured assemblies.

There are various approaches associated with the application of stress screens. Regardless of the approach used, the fundamental objective of ESS remains the same, i.e., to remove latent defects from the product prior to field delivery. The quantitative methods contained in MIL-HDBK-344 and the tri-service ESS guidelines extend this objective by focusing on the defects which remain in the product at delivery and their impact on field reliability.

11.2.4 Production Reliability Acceptance Testing (MIL-HDBK-781)

Reliability acceptance testing is performed on production hardware to determine compliance to specified reliability requirements. MIL-HDBK-781, "Production Reliability Acceptance Testing" contains all the essential procedures and requirements for designing an acceptance test plan for equipment that experiences a distribution of times-to-failure that is exponential. It defines test conditions, procedures and various test plans, and respective accept/reject criteria.

MIL-HDBK-781 has recently been completely revised to include detailed information for test planning and evaluation of data. The latest revision has been restructured to make extensive use of appendices to expand and clarify the various sections of the handbook. It clarifies the definition of mean-time-between-failures (MTBF) and the use of θ_0 (upper test MTBF) and θ_1 (lower test MTBF) , which are test planning parameters, and specifies the use of combined environmental test conditions (temperature, vibration and moisture)* based on the actual mission profile environments encountered during the equipment's useful life.

MIL-HDBK-781 is not intended to be used on a blanket basis, but each requirement should be assessed in terms of the need and mission profile. Appendices are designed so that the procuring activity may reference them along with specific parts of the handbook.

MIL-HDBK-781 covers requirements for preproduction qualification tests as well as production acceptance tests. Qualification tests are normally conducted after growth tests in the development cycle, using initial production hardware to make a production release decision. It should be

* Altitude may be included if the procuring activity determines that it is cost effective, but the cost of test facilities for combining altitude with the other environments would probably be prohibitive.

emphasized that qualification testing, conducted per MIL-HDBK-781, is to demonstrate reliability with statistical confidence, whereas reliability growth testing is performed to improve reliability. Depending on program requirements, funding, and other constraints, preproduction testing may maximize growth testing and minimize statistical testing (resulting in a high MTBF at a low confidence) or may minimize growth and maximize demonstration (resulting in a lower MTBF at a high confidence). Preproduction testing, including both reliability growth and qualification, was discussed in detail in Section 8.

Production reliability acceptance tests per MIL-HDBK-781 are described as “a periodic series of tests to indicate continuing production of acceptable equipment” and are used to indicate individual item compliance to reliability criteria. The tests are intended to simulate in-service evaluation of the delivered item or production lot and to provide verification of the inherent reliability parameters as demonstrated by the preproduction qualification tests. Therefore, an equipment would undergo qualification testing on preproduction hardware.

Once the specified reliability has been demonstrated, and after production begins, the lots produced would undergo reliability acceptance testing, usually at a stress less stringent than the demonstration test level, to indicate continuing fulfillment of reliability requirements.

Production Reliability Acceptance Testing per MIL-HDBK- 781 can be performed based on sampling an equipment from each lot produced as well as on all equipment produced. The test conditions, or stress profile, applied during the test should be measured (preferred) or estimated by the procuring activity and incorporated into the equipment specification. However, when the stress types and levels are not specified by the procuring activity and when measured environmental stresses for the proposed application or a similar application are not available for estimating, then the stress types and levels given in Table 11.2-8, taken from MIL-HDBK-781, should be applied. Table 11.2-8 provides a summary of combined environmental test condition requirements applicable to the following categories of equipment classification:

Category 1: Fixed ground equipment

Category 2: Mobile ground vehicle equipment

Category 3: Shipboard equipment
- sheltered
- unsheltered

Category 4: Equipment for jet aircraft

Category 5: Turbo-prop aircraft and helicopter equipment

Category 6: Air-launched weapons and assembled external stores

Figure 11.2-12, also taken from MIL-HDBK-781, illustrates a typical test cycle that shows the timing of the various conditions. MIL-HDBK-781 describes standard statistical test plans covering:

(1) Fixed length test plans (Test Plans IXC through XVIIC and XIXC through XXIC)

(2) Probability ratio sequential tests (PRST) (Test Plans IC through VIC)

(3) Short run high risk PRST plans (Test Plan VIIC and VIIIC)

(4) All equipment reliability test (Test Plan XVIIIC)

Accept/reject criteria are established on θ_1 and θ_0, where θ_1, the lower test MTBF, is an unacceptable MTBF based on minimum requirements. θ_0 is the upper test MTBF, or the acceptable MTBF. The ratio θ_0/θ_1 is defined as the discrimination ratio. Specifying any two of these three parameters, given the desired producer and consumer decision risks, determines the test plan to be utilized.

Test Plan XVIIIC, shown in Figure 11.2-13, can be used for 100% production reliability acceptance testing. This test plan is to be used when each unit of production (or preproduction equipment if approved by the procuring activity) equipment is to be given a reliability acceptance test. The plan consists of a reject line and a boundary line. The reject and boundary lines are extended as far as necessary to cover the total test time required for a production run. The equation of the reject line is $f_R = 0.72T + 2.50$ where T is cumulative test time in multiples of θ_1 and f is cumulative number of failures. The plotting ordinate is failures and the abscissa is in multiples of θ_1, the lower test MTBF. The boundary line is 5.67 failures below and parallel to the rejection line. Its equation is $f_B = 0.72T - 3.17$.

The test duration for each equipment shall be specified in the test procedure as approved by the procuring activity. The maximum duration may be 50 hours and the minimum 20 hours to the next higher integral number of complete test cycles. If a failure occurs in the last test cycle, the unit shall be repaired and another complete test cycle run to verify repair.

An optional nonstatistical plan can also be used for production reliability acceptance testing. Its purpose is to verify that production workmanship, manufacturing processes, quality control procedures, and the assimilation of production engineering changes do not degrade the reliability, which was found to be acceptable by the reliability qualification test. The test is to be applied to all production items with the item operating (power applied). The required test duration and number of consecutive, failure free, thermal test cycles (minimum of two) which each deliverable item must exhibit is specified by the procuring activity. The vibration, temperature cycling, and moisture environments together with any others which are deemed necessary may be applied sequentially. The equipment duty cycle and the sequence, duration, levels of the environments, and the vibration option to be used in this test require approval of the procuring activity and are submitted in accordance with the test program requirements.

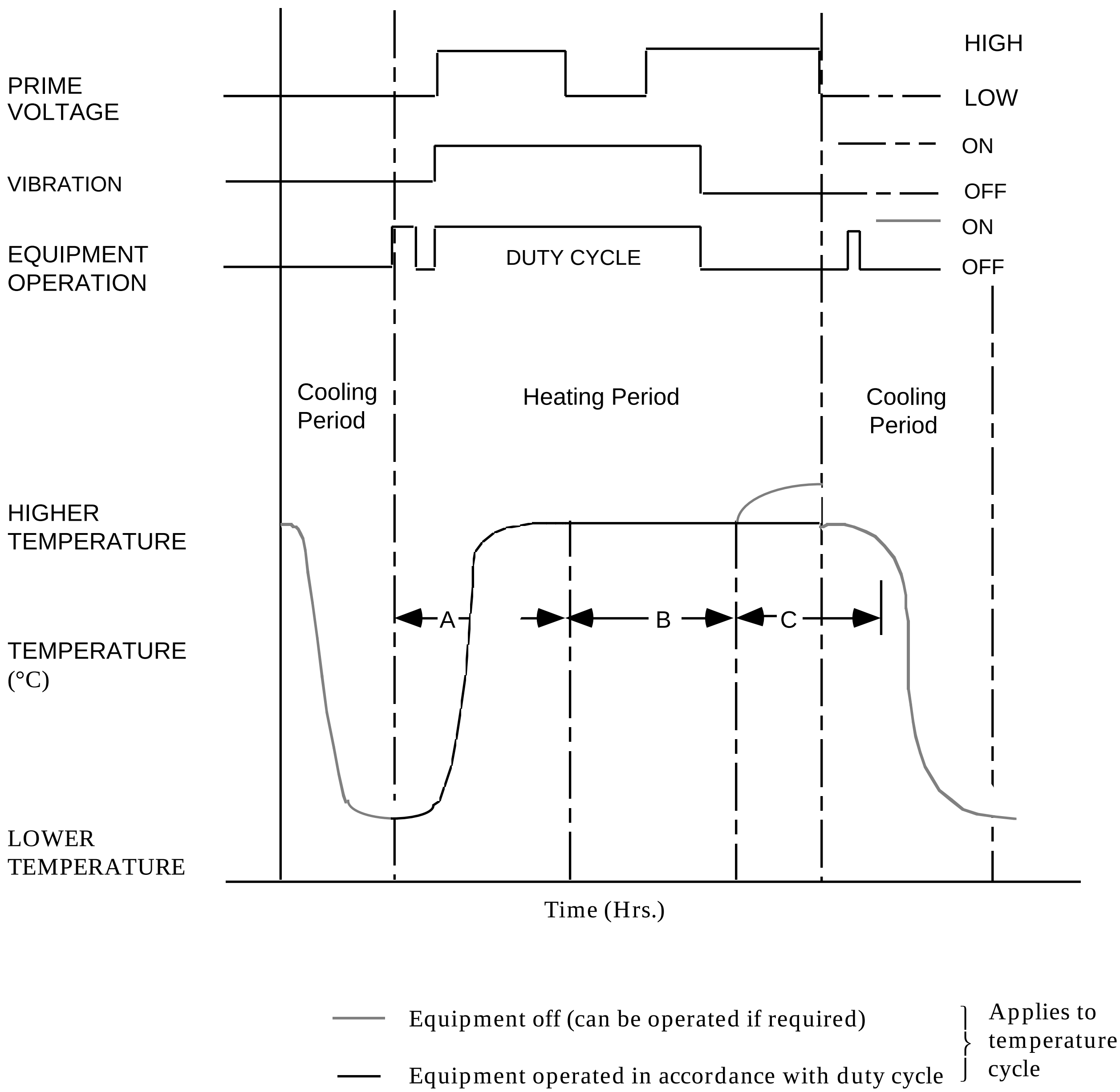
TABLE 11.2-8: TEST CONDITIONS MATRIX (TAKEN FROM MIL-HDBK-781)

Summary of Combined Environmental Test Condition Requirements

	FIXED GROUND	GROUND VEHICLE	SHIPBOARD SHELTERED	SHIPBOARD UNSHELTERED
ELECTRICAL STRESS				
Input voltage	Nominal +5%-2%	Nominal ± 10%	Nominal ±7%*	Nominal ± 7%*
Voltage cycle	high, nominal and low →	one per test cycle	————	————→
VIBRATION STRESS				
Type vibration	sinewave single frequency	swept-sine log sweep	swept-sine ** continuous	swept-sine** continuous
Amplitude	(See APPENDIX B for	stress levels)		
Frequency range***	20 to 60 Hz	5 to 500 Hz		
Application	20 minimum per equipment	sweep rate 15 minimum once/hr.	(See APPENDIX B ———	————→)
THERMAL STRESS (°C)	A B C ****	LOW HIGH	LOW HIGH	LOW HIGH
Storage temperature	- - -	-54 85	-62 71	-62 71
Operating temperature	20 40 60	-40 TO55	0 TO 50 (CONTROLLED)	-28 65
Rate off change	- - -	5°/min.	5°/min.	5°/min.
Maximum rate of change	- - -	10°/min.	10°/min.	10°/min.
MOISTURE STRESS				
Condensation	none	1/test cycle	See APPENDIX B	1/test cycle
Frost/freeze		1/test cycle		1/test cycle

	AIRCRAFT FIGHTER	AIRCRAFT TRANSPORT, BOMBER	AIRCRAFT HELICOPTER	AIRCRAFT TURBO-PROP	AIR-LAUNCHED WEAPONS AND ASSEMBLED EXTERNAL STORES
ELECTRICAL STRESS					
Input voltage range	nominal ± 10%	± 10%	± 10%	± 10%	± 10%
Voltage cycle	(nominal, high and	low voltage, one cycle	/thermal cycle or per	APPENDIX B)	
VIBRATION STRESS					
Type vibration	random	random	swept-sine log-sweep	swept-sine	swept-sine***
Amplitude	(————	———— SEE	APPENDIX B ——	————	————→)
Frequency range	200 - 2000 Hz	20 - 2000 Hz	5 - 2000 Hz****	10 - 2000 Hz	20 - 2000 Hz
Application	continuous	continuous	sweep rate 15 min. one/hr	continuous (See APPENDIX B)	continuous (See MIL-STD-1670)
THERMAL STRESS (°C)	LOW HIGH	LOW HIGH	LOW HIGH	LOW HIGH	LOW HIGH
Storage temperature (non-oper.)	-54 +71	-54 +71	-54 +71	-54 +71	-65 +71
Operating temperature range	(←————	—— SEE APPENDIX	B ————	————	————→)
Rate of change (min.)	5°/min.	5°/min.	5°/min.	5°/min.	
Duration (nominal)	3 1/2 hours	3 1/2 hours	3 1/2 hours	3 1/2 hours	
MOISTURE STRESS					
Condensation	(1/test cycle ----------	------------------------	------------------------	------------------------	--------------------------)
Frost/freeze	(1/test cycle ----------	------------------------	------------------------	------------------------	--------------------------)

* See MIL-STD-1399
** See MIL-STD-167-1
*** Frequency tolerance ±2 percent or ±0.5 Hz for frequencies below 25 Hz.
**** See 50.1.4 of Appendix B



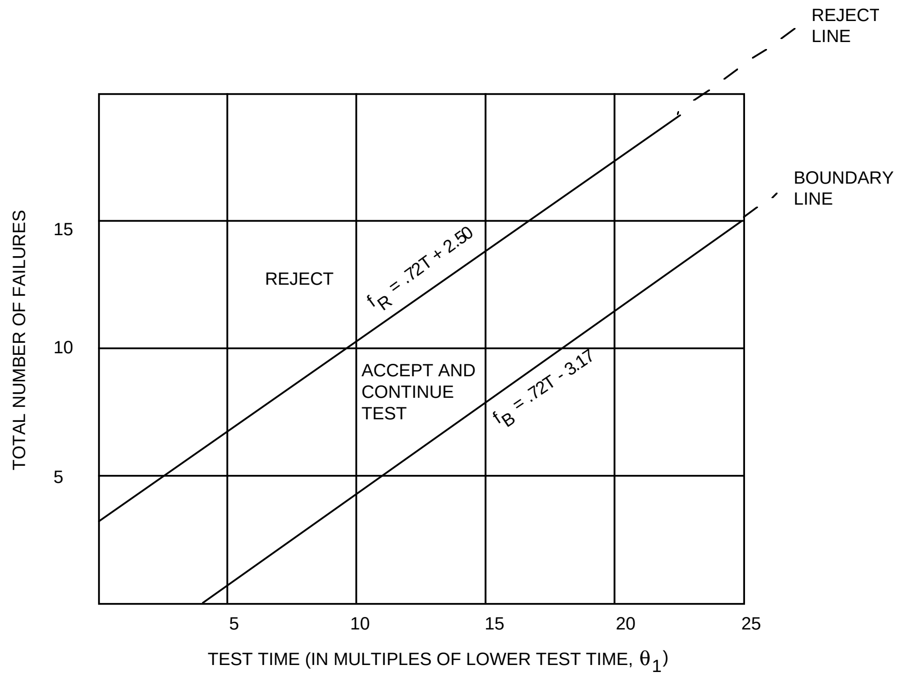
A. Time for chamber to reach stabilization at higher temperature

B. Time of equipment operation at higher temperature

C. Optional Hot Soak and hot start-up checkout

FIGURE 11.2-12: SAMPLE ENVIRONMENTAL TEST CYCLE

SECTION 11: PRODUCTION AND USE (DEPLOYMENT) R&M



Total Test Time*			Total Test Time*		
Number of Failures	Reject (Equal or less)	Boundary Line	Number of Failures	Reject (Equal or less)	Boundary Line
0	N/A	4.40	9	9.02	16.88
1	N/A	5.79	10	10.40	18.26
2	N/A	7.18	11	11.79	19.65
3	.70	8.56	12	13.18	21.04
4	2.08	9.94	13	14.56	22.42
5	3.48	11.34	14	etc.	etc.
6	4.86	12.72	15	.	.
7	6.24	14.10	16	.	.
8	7.63	15.49	.	.	.

* Total test time is total unit hours of equipment on time and is expressed in multiples of the lower MTBF. Refer to 4.5.2.4 for minimum test time per equipment.

FIGURE 11.2-13: REJECT-ACCEPT CRITERIA FOR TEST PLAN IXC

It must be emphasized that test criteria, including confidence level or decision risk, should be carefully selected and tailored from these documents to avoid driving cost or schedule without improving reliability. Some general guidelines, taken from MIL-HDBK-781 for planning and implementing production reliability acceptance testing are as follows:

- Production reliability acceptance testing must be operationally realistic, and may be required to provide estimates of demonstrated reliability.

- The statistical test plan must predefine criteria of compliance ("accept") which limit the probability that the item tested, and the lot it represents, may have a true reliability less than the minimum acceptable reliability. These criteria must be tailored for cost and schedule efficiency.

- Production reliability acceptance testing provide a basis for positive and negative financial feedback to the contractor, in lieu of an in-service warranty.

- Production reliability acceptance testing may require expensive test facilities to simulate the item life profile and operational environment; therefore, all equipment production reliability acceptance testing (100% sampling) is not recommended.

- Because it provides a basis for determining contractual compliance, and applies to the items actually delivered to operational forces, production reliability acceptance testing must be independent of the supplier, if at all possible.

- Sampling frequency should be reduced after a production run is well established, however, the protection that it provides for the government (and the motivation it provides for the contractor's quality control program) argues against complete waiver of the production reliability acceptance testing requirement.

Plans for performing production reliability acceptance testing are incorporated into the overall reliability test plan document, and should encompass the following considerations:

(1) Tests to be conducted

(2) Reliability level (i.e., MTBF) to be demonstrated, as well as the associated confidence level, and the relationship between demonstrated MTBF, confidence, test time, etc.

(3) Representative mission/environmental profile

(4) The number of units for test, expected test time, calendar time factors, and scheduling of effort

(5) The kinds of data to be gathered during the test

(6) Definition of failure (relevant, nonrelevant)

(7) Authorized replacement and adjustment actions

(8) Logs/data forms to be maintained that record number of units on test, test time accumulated, failures, corrective actions, statistical decision factors, and accept/reject criteria

11.2.5 Data Collection and Analysis (During Production)

The production reliability test and control program, once implemented in the factory, should continually be challenged relative to the effectiveness of the overall program, as well as that of the individual tests. Production screening and acceptance testing is a dynamic process which must be continually modified in response to experience. Test results and field experience data are monitored to determine the need to modify individual test criteria and conditions to reduce the sampling frequency of acceptance tests and to identify the possibility of applying earlier screen tests where the test costs are less and the potential for cost avoidance is higher. It should be emphasized that the production program, as initially planned, represents a baseline for applying the tests. A production screen test, for example, like any quality inspection, must be adjusted depending on the results of subsequent higher level tests or field performance. However, the extent and nature of any changes should be determined only through careful review and analysis of the subsequent failures.

A data system supported by failure analysis and corrective action is established to maintain visibility over the effectiveness of the production test program as well as all tests including development, qualification, and production. The data system is designed to compile test and failure data and to provide information that would provide a basis to change the test program as necessary to minimize cost and maximize effectiveness. A failure reporting, analysis and corrective action system (FRACAS) is an essential element of the production test program as well as the overall reliability control program. A well designed FRACAS system will provide a uniform mechanism for reporting failures, determining causes and remedies, and making these findings known to the appropriate engineers and designers to enable them to formulate and implement corrective action and, specifically, to ascertain whether or not to design and implement improved inspection, screening and acceptance tests.

Section 8 of the handbook describes failure reporting, analysis, corrective action, and the provisions necessary to assure that failures are accurately reported, thoroughly analyzed, and that corrective actions are taken on a timely basis to reduce or prevent recurrence.

The results of production acceptance test, screening and inspection results, as well as failure reports and analyses from the FRACAS program, are compiled and incorporated into the data system. Maintaining accurate and up-to-date records through a formal data recording and analysis system is particularly essential in tracking and assessing field reliability performance. Comparative evaluation between predicted reliability estimates and actual field reliability provides criteria for improving production acceptance testing (including the screening and burn-in testing procedures) to assure that the most cost effective test program is developed and applied. This is especially important for new systems where changing performance and reliability characteristics would be expected as a result of design and manufacturing improvements.

A properly designed and operating data system would provide the following information as it pertains to production testing:

(1) Identification of hardware subjected to production tests

(2) Total cumulative operating time for each hardware item including the last operating time interval of failure free operation and acceptance test completion dates

(3) Sampling frequency of reliability acceptance tests

(4) Failure reports of hardware discrepancies including description of failure effects and accumulated operating hours to time of failure

(5) Failure analysis reports of hardware discrepancies including cause and type of failure modes

Also, cumulative plots of screening and burn-in failure events versus time can be prepared and maintained and periodic summary reports submitted to engineering and management activities that provide:

(1) Failure/reject rates by test type and level

(2) Screen test efficiency factors

(3) Responsible failure mechanisms

(4) Recommended or accomplished corrective actions

(5) General product reliability analysis that correlates design predictions with test results and field experience of parts, contamination of surfaces or materials, poor soldering of parts, improper securing of component elements, and bending or deformation of materials

These defects, as mentioned earlier, whether intrinsic to the parts or introduced during fabrication can be further isolated into quality and reliability defects. Quality defects are not time dependent and are readily removed by conventional quality control measures (i.e., inspections and tests). The more efficient the inspection and test the more defects that are removed. However, since no test or inspection is perfect, some defects will escape to later manufacturing stages and then must be removed at a much higher cost or, more likely, pass through to field use and thus result in lower actual operational reliability with higher maintenance cost.

11.2.6 Monitor/Control of Subcontractors and Suppliers

The monitoring of subcontractors is a critical, but often overlooked function of a successful reliability program. End product reliability and its life cycle cost can be adversely affected if a sub-tier vendor or major subcontractor does not fully comply with the applicable reliability program requirements.

The requirements for the monitoring of subcontractors and the monitoring of suppliers often differs due to the nature of the product being furnished and may therefore frequently be defined separately.

11.2.6.1 Major Subcontractor and Manufacturer Monitoring

Development-phase subcontractor monitoring is accomplished by reviewing design data, reliability data, parts selection, non-standard parts requests, failure reports, periodic attendance at design reviews and participation in reliability problem resolution. Production-phase monitoring consists of verifying adherence to the Quality Assurance (QA) standard established between the prime contractor and the subcontractor. It should include the review of process control, production control, personnel qualifications, workmanship and participation in the FRACAS (see section 8.2).

Normally, except for off-the-shelf procurements, the requirements imposed on the manufacturer of a unit/major assembly is as specified in the prime item/system specification.

Supplier monitoring/control requires detailed inspection of the material being furnished, verification of QA procedures, critique of manufacturing processes, periodic inspection to verify adherence to the quality standard, identification of problems, incoming inspection, testing and performance tracking.

Monitoring of Parts Suppliers requires review of vendor performance in addition to the tasks noted.

11.2.6.2 Establishing Vendor Capability and Program Reviews

The most direct method of determining a vendor capability is to review past performance. If this data is not available or is incomplete, a facility survey should be performed.

Program reviews should be conducted on a continuous basis for the life of the contract. It is essential to develop a free exchange of information and data so that the prime contractor has maximum program visibility into a vendor's process methods and performance. These reviews verify that the manufacturing process is under control, and that: workmanship, personnel certification training, and testing, as defined in the equipment specification and QA manual, is being implemented correctly.

Failure report data from production tests (Burn-in, ESS, PRAT, etc.) received from a vendor or as a result of in-house testing should be reviewed for failure trends and possible corrective action.

11.2.6.3 Supplier Monitoring

Monitoring and verification require that the prime contractor and the selected vendors have a complete and mutual understanding of the standards and quality requirements imposed. A requirements data package should be supplied to each vendor. These data sets then form the foundation for mutual agreement regarding the requirements of a purchase. It is also essential to establish measurement compatibility between the prime contractor's inspection department and the vendor's inspection department should conflicts arise (i.e., a part tests good at vendor final inspection and fails incoming inspection).

Monitoring requirements may vary with the type of procurement (off-the-shelf purchase, etc.). Therefore it is important to assess and plan for the effort that will be required, to define the monitoring requirement associated with the various types of procurement. For example, if component parts are procured from a distributor then the monitoring should consist of verifying that QA and Reliability requirements are developed, that Certificates of Compliance are available, that the Defense Material Administration is monitoring, etc.

11.3 Production Maintainability Control

As was previously indicated for reliability, the inherent design maintainability of an equipment/system can also be degraded during production unless adequate controls are specified and applied to prevent this degradation. This topic is addressed in detail in a companion document MIL-HDBK-470A, "Military Handbook: Designing and Developing Maintainable Products and Systems."

11.4 Reliability and Quality During Shipment and Storage

Electronic components and equipment are subject to change, damage, deterioration and performance degradation during shipment and while in storage. Consequently, the identification of significant defects, the quantification of the rate at which defects occur, and the analysis of deterioration influenced by shipment and storage environments, dormancy, storage testing, and environmental cycling effects are essential to minimize performance degradation and to assure the designed hardware reliability. Specific inspections and analyses to predict the effects of shipment and storage, to assess the in-storage functional status of component and equipment items, and to control deterioration mechanisms are performed as part of the overall life-cycle reliability program. Included are efforts applicable to:

(1) New Items - determine the environmental conditions needed for proper storage and the effects of shipment, storage and handling on reliability.

(2) Items in Storage - generate storage reliability control techniques covering receipt, storage and prior-to-issue phases of material and equipment items.

The control efforts include identifying components and equipment (and their major or critical characteristics) which deteriorate during shipment and with storage and preparing procedures for in-storage cycling inspection to assure reliability and readiness. The inspection procedures are to identify the number of items for test and the acceptable levels of performance for the parameters under test. Results of these efforts are used to support long term failure rate predictions, design trade-offs, definition of allowable test exposures, retest after storage decisions, packaging, handling, or storage requirements, and refurbishment plans.

11.4.1 Factors Contributing to Reliability Degradation During Shipment & Storage

Defects can be induced during shipment because (1) the packing protection was not adequate for the mode of transportation, (2) the container or other packaging material did not meet specification requirements, or (3) the equipment was roughly handled or improperly loaded.

Electronic components age and deteriorate over long storage periods due to numerous failure mechanisms. In particular, the electrical contacts of relays, switches, and connectors are susceptible to the formation of oxide or contaminant films or to the attraction of particulate matter that adheres to the contact surface, even during normal operation. During active use, the mechanical sliding or wiping action of the contacts is effective in rupturing the films or dislodging the foreign particles in a manner which produces a generally stable, low resistance contact closure. However, after long periods of dormant storage, the contaminant films and/or the diversity of foreign particles may have increased to such an extent that the mechanical wiping forces are insufficient for producing a low resistance contact.

The formation of contaminant films on contact surfaces is dependent on the reactivity of the control material, its history, and the mechanical and chemical properties of the surface regions of the material. Gold is normally used whenever maximum reliability is required, primarily because gold is almost completely free of contaminant oxide films. Even gold, however, is susceptible to the formation of contaminant films by simple condensation of organic vapors and the deposition of particulate matter. Silver is highly susceptible to the sulfide contaminants that abound in the atmosphere, as are alloys of copper and nickel. Shipping and storage of these systems in paper boxes should be avoided because such boxes contain small amounts of sulfur. Particulate contamination can also lead to corrosive wear of the contact surfaces when the particle is hygroscopic. With this condition, water will be attracted to the contact surface and can lead to deterioration through corrosive solutions or localized galvanic action. The source of such particles can be directly deposited airborne dust or wear debris from previous operations.

Another failure mode which may become significant after long term storage is the deterioration of lubricants used on the bearing surfaces of relays, solenoids, and motors. Lubricants can oxidize and form contamination products. Similarly, lubricants can also attract foreign particles, particularly when exposed to airborne dust, and can lead to lubrication failures and excessive wear.

Over a period of time, many plastics (such as those used in the fabrication of electronic components, i.e., integrated circuits, capacitors, resistors, transistors, etc.) lose plasticizers or other constituents which may evaporate from the plastic, causing it to become brittle, and possibly, to shrink. This can cause seals to leak, insulation to break down under electrical/mechanical stress, and other changes conducive to fatigue and failures. Additionally, plastics may continue to polymerize after manufacture. That is, the structure of the molecules may change without any accompanying change in chemical composition. This will result in change in characteristics and physical properties.

Many materials slowly oxidize, combine with sulfur or other chemicals, or break down chemically over a period of time. These changes may affect electrical resistivity, strength, etc. In addition, many of these materials when exposed to condensed moisture or high humidity conditions may, through a leaching process, lose essential ingredients such as fire retardant additives, thereby causing a hazard to slowly develop. Other materials, such as explosives and propellants, may become unstable over time, posing a safety hazard.

Many component parts and assemblies are sensitive to contaminants and, thus, are sealed during manufacture. These seals will often leak, partly as a result of flexing due to changing temperature and atmospheric pressure, allowing air, moisture or other contaminants to reach the active portions of the component. This leakage can be so slow that the effects may not be discernible for years, but ultimately significant changes can occur.

Finally, the methods/materials of preservation, packaging, and packing (PP&P) used in the storage of components and equipment, i.e., cardboards, plastic bags, polystyrenes, etc., themselves may react with the items stored and cause decomposition and deterioration when left dormant for long durations.

Rough handling during shipment and depot operations, aging, and deterioration mechanisms as discussed above can, if uncontrolled, lead to a variety of component and equipment failure modes. A summary of some of the failure modes encountered with electronic components during storage is given in Table 11.4-1. Protective measures must be applied to isolate the components from the deteriorative influences in order to eliminate or reduce failure modes such as those listed in Table 11.4-1 and others that can be induced during shipment and storage.

11.4.2 Protection Methods

Proper protection against damage to and deterioration of components and equipment during shipment and storage involves the evaluation of a large number of interactive factors and the use of tradeoff analysis to arrive at a cost effective combination of protective controls. These factors can be grouped into four major control parameters: (1) the level of preservation, packaging and packing (PP&P) applied during the preparation of material items for shipment and storage; (2) the actual storage environment; (3) the need and frequency of in-storage cyclic inspection; and (4) the mode of transportation. These parameters, as depicted in Figure 11.4-1 (circled numbers), must be evaluated and balanced to meet the specific characteristics of the individual equipment and material items. The significance of each of the three parameters is as follows:

(1) Preservation, packaging and packing (PP&P) is the protection provided in the preparation of material items for shipment and long term storage. *Preservation* is the process of treating the corrodible surfaces of a material with an unbroken film of oil, grease, or plastic to exclude moisture. *Packaging* provides physical protection and safeguards the preservative. In general, sealed packaging should be provided for equipment, spare parts, and replacement units shipped and placed in storage. *Packing* is the process of using the proper exterior container to ensure safe transportation and storage.

    Various levels of *PP&P* can be applied, ranging from complete protection against direct exposure to all extremes of climatic, terrain, operational, and transportation environments (without protection other than that provided by the PP&P) to protection against damage only under favorable conditions of shipment, handling and storage. A military package as defined per MIL-E-17555, "*Electronic and Electrical Equipment, Accessories, and Provisioned Items (Repair Parts): Packaging of;*" is the degree of preservation and packing which will afford adequate protection against corrosion, deterioration, and physical damage during shipment, handling, indeterminate storage, and worldwide redistribution. A minimum military package is the degree of preservation and packaging which will afford adequate protection against corrosion, deterioration and physical damage during shipment from supply source to the first receiving activity, for immediate use or controlled humidity storage. Many times a minimum military package conforms to the supplier's commercial practice.

TABLE 11.4-1: FAILURE MODES ENCOUNTERED WITH ELECTRONIC COMPONENTS DURING STORAGE

COMPONENT	FAILURE MODES
Batteries	Dry batteries have limited shelf life. They become unusable at low temperatures and deteriorate rapidly at temperatures above 35°C. The output of storage batteries drops as low as 10 percent at very low temperatures.
Capacitors	Moisture permeates solid dielectrics and increases losses which may lead to breakdown. Moisture on plates of an air capacitor changes the capacitance.
Coils	Moisture causes changes in inductance and loss in Q. Moisture swells phenolic forms. Wax coverings soften at high temperatures.
Connectors	Corrosion causes poor electrical contact and seizure of mating members. Moisture causes shorting at the ends.
Relays and Solenoids	Corrosion of metal parts causes malfunctioning. Dust and sand damage the contacts. Fungi grow on coils.
Resistors	The values of composition-type fixed resistors drift, and these resistors are not suitable at temperatures above 85°C. Enameled and cement-coated resistors have small pinholes which bleed moisture, accounting for eventual breakdown. Precision wire-wound fixed resistors fail rapidly when exposed to high humidities and to temperatures at about 125°C.
Semiconductors, Diodes, Transistors, Microcircuits	Plastic encapsulated devices offer poor hermetic seal, resulting in shorts or opens caused by chemical corrosion or moisture.
Motors, Blowers, and Dynamotors	Swelling and rupture of plastic parts and corrosion of metal parts. Moisture absorption and fungus growth on coils. Sealed bearings are subject to failure.
Plugs, Jacks, Dial-Lamp Sockets, etc.	Corrosion and dirt produce high resistance contacts. Plastic insulation absorbs moisture.
Switches	Metal parts corrode and plastic bodies and wafers warp due to moisture absorption.
Transformers	Windings corrode, causing short or open circuiting.

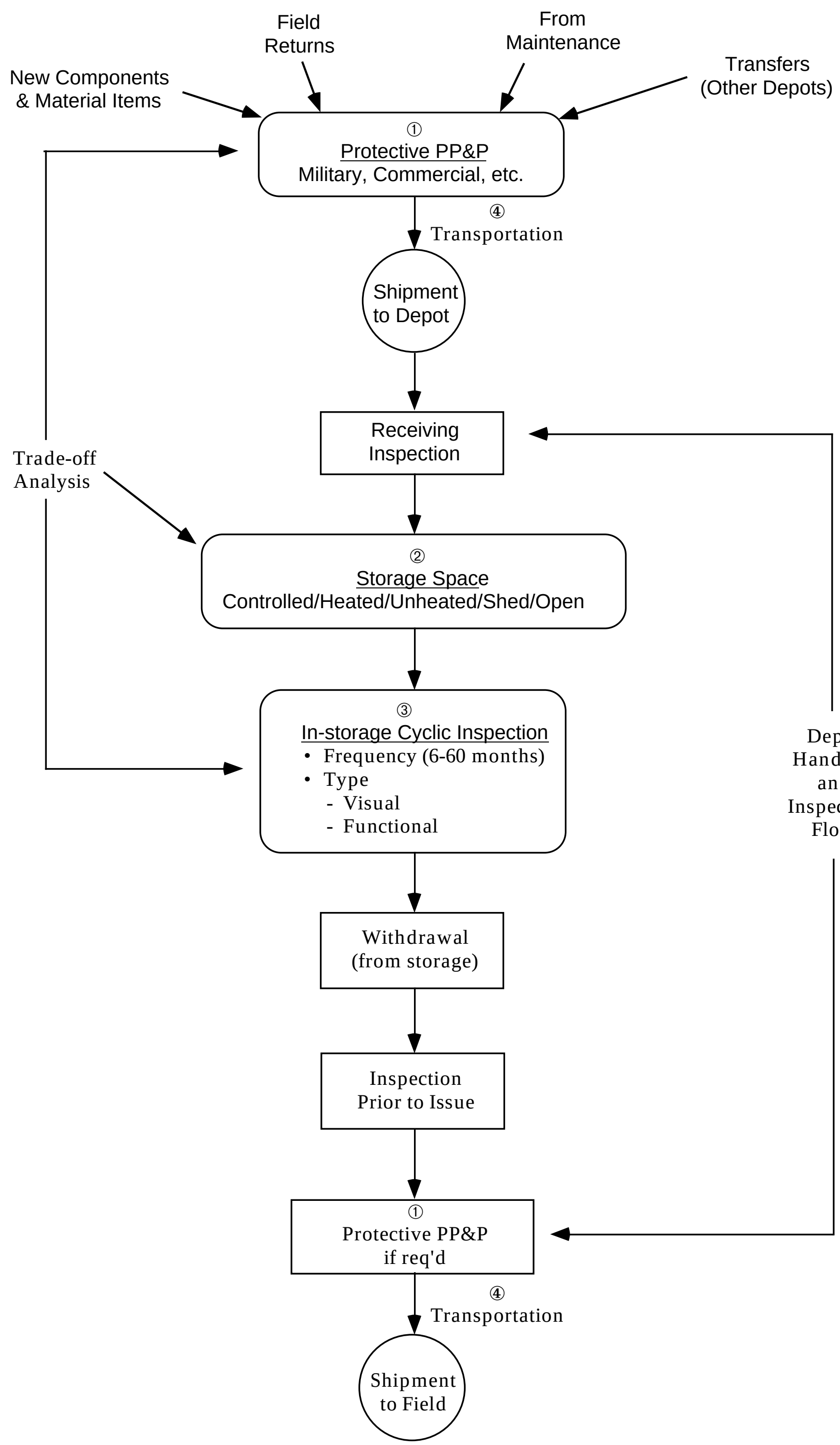


FIGURE 11.4-1: PROTECTIVE CONTROL DURING SHIPMENT AND STORAGE

(2) The storage environment can vary widely in terms of protection afforded. However, whenever possible, electronic hardware should be stored in dry, well ventilated warehouses, where the temperature of the air surrounding the equipment can be regulated so that it does not fall to dewpoint values at night. Storage in controlled temperature/humidity buildings is of course, ideal. If equipment is stored in bins, it is important that it be placed above floor level. The military has several types of storage areas. These include warehouse space with complete temperature and humidity control, warehouse space with no humidity and temperature control, sheds, and open ground areas that are simply designated for storage.

(3) In-storage scheduled cyclic inspection is the key to assuring the actual reliability of components and equipment during storage. In-storage cycling inspections are designed to detect performance degradation, deterioration, and other deficiencies caused by extended periods of storage and improper storage methods. The inspections are to identify those items which require corrective packaging (or further storage control) or condition reclassification to a lesser degree of serviceability. The inspections are performed at intervals derived from shelf life periods and the level of protective packaging and storage afforded the material items. It should be noted that all items when originally placed in storage are ready for issue and that all applicable preservation, packaging and packing (PP&P) requirements have been met. In-storage cycling inspection is part of the depot's overall inspection system (see Figure 11.4-1) that includes inspection of items at receipt as well as prior to issue.

In general, shipment and storage degradation can be controlled in terms of the above-mentioned three parameters. The planning and specification of shipment and storage requirements for new component and equipment items (as well as the reestablishment of requirements for existing items in storage) must take into account economic choices between the various factors within these parameters to arrive at the most cost effective balance that meets reliability and readiness objectives.

(4) The Mode of Transportation greatly influences the level of PP&P needed for an item. The modes of transportation used for military systems are primarily:

- aircraft
- surface ship
- rail
- truck

Each mode is characterized by a unique set of environmental factors. Truck and transport rail, for example, pose a certain temperature and vibration spectrum than do aircraft or surface ships. Exposure times also vary; item shipped by air are exposed to the environmental stresses of transport for a much shorter time than items transported by rail or surface ship. Items shipped by rail may sit in switching yards or sidings for days under all kinds of climatic conditions. Similarily, an item shipped by air may sit crated on the tarmac under extremes of heat or cold. Items aboard ships may be exposed to highly corrosive salt water spray.

Complicating matters is the fact that most items are not transported from origin to delivery point via a single mode of transportation. An item may, for example, be picked up at its point of origin by truck, driven to a rail loading dock, taken by train to a seaport, sent by ship to another port, downloaded to a truck, and then delivered to its final destination. Such multi-modal transportation imposes a greater variety of environmental stresses. In addition, the handling involved in switching between modes imposes its own set of stresses. The level of PP&P must be sufficient to protect the item against the most severe stresses to which it will be subjected throughout the transportation process.

11.4.3 Shipment and Storage Degradation Control (Storage Serviceability Standards)

Since electronic components and equipment are subject to damage, deterioration and performance degradation if unprotected during shipment and left uncontrolled for long periods of dormant storage, organizations have established programs to control the parameters defined above. The Army, for example, has established the Care of Supplies in Storage (COSIS) program (Ref. [4]). The program assures that material is maintained in a condition to meet supply demands at a minimum cost in funds, manpower, facilities, equipment, and materials. COSIS by definition is "a Department of the Army (DA) program to perform specific tasks to assure that the true condition of material in storage is known, properly recorded, and the material is provided adequate protection to prevent deterioration. The distinction between COSIS-related actions and actions that might otherwise fall into the broad category of care given material in storage is that COSIS concerns itself with the in-storage inspection, minor repair, testing, exercising of material and the preservation, packaging and packing (PP&P) aspects of the efforts."

A major and most significant element within the COSIS program is the Storage Serviceability Standards (SSS) documents controlled by participating Army commodity commands as required by DARCOM-R 702-23, *"Product Assurance - Storage Serviceability Standards (SSSs),"* (Ref. [5]). The SSS documents consolidate and establish the depot quality control and reliability management procedure for inspection, testing, and/or restoration of items in storage. They encompass preservation, packaging, packing (PP&P) requirements, storage environment criteria, as well as inspection requirements during the storage cycle to determine material serviceability and the degree of degradation that has occurred. They are applicable to shelf life items as well as all items that are considered sensitive to shipment and storage deterioration. In the case of shelf life items, specifically those items whose shelf life is considered extendible, the standards are used to determine if the items have retained their original characteristics and are of a quality level which warrants extension of their assigned time period.

Figure 11.4-2 illustrates conceptually the basic technical approach in the preparation of the standards. The figure shows that the storage serviceability standards are formatted into two documents (per Ref. [5]). The first, which is based on Appendix A of Ref. [5], specifies PP&P levels, storage type and those tests, criteria and other provisions that can be coded easily into a computerized format. The second, which is based on Appendix B of Ref. [7], specifies applicable supplementary tests including functional performance, detailed visual and other special tests that cannot be coded easily into a computerized format but are necessary to assess the readiness of the stored items.

The form for the storage serviceability standards (see Figure 11.4-2 and Appendix A of DARCOM-R 702-23) contains in coded format the following data:

Federal Stock Number (FSN) - the federally assigned stock number for the item.

Item Name - provides a brief description of the item.

Quality Defect Code for Inspection (QDC) - defines potential storage-induced defects. The assigned defect codes cover preservation, packaging, marking, and storage as well as material deficiencies. Cyclic inspections are performed to accept or reject material relative to the defects identified by this code. A three-digit code is used, where the first digit identifies the severity of the defect (critical 0, major 1, or minor 2), and the second and third digits (see Table 11.4-2) identify a specific class of defects. For example, the code 113 would indicate a major defect (1) due to (13): container damaged or deteriorated. Complete definitions for quality defect codes applicable to the acceptance/rejection of material items inspected during the various depot inspection and testing phases (i.e., on receipt, audit, scheduled cyclic, special, etc.) are provided in AMCR 702-7 (Ref. [6]).

Inspection Level (IL) - determines the relationship between item lot or batch size and sample size for inspection. The inspection level is used in conjunction with the acceptable quality level (AQL) to form the sampling plan. (The sampling plan provides accept/reject criteria for individual item inspections).

Acceptable Quality Level (AQL) - the maximum percent defective (or the maximum number of defects per hundred units) that for purposes of sampling inspection can be considered satisfactory.

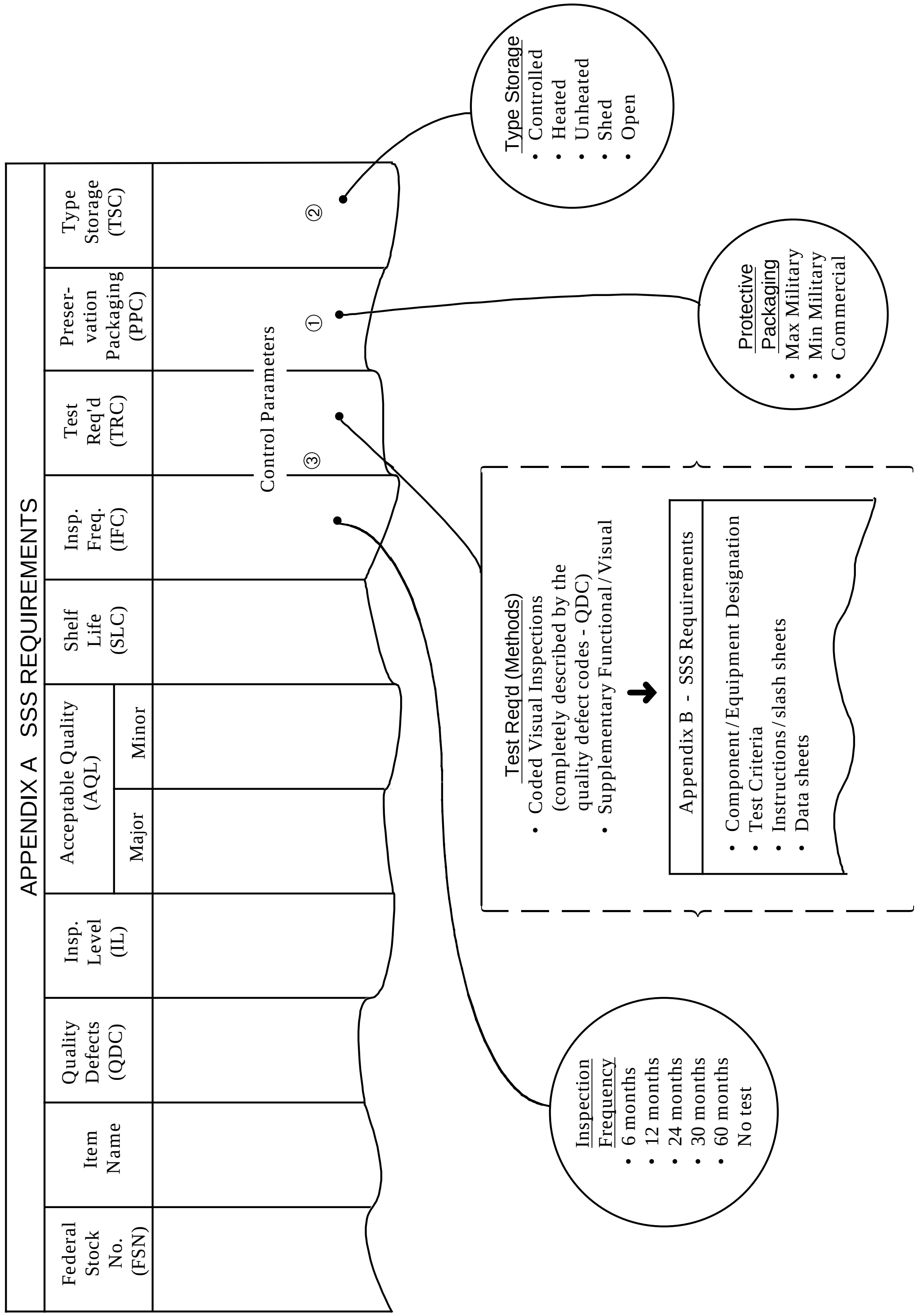


FIGURE 11.4-2: TECHNICAL APPROACH TO STORAGE SERVICEABILITY STANDARDS (SSS)

TABLE 11.4-2: STORAGE-INDUCED QUALITY DEFECTS

Category	Second & Third Digit (QDC)
Preservation Inadequate	02
Container Damaged or Deteriorated	13
Containers, Boxes, Crates, or Pallets Damaged or Deteriorated	23
Markings Illegible	33
Loose or Frozen Parts (out of adjustment)	40
Damaged Parts (cracked, chipped, torn)	41
Leakage (liquid)	45
Bonding Deterioration (soldering, welding, etc.)	48
Contamination (dirt, sludge, moisture, foreign matter)	50
Excessive Moisture (fungus, mildew, rot)	51
Shelf-life Data Exceeded	55
Failed Test Requirements (failed supplementary tests functional/visual)	62
Improper Storage Space	86
Corrosion, Stage 1 (or more)	90

Shelf Life (SLC) - describes deterioration characteristics versus time. Shelf life periods for deteriorative material range from 1 month to 60 months. The condition of a shelf-life item is evaluated during cyclic inspection in terms of time remaining and downgraded if necessary.

Inspection Frequency (IFC) - defines the elapsed time between cyclic inspections. Inspection periods range from 6 months to 60 months.

Test Required (TRC) - describes the method by which an item is to be inspected or tested.

Preservation Packaging (PPC) - describes the preferred level and/or most cost effective level of protection for each item. After an item has been inspected and accepted, the packaging/preservation is to be restored to its pre-inspection level. Further, the date of repackaging as well as the date of original packaging is stamped on the package.

Type Storage (TSC) - indicates the preferred or most cost effective storage condition.

In order to prepare standards for new or existing material items, criteria for specifying cost effective tests and control provisions are first established. The criteria (and the subsequent standards) should provide for the inspections to be performed frequently enough to detect potential problems but not so often as to dilute the total depot inspection effort and compromise other items in storage which may be more critical and require higher inspection frequencies. To be effective, the criteria must take into account:

(1) Material deterioration

(2) Application risk and criticality

(3) Cost

(4) Material complexity

(5) Preservation/packing and packaging (PP&P)

(6) Storage environment

The Army has developed general criteria and a material weighting factor technique as part of a complete standard preparation process that takes into account these factors (Ref. [7]). The process, which is illustrated in Figure 11.4-3, focuses on the three major control parameters: (1) protective packaging level, (2) storage type, and (3) cyclic inspection (frequency and method). The process involves first defining the level of packaging and storage (preferred) from a review of material deterioration properties and then determining inspection frequency by evaluating deterioration, application risk, criticality and other factors in light of the defined packaging and storage environment levels. It is an iterative process that involves tradeoff analysis to define an optimum set of requirements. It emphasizes and uses to the maximum extent the visual coded inspection criteria, i.e., QDC, to detect material failure and/or defects due to corrosion, erosion, and other deficiencies resulting from improper storage methods, extended periods of storage, and the inherent deterioration characteristics of the material item. The technique is sufficiently flexible to make allowances for available storage facilities if they differ from the preferred through the adjustment of inspection frequency.

In the initial preparation of the standards, the type and level of storage space and packaging methods are considered as fixed parameters (although iterative) where the preferred levels are defined based on material deterioration properties. Therefore, the element which provides the overall stimulus for the control and assurance of the readiness of stored components and equipment is the type and frequency of inspection. A ranking is assigned to each item that accounts for material deterioration and the other factors depicted in Figure 11.4-3 and is used as the basis to determine first the need for inspection and then, if needed, the frequency and type of inspection.

To effectively manage the depot cyclic inspection program, priorities are established as indicated in Figure 11.4-3. Items classified as definite shelf-life are given priority and subjected to cyclic inspection. Other indefinite shelf-life items that are considered particularly sensitive to deterioration are also subject to cyclic inspection. Definite shelf-life items are those possessing intrinsic deterioration characteristics that cannot be eliminated (or minimized) by storage and packaging controls. They are further classified into nonextendible (Type I) and extendible (Type II) materials. Indefinite shelf-life items, on the other hand, include items that do not deteriorate with storage time, as well as items that are sensitive to deterioration as a result of induced external failure mechanisms. The relationship between these types of material item classification and their relative deterioration level is illustrated in Figure 11.4-3. Figure 11.4-4 shows the nonextendible life characteristic of Type I material, the extendible shelf-life characteristic of Type II material, and the relative indefinite shelf-life characteristic of all other stored material.

Figure 11.4-5 presents a matrix that can be used to determine inspection frequency (IFC) and to optimize in-storage inspection coverage. The matrix includes:

(1) The most deteriorative items to the least deteriorative in terms of a total ranking factor that accounts for deterioration, complexity, cost, accessibility and criticality

(2) All combinations of depot storage and packaging conditions ranging from the most protective (containerized package and a controlled humidity environment) to the least protective (commercial package and an open area)

Application of the matrix to a given material item involves assigning appropriate values to each of the weight factors depicted in Figure 11.4-5 in order to arrive at a total ranking. This ranking represents a rough measure of the overall deterioration/cost sensitivity of the item to the storage environment. The ranking is then entered in the proper weight column of the matrix to determine inspection frequency for any desired combination of packaging and depot storage protection level.

SECTION 11: PRODUCTION AND USE (DEPLOYMENT) R&M

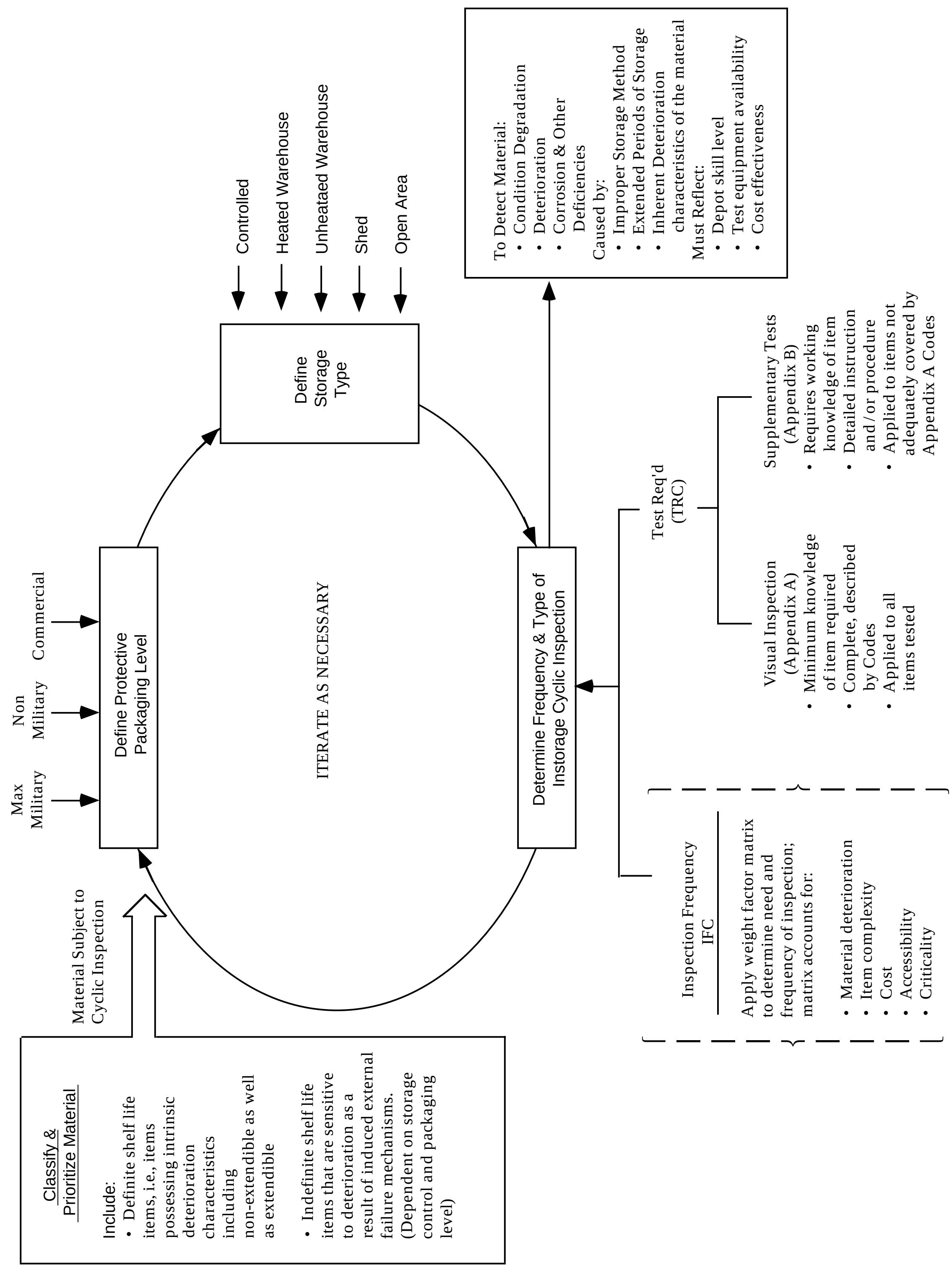


FIGURE 11.4-3: STORAGE SERVICEABILITY STANDARD PREPARATION PROCES

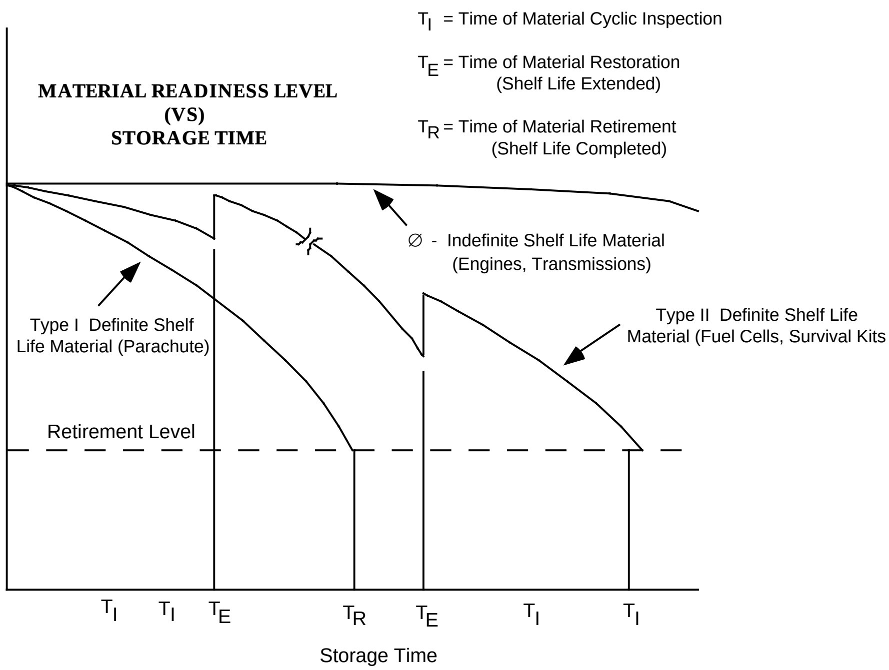


FIGURE 11.4-4: DETERIORATION CLASSIFICATION OF MATERIAL

STORAGE PROTECTION		
Storage Environment	Packaging Level	Test Frequency Code

For new items, the matrix allows broad tradeoffs to be made to arrive at the optimum balance of packaging, storage, and inspection requirements. Also, the combining of deterioration with cost and the other weight factors via the matrix approach allows the specification of cost effective inspection periods. This cost effectiveness is illustrated by considering two items one of which exhibits low deterioration properties but the cost and other factors are high, and the other which exhibits high deterioration properties but the total of the other factors is low. A relatively low cost or nominal test inspection frequency may be computed for both items that reflects an effective balance of all factors; whereas, if only deterioration was considered in computing the test periods, over-inspection (excessive cost) of the high deterioration item and under-inspection (low readiness assurance) of the low deterioration items would most likely result. Of course, for those items where all factors including deterioration and cost are high, frequent inspection would be required to ensure the readiness of material and for those items where deterioration and the other factors are low, less frequent inspections would be required.

The matrix approach also provides flexibility for regulating the number and type of items subjected to cyclic inspections by adjustment of the weight assigned to the factors that relate the material to the storage environment.

As previously indicated, an inspection time period is originally set based upon preferred storage environment and packaging methods specified in the TSC and PPC columns of Figure 11.4-2. However, many times an item will be stored and packaged at a different level. In that case an adjustment is made to its inspection time periods to maintain the same state of readiness based on the data provided in the inspection frequency matrix.

11.4.3.1 Application of Cyclic Inspection During Storage to Assure Reliability and Material Readiness

Critical to the control of reliability during storage is the proper application of cyclic inspections and tests. The purpose of in-storage cyclic inspection is to assess component/equipment reliability and readiness for use, to detect deterioration while in storage, and to furnish data for any necessary condition reclassification action. A knowledge of the component or equipment item, particularly its deterioration properties and risk attributes, is necessary to plan and specify optimum in-storage cyclic inspection requirements. The inspections must be practical and maintain an overall cost effective posture that reflects readily available depot test equipment and skills.

SECTION 11: PRODUCTION AND USE (DEPLOYMENT) R&M

In-storage cyclic inspection generally includes two basic types as indicated in the previous subsection. The first type is based on subjective visual inspections where material acceptance is completely described by codes covering quality defects (and included in the QDC column of the Storage Serviceability Standard). A minimum knowledge of the items is required to specify the criteria and perform the inspections. These coded requirements apply to all items tested. Figure 11.4-6 illustrates some of the quality defect codes and shows that the assigned codes cover preservation, packing, marking and storage, as well as material deficiencies. The figure indicates that there are basically three levels of inspection corresponding to (1) the outer package or container, (2) the inner packing, and (3) the item itself. If a defect is not considered critical, major, or minor at the time of inspection but (due to inspector experience) is expected to become critical, major or minor prior to the next cyclic inspection, it is identified as such, considered as a cause for rejection, and counted relative to the item's sampling plan criteria. Defects of a trivial nature are not considered as cause for rejection of a lot, unless some reduction in usability or function of items is expected prior to the next scheduled inspection. For example, nicks, dents, or scratches that do not break coatings or paint films are considered trivial deficiencies.

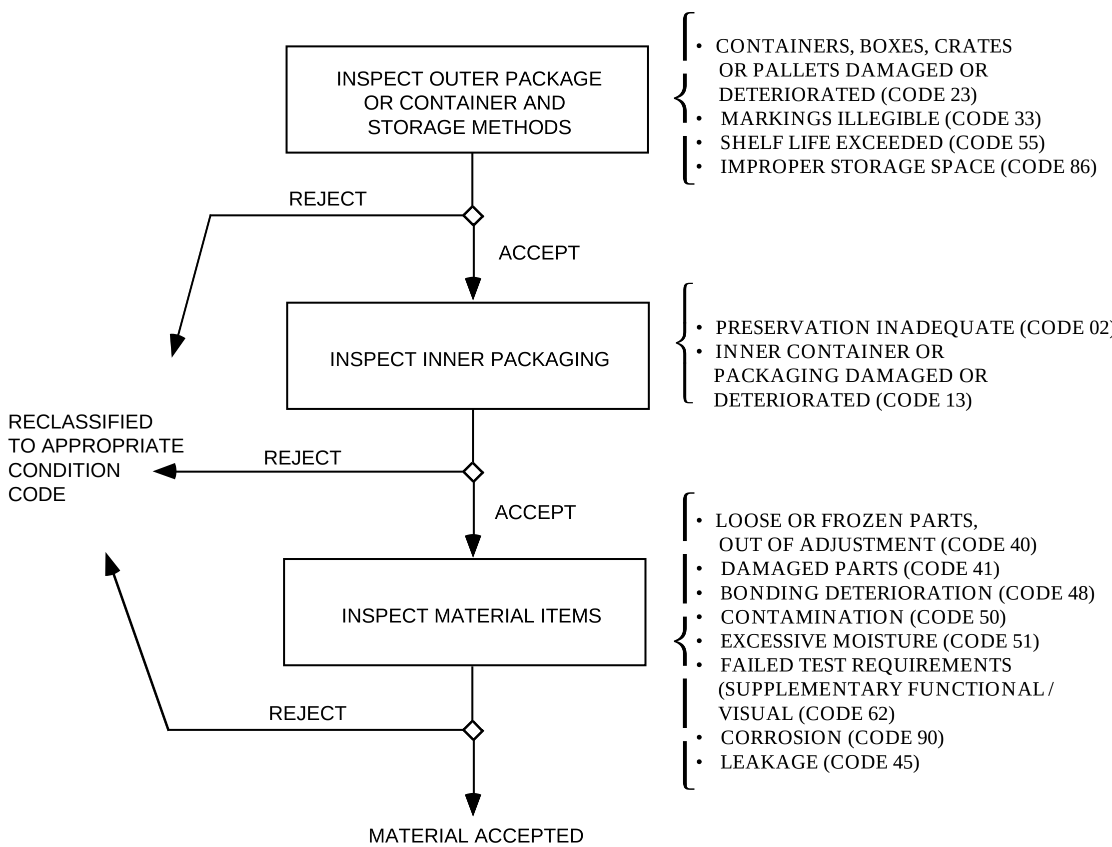


FIGURE 11.4-6: CODED QUALITY INSPECTION LEVELS

The second type of in-storage inspection involves supplementary requirements that are applied to items that cannot adequately be inspected by the visual coded requirements. They generally include functional tests (derived from technical manuals) and/or special, more-detailed visual inspections. Special test and/or inspection procedures complete with acceptance criteria are prepared for these items and included in Appendix B to the SSS. Emphasis is placed on defining viable test or checkout procedures that can be applied simply and quickly to the stored material items to assure that they perform satisfactorily with only a minimal level of evaluation, support, and guidance. These supplementary tests can be applicable to parts, material, equipment, or complete systems, including shelf-life items as well as other items that are storage sensitive.

The supplementary tests are not intended to represent a complete and detailed inspection or checkout of the item to determine compliance to specified requirements. The tests are designed to verify operability and are to be based on a “go/no-go” concept, fully utilizing end item functions to indicate functional readiness for service and issuance.

The functional tests are designed such that they do not require external and specialized test equipment except common and readily available equipment found at the depots and other installations (power supplies, volt-ohmmeters, etc.). The functional tests in general involve first checking the operational mode of all indicators such as dial lamps, power lights, meters, and fault lights as applicable and then applying a simple procedure that exercises some or all of its functions to verify operational status. Many times the equipment can be tested as part of a system. For example, two radio (receiver/transmitter) sets could be tested as a system pair by positioning the sets a certain distance apart (e.g., 25 feet). One is placed in the receive mode and the other in the transmit mode, and all associated hardware and interconnecting cables are attached. An audio (spoken word) input is applied to the set in the transmitting mode, and the set in the receive mode is checked for reception. The process is repeated with the transmitter/receive modes reversed.

The functional test procedures for a given equipment item can be derived from a review of the equipment's maintenance and/or operating manuals. These manuals describe the operational sequence, the turn-on and shut-down procedure, and the equipment operational test and checkout procedure necessary for complete verification of equipment operational status. Consequently, they provide a sound basis for deriving a simplified and cost effective functional test that is suitable for assessing reliability during storage.

11.4.4 Data Collection and Analysis (During Storage)

The shipment/storage test and control program, like the production test program, must be continually challenged relative to the effectiveness of the overall program as well as the individual tests. In-storage cyclic inspection must also be considered as a dynamic test where the test methods, frequencies, and criteria are adjusted to reflect actual depot and field experience. In-storage data (reject rate, quality discrepancy reports, causal data, etc.) generated during the implementation of the cyclic inspections should be compiled, reduced, thoroughly analyzed, and fed back to item management and engineering activities in a form that will provide a basis to:

(1) Determine the effectiveness of the shipment/storage degradation control program to meet reliability and readiness objectives

(2) Eliminate the causes for deficiencies

(3) Revise item inspection or protective packaging and storage level requirements, if necessary

11.5 Operational R&M Assessment and Improvement

Electronic systems are also subject to damage and performance degradation during operation and maintenance. Consequently, operational systems are continually assessed to ensure that they are performing in accordance with expectation and to identify areas where improvements can be incorporated to minimize degradation, improve R&M, and reduce life cycle costs. This time period is most significant because it is here that the true cost effectiveness of the system and its logistic support are demonstrated and historical R&M data are gathered and recorded for use on future products. The effort includes:

(1) Assessing R&M performance from an analysis of operation/failure data, identifying operation/maintenance degradation factors, and comparing actual R&M with that predicted and demonstrated during acquisition

(2) Identifying systems, equipment and other hardware items that exhibit poor reliability, require extensive maintenance and are prime candidates for cost effective improvements

(3) Evaluating the impact on R&M of system changes and corrective action implemented in response to operational failures

11.5.1 Factors Contributing to R&M Degradation During Field Operation

Degradation in reliability can occur as a result of wearout, with aging as the dominant failure mechanism. Defects can also be induced into a system during field operation and maintenance. Operators will often stress a system beyond its design limit either to meet a current operational need or constraint or inadvertently through neglect, unfamiliarity with the equipment, or carelessness. Situations occur in which a military system may be called upon to operate beyond its design capabilities because of an unusual mission requirement. These situations can cause degradation in inherent R&M parameters. Operational abuses due to rough handling, extended duty cycles, or neglected maintenance can contribute materially to R&M degradation during field operation. The degradation is usually the result of the interaction of man, machine and environment. The translation of the factors which influence operational R&M degradation into corrective procedures requires a complete analysis of functions performed by man and machine plus environmental and/or other stress conditions which degrade operator and/or system performance.

Degradation in inherent R&M can also occur as a result of poor maintenance practices. Studies have shown that excessive handling brought about by frequent preventive maintenance or poorly executed corrective maintenance (e.g., installation errors) have resulted in defects introduced into the system, with resultant degradation of R&M. Some examples of defects resulting from field maintenance, depot overhaul, or reconditioning are due to:

(1) Foreign objects left in an assembly

(2) Bolts not tightened sufficiently or overtightened

(3) Dirt injection

(4) Parts replaced improperly

(5) Improper lubricant installed

Also, during unscheduled maintenance, good parts are replaced in an effort to locate the faulty parts. In many cases, the good parts are written up as defective instead of being reinstalled. These parts often are returned to the depot for repair or discarded, resulting in a reported field failure rate that is higher than is actually occurring.

Several trends in system design have reduced the need to perform adjustments or make continual measurements to verify peak performance. Extensive replacement of analog with digital circuitry, inclusion of more built-in test equipment, and use of fault-tolerant circuitry are indicative of these trends. These factors, along with greater awareness of the cost of maintenance, have brought changes for ease of maintenance whose by-product has increased system R&M. In spite of these trends, the maintenance technician remains a primary cause of R&M degradation. The effects of poorly trained, poorly supported or poorly motivated maintenance technicians on R&M degradation require careful assessment and quantification.

The operation and maintenance induced defects are factors that must be carefully considered and taken into account in the assessment and control of operational R&M. In general, the environmental factors considered in prediction techniques account for the added stress provided by operation within that environment. However, the environmental stresses imposed during field maintenance may be other than what was anticipated during the original prediction. For instance, a subassembly removed for repair in a desert area may be placed in direct sunlight while awaiting transfer. Component temperatures may exceed those experienced during normal operation for an extended period, thus reducing their life expectancy. Mechanical stresses imposed on components during removal, repair, and reinsertion may exceed that designed for a given environment. Therefore, field and depot requirements and procedures must include criteria for controlling the reliability and quality of the repair/overhaul action to minimize potential maintenance induced defects in order to achieve an actual field R&M that approaches that predicted and demonstrated during acquisition.

11.5.2 Maintenance Degradation Control (During Depot Storage)

Depot maintenance activities include complete overhauling, partial rebuilding, product improvement and retrofit, calibration, and the performance of highly complex repair actions. In addition, the depot normally stores and maintains the supply inventory. Physically, depots are specialized fixed installations that contain complex and bulky production and test equipment, and large quantities of spares under environmental storage control. Depot facilities maintain high volume potential and use assembly line techniques with relatively unskilled specialists in key areas such as condition evaluation, fault diagnosis, and quality control and inspection.

Since the R&M of hardware items can be materially degraded during maintenance and depot operations, engineering plans and analyses are performed and R&M controls implemented to assure performance and to eliminate defects due to workmanship and the various other factors that would, if uncontrolled, lead to poor quality and R&M degradation.

Control efforts for a given hardware item start with the preparation of a Maintenance Plan during development as part of logistic support analysis (LSA); they continue into the operational and maintenance phase with the establishment of specific criteria and special maintenance and restoration procedures which must be followed to avoid R&M degradation and to retain the inherent R&M of the item. Possible deviations from the Maintenance Plan are described and related to their potential effect on operational R&M. Specifications are prepared and incorporated into a maintenance/depot requirement document including provisions covering:

(1) Life cycle reconditioning performance/quality parameters and acceptance criteria

(2) Types and kinds of material approved for use during overhaul, repair, and reconditioning

(3) Acceptable workmanship standards and techniques

(4) Quality and reliability assurance inspection, tests, analysis methods, and controls

The intent of the maintenance requirement document is to ensure that quality and R&M measures reflect adequate, viable, and practical acceptance criteria and procedures that can be implemented most cost effectively by depot personnel during the repair, overhaul, or reconditioning of the hardware items.

Some of the areas that are evaluated, controlled and reflected into the maintenance documentation from a reliability and quality standpoint are listed in Table 11.5-1. These include reviewing the technical accuracy and adequacy of instructions covering equipment checkout, calibration, alignment, and scheduled removal and replacement. In addition, all disassembly, cleaning, inspection, testing, repair, replacement, re-assembly, troubleshooting, preventive maintenance checks and services, and maintenance processes and procedures are evaluated.

Criteria are also established that recognize the fact that hardware in field use (as well as during storage) deteriorates due to age, environment, and storage conditions. When deterioration begins to take effect, the quality level of the material will decline below that which was initially specified during procurement. Although the effectiveness and adequacy of the reconditioning operations and controls will minimize the decline, the resultant quality level of the reconditioned material will usually be lower than that initially specified. The depot requirements include maintenance quality level requirements that reflect:

(1) Minimum deterioration, which is lower than the initially specified value

(2) Criteria that indicate the quality limits beyond which repair is not economically achievable

(3) Acceptance criteria for reconditioning cycles(s) at predetermined storage and use milestones

TABLE 11.5-1: DEPOT MAINTENANCE REQUIREMENT AREAS

Inspection and Test Equipment - Test equipment used to determine performance of depot maintenance specifications and requirements

Material Quality - Quality level of parts and material used for replacement, repair or modification

Pre-shop Analysis - Extent of overhaul required. Included in the analysis would be procedural instructions as well as a detailed checklist to aid in the evaluation of the items for determining extent of cleaning, repair, modification or replacement

In-Process Inspection - In-process inspection requirements, including procedural as well as accept/reject criteria associated with each overhaul operation such as disassembly, cleaning, repair, replacement and modification, as applicable

Diagnostic and Automated Test Equipment - Diagnostic and automated test equipment (such as NDT, magnetic particle, dye penetration, etc.) used to determine the adequacy of repair, overhaul or reconditioning

Repair - Total sequential, step-by-step instructions and specifications used for repair, replacement, reclamation, rework or adjustment for hardware items

Assembly/Disassembly - Total step-by-step instructions used to assemble/disassemble the hardware item

Calibration - Level and method of calibration for all equipment and instrumentation

Final Performance Check - Techniques and methods to assure total satisfactory performance of the hardware item in accordance with the established criteria

In addition, a process analysis similar to that described in Sections 11.2 and 11.3 to determine and control R&M degradation introduced by manufacturing can also be applied to determine and control degradation introduced by the reconditioning and overhaul process. This analysis would identify processing and inspection steps that can be improved to reduce R&M degradation and determine the need to incorporate controlled screening and burn-in tests as described in Section 11.2.

11.5.3 Maintenance Documentation Requirements

An important factor in controlling R&M degradation during deployment is the availability of adequate maintenance documentation for the equipment/system. System maintenance documentation includes the written, graphical, and pictorial data which should be supplied with the system for use by operators and maintenance personnel to accomplish both the routine preventive maintenance tasks and the corrective repair procedures identified in the Maintenance Plan for the system. This documentation should reflect the maintenance concept and repair policies established for the system. In general, system operation and maintenance documentation should be a completely integrated package providing clear-cut direction leading from failure detection to fault isolation and repair procedures and should be presented in a format and style designed for ready access and updating as changes are introduced.

Four types of data represent the minimum package which should be provided with an operating system if it is to be successfully operated and maintained in accordance with the Maintenance Plan. These working documents should be instructional and factual. The four categories of maintenance documentation required to successfully implement the Maintenance Plan are described as follows:

(1) Functional Description and Operating Instructions for Each System - Data in this category includes: a description of the capabilities and limitations of the installed system; a technical description of system operation, including its operating modes and alternate modes; step-by-step turn-on and manual operating procedures; "confidence" checks normally employed to verify that equipment is performing satisfactorily.

(2) Equipment and Installation Description - Data in this category must provide an accurate, up-to-date description of the hardware as it is installed in the weapons system. Minimally, it should consist of: A complete set of functional flow or logic diagrams; a complete set of schematic diagrams for electrical layout, electronics, hydraulics, pneumatics, etc.; parts data containing reference information in sufficient detail to permit reordering or fabrication of the individual parts within the system; and the necessary instructions for installing and checking out installed/retrofitted equipment.

(3) Maintenance Aids (Troubleshooting) - This category presents the specific data required by the technician for localizing a fault to a replaceable item and for checking out the system after repair. Included are:

    (a) Methods for system-level fault isolation when the system is "up" but operating in a degraded mode; use and interpretation of system readiness test results

(b) Method of system level fault isolation when the system is totally down; use and interpretation of fault isolation tests and monitoring of console displays

(c) Procedures for functional equipment level fault isolation; based on fault sensing indicators supplemented, as required, by test point measurements using built-in test equipment

(d) Equipment-level isolation techniques the use of which will permit identification of the problem area to a single module or replaceable part

(e) Routine tests, adjustments, alignment, and other "preventive" procedures which are performed at periodic intervals

(4) Ready Reference Documentation - This documentation is limited to that information routinely required by the technician in a given equipment repair area. The documentation should be easily usable in the work area - i.e., capable of being held with one hand, remaining open to a given section, permitting easy replacement or additions, and suitable for storage in the work area. It should contain only routine checkout, alignment, and preventive maintenance procedures; fault monitoring interpretation and replacement data; supplemental troubleshooting techniques required to complement the automatic fault detection and isolation system; and item and unit spare parts ordering data keyed to system identity codes.

11.5.4 Data Collection and Analysis (During Field Deployment)

A new system or equipment begins to accrue valuable experience data with its initial introduction into the field. These data, accurately recorded and consistently reported, provide the final basis for judging suitability of the system for continuing deployment. Thereafter, the reporting system can become the essential basis for an effective R&M feedback loop if provisions are made for continuous reporting and periodic analysis of maintenance experience data throughout the deployment phase and if formal procedures are established for progressive correction of discrepancies revealed by the analysis. On the other hand, if the reporting system is not fully exploited analytically and applied dynamically in a formal corrective action program, the R&M feedback loop is short circuited and serves no purpose other than logistic surveillance.

Data required to effectively assess, monitor, control and improve the R&M of fielded systems and equipment items include hours of operation (and appropriate data on operating characteristics), performance measures and assessments, application environmental factors, and, most important, failure and maintenance data. The feedback of information obtained from the analysis of failure during actual use is essential to reliability growth. The focus of the data collection should be on tracking failure modes, not symptoms.

Development of a formal and well-documented field data recovery, analysis and feedback system is a key element in an effective R&M program. The data recovery and feedback program is designed to be compatible with and incorporate data from other data collection efforts during acquisition and storage. An effective data system provides output information that can be used for:

(1) R&M assessments

(2) R&M tracking

(3) Comparative analysis and assessments

(4) Determination of the effectiveness of R&M tasks and management concepts

(5) Identification of critical equipment, components and problem areas

(6) Compilation of historical component failure rates for design predictions

Plans are prepared that describe the specific mechanisms for collecting operation, maintenance and installation data at field sites, depots, and disposal areas as well as during factory test for feedback. Included are detailed instructions, document forms, and the delineation of responsibilities for implementation. Furthermore, the system must be planned such that it is compatible with standard military data systems. It should be noted that during acquisition the data system is primarily the responsibility of the system equipment developer where control by the military is established through reporting of summary data and deliverable data items.

During operation, military maintenance data collection systems are used to record and accumulate ongoing data. These programs, including the Army's TAMMS (The Army Maintenance Management System), the Navy's 3M and the Air Force's REMIS and other maintenance data collection systems, are primarily maintenance oriented. Maintenance actions are reported and processed in a computer data bank at three levels: equipment, assembly board, and piece-part. For each entry, the failure indicator is reported along with codes identifying such things as the base command and the equipment nomenclature. They do not, however, report operating time. Moreover, the field use environment and the field maintenance environment are not adequately quantified to ensure consistent interpretation of field data. Thus, field reliability cannot be assessed using data from only the military systems. In order to assess reliability and to compare the attained field reliability with that specified and estimated during acquisition, both equipment/system failure (or maintenance) data and their associated operating time(s) are required. The associated equipment/system operating time must be estimated or obtained directly from the operational units themselves. Operating times are recorded in station logs and the equipment inventory, with associated records of uptime, storage time and maintenance times, by month.

In addition to the previously mentioned maintenance data collection systems, the Department of Defense instituted the Reliability Analysis Center (RAC), a DoD Information Analysis Center, which functions as a focal point for the recovery of reliability test data and experience information on electronic, electrical, and electromechanical components, and R&M data on the equipments/systems in which these components are used. Reliability experience information is disseminated by the RAC through reliability data compilations, handbooks and appropriate special publications to upgrade and support system reliability and maintainability.

These publications cover the following:

(1) Nonelectronic Parts Reliability Data (NPRD)

(2) Nonoperating Reliability Databook (NONOP-1)

(3) Failure Mode/Mechanism Distributions (FMD)

The publications are updated and reissued periodically, deleting outdated data entries and incorporating new acquisitions from the latest technologies and applications. For additional information on the RAC, as well as other specialized DoD Information Analysis Centers, see Reference 9.

11.5.5 System R&M Assessment

Once an equipment/system is deployed, its R&M performance is periodically assessed based on the analysis of collected field operational/failure data as described in the previous section, as well as information derived from other sources. Programs have been established to assess R&M in a manner so as to yield consistent and accurate data and information that can be fed back into the product improvement process as well as to provide a "lessons learned" information base for subsequent acquisitions. The programs are designed to provide data and information that can be used to:

(1) Uncover problem areas, effect timely corrective action, and provide a solid basis for system R&M improvement programs.

(2) Determine the effectiveness of design, test and program concepts applied during system acquisition.

(3) Track the performance and, in particular, the R&M of the fielded system.

Application of the feedback loop to service evaluation of R&M and correction of R&M problems is accomplished in five major steps, the last of which becomes the first step in a repetition of the cycle:

(1) Acquisition of Required Data - Use the data collection and reporting system to acquire the basic service use experience data, supplemented as necessary by system configuration and engineering data, and operational information to ensure correlation between reported maintainability experience and the conditions under which the experience data was accrued.

(2) R&M Assessment - Analyze the reported experience data to derive a measure of the R&M parameters (e.g., failure rate, MTBF, mean corrective maintenance time, ($\overline{M}_{ct}$), maximum corrective maintenance time ($M_{max_{ct}}$), maintenance manhours per operating hour, logistics delay time, etc.) at system, subsystem, equipment, major component, and lower levels, corresponding to the levels to which R&M was allocated, specified, and demonstrated during the development phase.

(3) Problem Definition - Identify, investigate, and describe the underlying problems which account for major discrepancies or deficiencies noted in the analysis of (2) above in terms amenable to translation into corrective action as design changes, documentation changes, maintenance or logistics procedural changes, etc., as appropriate. Introduce on a limited sampling basis such supplementary data recording forms, time clocks, instrumentation, and reporting instructions as required for the assessment of R&M where the field values greatly exceed predicted or demonstrated values.

(4) Corrective Action Assignment - Formally assign corrective action responsibility accompanied by problem descriptions developed under (3) above with specified criteria for verifying achievement of corrective action objectives.

(5) Follow-Through - Reassess R&M as in (2) above to evaluate effectiveness of corrective actions, to compare R&M trends relative to established improvement objectives, and to reevaluate problems identified in earlier assessments. This step begins the assessment cycle all over again.

Department of the Army, Readiness Command (DARCOM) Regulation 702-9 (Ref. [10]) defines the policies and procedures of a formal R&M System Assessment Program established by the Army. This regulation requires that assessments be performed in order to determine whether the fielded system has satisfied user needs for mission performance and logistic support. They are conducted in order to identify and take corrective action on problems which are degrading user satisfaction, operational readiness, and life cycle cost. Through the performance of such assessments the Army determines how a system is operating, uncovers and corrects problems in system operation and support, and thus helps achieve complete user satisfaction.

As presently structured, the System Assessment Program includes the assessment of all aspects of fielded system operations including:

(1) Technical
- A narrative description of the system and its support equipment
- Original design objectives
- The results of development and operational tests
- Corrective action results

(2) Operational
- Initial field performance parameter values
- Changes incorporated into the fielded system (e.g., payload, accuracy, reliability, availability, and maintainability)
- Present field performance parameter values

(3) Environmental
- Individual component shelf-life values
- The reliability of components which require storage stockpile testing
- The effect stored components are having on overall system reliability

(4) Human Factors
- The user's opinion of the adequacy of the system
- The quantity of personnel, by military occupational specialty
- The quality of personnel, by military occupational specialty

(5) Support
- Current problems
- Development initiatives for replacement
- Effectiveness of the present logistic support system
- Improvement actions required
- System improvement plans

DARCOM Regulation 702-9 states that maximum use will be made of existing field data to assess these areas. Other data sources include

(1) Sample data collection programs

(2) Field visits and surveys

(3) User questionnaires

(4) User conferences

(5) Logistic personnel and field maintenance technicians

11.5.6 System R&M Improvement

In addition to optimizing R&M during acquisition through aggressive design, development, and production programs, substantial R&M growth potential exists during deployment. Some of this growth occurs naturally as the operations and maintenance personnel become more familiar with the equipment. However, to accelerate the growth rate and achieve significant increases in operational R&M requires the application of a closed-loop process of positive corrective action based on analysis and assessment of field R&M data. For newly deployed equipment, this closed-loop process can achieve significant reliability improvement, especially when used within the context of a total, disciplined system assessment program as discussed in the previous subsection. Reliability growth is based upon the iterative process of monitoring system operation to identify actual or potential sources of failures, to redesign out the failure source, and to fabricate and apply changes which improve system reliability. As such, reliability growth can be applied during development, production, or during operation. For fielded systems, the reliability growth process is a valuable tool to attain reliability improvements and achieve savings that could more than offset the cost of the reliability improvement program. The process is also performed during field deployment to eliminate problem areas not evident during the development phase.

The R&M improvement program must work in conjunction with the data collection and assessment programs (as discussed in the previous section) in a total integrated process consisting of data collection, system assessment and improvement selection, development, and implementation to achieve reliability growth in the field.

As described in more detail in the previous section, the program is an iterative feedback process consisting of the following steps:

(1) Acquisition of required data

(2) R&M assessment

(3) Problem definition

(4) Corrective action assignment

(5) Follow through to evaluate effectiveness of corrective action(s)

SECTION 11: PRODUCTION AND USE (DEPLOYMENT) R&M

The action of improving system reliability involves a systematic review of several concepts which appear from the backup data to be most useful for reliability cost tradeoff considerations, among which are:

(1) The reduction of failure rates by operating components at reduced (derated) stress levels, accomplished by selecting components which have ratings well in excess of those required for their system application.

(2) The use of improved components for which reliability has been significantly increased through special manufacturing techniques, quality control procedures, and testing methods.

(3) Design simplification to eliminate parts or components.

(4) The substitution of functionally equivalent items with higher reliability.

(5) The overall reduction of failure rate through increased control of the internal system environment, e.g., through reduction of ambient temperature, isolation from handling effects, and protection from dust.

(6) The provision of design features which enable prediction of incipient failures and permit remedial action to be taken before an operational failure occurs.

(7) The provision of design features which reduce the probability of human-initiated errors.

(8) The provision of multiple identical parts, paths or higher functional levels (redundancy) in order to prevent a system failure in the event that one element fails.

(9) The reduction of failure rate through increased control of the environment external to the equipment, as through reduction of ambient temperature, isolation from handling effects, isolation of operator from ambient noise, and protection of equipment from dust.

(10) The implementation of controlled screening and burn-in tests for the purpose of significantly reducing incipient failures due to undetected defects in workmanship or components.

Similarly, maintainability (MTTR) can be improved by incorporating improved use of maintenance practices, providing higher quality technical manuals and maintenance aids or possibly better training to improve the skill level of the technicians.

Computing the impact of the improvement recommendations which appear most useful for cost tradeoff consideration on MTBF, MTTR, overall downtime and system performance, using the methods and techniques previously described, and determining the total cost for their implementation is an essential step in evaluating the effectiveness of the improvement.

Critical to the analysis process is the ability to assess quantitatively the cost effects of reliability and maintainability. The cost of each recommended change must take into account total cost throughout the life cycle of the system and accordingly must include cost elements associated with design, manufacture, procurement, installation, and field use (i.e., operation, maintenance, and logistics).

The final step is to compute cost/benefit factors, i.e., develop a numeric for each R&M recommendation which reflects the total cost of the change, its impact on system performance, and the cost avoidance to be realized over a given time period by their implementation. This will allow the determination of those change recommendations which have maximum cost effectiveness. (See Section 8.1 for a discussion on reliability data collection and analysis). The recommended changes can then be presented in an improvement plan in prioritized order of cost effectiveness, as defined by the computed cost/benefit factors.

11.6 References For Section 11

1. Schafer, R.E., A.E. Sari and S.J. Van DenBerg, Stress Screening of Electronic Hardware. RADC-TR-82-87, May 1982, (AD-A118261).

2. Environmental Stress Screening Guidelines. The Institute of Environmental Sciences, Library of Congress Catalog Card No. 62- 38584, 1981.

3. Navy Manufacturing Screening Program. NAVMAT P-9492, Naval Material Command, May 1979.

4. Care of Supplies in Storage (COSIS). Army Regulation AR 740-3, 1993.

5. Product Assurance - Storage Serviceability Standards (SSS). Army Regulation DARCOM-R 702-23.

6. Product Assurance Depot Quality Assurance System. Army Regulation AMC-R 702-7.

7. Army Supply Bulletin SB740-99-1, Storage Serviceability Standard for TSARCOM Material.

8. Maintenance of Supplies and Equipment, AMC Guide to Logistics Support Analysis. AMCP 750-16, Headquarters, Army Materiel Command, January 1978.

SECTION 11: PRODUCTION AND USE (DEPLOYMENT) R&M

9. Directory of the Department of Defense Information Analysis Centers. Defense Technical Information Center, Defense Logistics Agency, Cameron Station, Alexandria, VA 22304-6145, April 1995.

10. DARCOM Regulation 702-9, Department of the Army, September 1977.

11. Environmental Stress Screening Guidelines. Tri-Service Technical Brief, 002-93-08.

12. Environmental Stress Screening Guidelines for Assemblies. 1984 & 1988, The Institute of Environmental Sciences.

13. Environmental Stress Screening Guidelines for Assemblies. 1990, The Institute of Environmental Sciences.

14. Environmental Stress Screening Guidelines for Parts. 1985, The Institute of Environmental Sciences.

15. Impact of Nonoperating Periods on Equipment Reliability, RADC-TR-85-91.

16. Kinlaw, Dennis C., Continuous Improvement and Measurement for Total Quality: A Team-Based Approach, Pfeiffer & Company/Business One Irwin, 1992.

17. Wilson, Lawrence, H., Eight-Step Process to Successful ISO 9000 Implementation: A Quality Management System Approach, ASQC Quality Press, Milwaukee, WI, 1996.

18. Prescott, Jon, "What the $75 Books Don't Tell you about ISO 9000 Documentation," Reliability Review, Volume 15, Number 2, June 1995.

19. Breitenberg, Maureen, "Questions and Answers on Quality, the ISO 9000 Standard Series, Quality System Registration and Related Issues," NISTIR 4721, National Institute of Standards and Technology, April 1993.

12.0 RELIABILITY MANAGEMENT CONSIDERATIONS

12.1 Impacts of Acquisition Reform

As discussed in Section 4.0, recent Acquisition Reform policies have resulted in the elimination of many historical reliability standards from the DoD procurement process. Past versions of this handbook heavily relied on MIL-STD-785 (canceled 30 July 1998), *Reliability Program for Systems and Equipment Development and Production,* to explain and provide guidance on the makeup, planning and management of a reliability program. However, under new reforms in acquisition, such standards can no longer be levied as a requirement on the system development contractor. In the past, the procuring agency was able to develop a statement of work (SOW) that specifically stated the contractor was to develop a Reliability Program Plan, and further, which reliability tasks from MIL-STD-785 (canceled 30 July 1998) were targeted to be performed to meet stated quantitative reliability requirements for the system to be procured. Now, as part of the cited reform policies, MIL-STD-785 has been canceled as of 30 July 1998, and military standard documents, with some exceptions, may not be imposed without a waiver. On the other hand, there is nothing in the latest acquisition reform language that prevents the system developer from proposing to use any current or previously existing military standard or handbook as the basis for implementing a design approach or program as part of an overall development approach.

12.1.1 Acquisition Reform History

On June 29, 1994, Secretary of Defense William Perry issued a five-page memorandum, “Specifications & Standards - A New Way of Doing Business.” The intent of the memorandum can be summarized as three "overarching” objectives:

(1) Establish a performance-oriented solicitation process

(2) Implement a document improvement process

(3) Create irreversible cultural change in the way DoD does business

The DoD is working to streamline the way in which procurement is managed and to adopt commercial practices whenever possible. It is reassessing and trying to improve the way it does business to decrease costs and increase customer satisfaction.

As will be explained, military standards and specifications may be cited for guidance in a Department of Defense solicitation but **shall not** be cited as requirements unless a waiver is granted. Commercial standards may be cited for guidance. Although not specifically prohibited by policy at the time this handbook was written, commercial standards should not be mandated as requirements. Given the spirit of the new acquisition policy, mandating a commercial standard is no different than mandating a military standard. In either case, the procuring agency would be telling the bidding contractor what to do and how to do it, at least to the extent that the cited standard provides suggestions on the tasks and activities needed for reliability. **The main objective of the new policy is to use performance specifications.** Only when performance specifications are inadequate for fully describing what the government wants should commercial requirements for military-unique items used in multiple programs or applications. MIL-STD-961 includes guidance on the format and content of SPSs.

(d) Program-Unique Specifications. This type of specification, also called a system specification, establishes requirements for items used for a particular program or weapon system. Little potential exists for using these specifications in other programs or applications. They should be performance-based but may include a blend of performance and detail design requirements. They are restricted to items for which the preceding categories of performance specifications are not applicable.

(3) Performance specifications are also categorized by the type of item being acquired. Those used to acquire materials are called material specifications, to acquire components are called component specifications, and to acquire systems are called system specifications. The Department of Defense has issued a guide

solicitation unless a waiver is granted.

(d) Standard Practice Standard. A standard practice standard is one that specifies procedures on how to conduct certain functions or operations. These procedures are not related to manufacturing processes. It has not yet been decided if standard practice standards may be cited as requirements in a Department of Defense solicitation without a waiver.

(2) Handbooks. A handbook is a guidance document that provides engineering or technical information, lessons learned, possible options to resolve technical issues, classification of similar items, interpretive direction and techniques, and other types of guidance or information. The purpose is to help the customer or the seller to design, construct, select, manage, support, or operate systems, products, processes, or services. Military handbooks **shall not** be cited as a requirement in a Department of Defense solicitation, contract, specification, standard, drawing, or any other document.

12.1.1.3 Overall Acquisition Policy and Procedures

The primary documents governing defense acquisition are DoD Directive 5000.1 and DoD Regulation 5000.2-R. Both documents were revised as a result of Defense Acquisition Reform. A third document, DoD 5000.2-M has been canceled. The revisions of 5000.1 and 5000.2-R incorporate new laws and policies, separate mandatory policies and procedures from discretionary practices, and integrate acquisition policies and procedures for weapon systems and automated information systems. In addition to the two documents, an Acquisition Deskbook is available to DoD procuring activities. The Deskbook is an automated repository of information consisting of a Desk Reference Set, a Tool Catalog, and a forum for information exchange. The Reference Set consists of mandatory Guiding Principles, discretionary Institutionalized Knowledge, and Sage Information (expert wisdom and lessons learned). Information about the Acquisition Deskbook can be obtained using the Internet:
<http://deskbook.osd.mil/deskbook.html>.

The major themes of the new acquisition documents are teamwork, tailoring, empowerment, cost, commercial products, and best practices. These themes can be summarized as follows: acquisition should be a team effort among all concerned in the process, the acquisition approach for a specific system should be tailored based on risk and complexity, acquisition will be conducted with a customer focus, cost will be an independent variable in programmatic decisions, commercial products should be used when practical, and acquisition is now more closely modeled on best commercial business practices.

12.1.1.4 Impacts on Reliability Management

Despite the recent changes in Acquisition Reform policy, reliability management methods and concerns have not changed dramatically. The major change is in how the reliability program tasking is defined, and the greater emphasis on the use of Commercial-Off-The-Shelf (COTS) and Nondevelopmental Item (NDI) equipment. It is now the contractor or supplier who has to decide what needs to be done to cost effectively achieve a stated reliability capability. Further, the government or other customer must evaluate which, of potentially many different approaches provides the best value. As will be discussed in this section, it is still important to the contractor to develop a reliability program plan and manage reliability as an integral part of the product design and development effort. For the customer, greater emphasis must be put on defining the required levels of reliability, availability and maintainability that are needed to meet performance expectations. This will include defining product usage profiles, the maintenance concept, operating and maintenance environments, and other life cycle factors such as storage and handling conditions. This information is essential if the contractor is to define a reliability program that meets stated requirements within cost and schedule constraints.

12.2 Reliability Program Management Issues

In managing a reliability effort, whether as a customer or as a supplier, there are several key issues that must be addressed. For any product or system, the key issues from any customer's perspective are:

(1) What measures of reliability are important?
(2) What levels of reliability are necessary to meet my needs?
(3) How will it be ensured that the required levels of reliability have been achieved?
(4) What reliability activities are the most effective for the product or system, such that the reliability program objective is achieved? Note: Even when the contractor selects the reliability activities, program offices must be able to judge which activities are applicable to their particular acquisition. Such judgement allows the acquisition staff to determine the risks associated with a contractor's proposed effort and, if necessary, negotiate changes.

From a supplier's perspective, the key issues are:

(5) What reliability activities are the most effective for the product or system, such that the reliability program objective is achieved?
(6) What reliability design goals are appropriate to ensure that customer's needs are met?
(7) What design approaches will be most effective in achieving the required reliability in the expected environmental and usage profiles?
(8) What tasks can be effectively used to assess progress towards reliability goals and requirements?
(9) What are the most appropriate means to determine if the reliability objective has been achieved?
(10) How can the designed-in reliability be retained during manufacturing and operational use, thereby ensuring reliable performance?

Each of the above issues must be addressed as part of meeting the basic objectives of product reliability which are: understanding the customer's requirements, meeting the requirements and demonstrating the requirements have been met.

In a commercial world, the customer is not usually concerned with the second set of issues - they are left to the seller to confront. If the seller does a poor job, the customer will go elsewhere for the product. Thus, competition in the marketplace provides a strong incentive to "do it right." In the military world, the level of competition is often much lower than in the commercial world. If dictated by the nature of the product (e.g., used only by the military), the risks (e.g., very high with unproved technologies being used), and the type of acquisition (e.g., totally new development), it will be necessary for the government customer to take more of an active role in addressing the second set of issues. (Some industrial customers also may be involved with the second set of issues, especially those dealing with measuring progress and determining the achieved level of design reliability). The form that this role takes, however, has changed.

Previously, by imposing standards and specifications, the military customer could force contractors to use certain analytical tools and methods, perform certain tests in a prescribed manner, use parts from an approved list, and so forth. The objective under Defense Acquisition Reform is not to tell contractors how best to design and manufacture a product. The responsibility for making such decisions has shifted from the government to the contractor. None-the-less, military customers are still more likely to be aware of the second set of issues than are commercial customers. Consequently, specifications issued by the government will probably continue to be more detailed than those issued by commercial organizations. Of course, when COTS products or non-developmental items (NDI) (Ref. [2]) are being procured, a more commercial approach to acquisition by the government is appropriate.

12.3 Reliability Specification Requirements

It is through the solicitation that a customer describes a needed product and solicits bids from competing sources to develop the product. Typically, a solicitation consists of a specification and a statement of objectives (SOO) or statement of work (SOW). (Note: Military solicitations must be issued in accordance with the Federal Acquisition Regulations).

(1) The specification should be a performance specification, one that states requirements in terms of the required results with criteria for verifying compliance but does not state the methods for achieving the required results.

Traditionally, a military or commercial acquisition has only one specification. Some customers, however, have adopted a new approach to specifications. They issue an initial specification and then work with each prospective bidder to develop a specification unique to that bidder. In that way, multiple specifications are developed. The specifications reflect the technical capability of each bidder, and one bidder's specification may be more demanding than others, although all must meet the customer's needs. The bidder whose specification and price represents a best-value is awarded the contract.

In some cases, the customer does not provide a specification. For example, the general public does not provide automobile manufacturers with specifications for a vehicle. Instead, the automobile manufacturers must develop their own specifications based on such considerations as: federal, state, and other government laws and regulations, benchmarking of competitors' products or market surveys and opinion polls.

(2) The SOW normally includes constraints, assumptions, and other criteria that the bidders must consider in developing and manufacturing the product. For example, the customer should identify how the product will be used (operating concept) and supported (support concept).

The SOW may also include specific activities or tasks required by the customer. In the past, the SOW included with a military solicitation almost always identified specific tasks, such as "perform a Failure Modes and Effects Analysis." As stated earlier, the approach under Defense Acquisition Reform is to allow the bidders to identify planned activities and to explain why, how, and when these activities will be performed. Commercial customers seldom specify specific tasks but are, of course, free to do so.

Instead of the traditional SOW, some procuring agencies use a statement of objective (SOO). Considered more in keeping with the spirit of acquisition reform, the SOO is concise and written to allow the contractor as much flexibility as possible in responding to the solicitation. A typical SOO has five sections: Objective of the Program (Solicitation), Objective (Purpose) of the Contract, Scope of the Contract, Work to be Accomplished under the Contract, and Program Control. The SOO is included as an attachment to an RFP, typically appended to Section L. Normally, the government will ask offerors in response to the SOO to prepare and provide a SOW in their proposals. Specific efforts defined in an offerors SOW shall be traceable to the SOO.

12.3.1 Template for Preparing Reliability Section of Solicitation

In developing the reliability portion of a procurement package, two distinct areas must be covered. These areas are performance-based requirements and programmatic and reporting requirements.

Performance-based requirements for reliability that may be placed in a specification include but are not limited to: Probability of Success P(S), Mission Reliability $R(t_m)$ or MTBF. In the case of programmatic and reporting requirements, the customer may require the seller to prepare and submit reports describing the results of analyses, tests, and other activities conducted by the contractor and described in the reliability program plan to design and manufacture a reliable product.

For NDI and COTS items the customer may require the seller to furnish operational data and the results of testing to substantiate reliability claims. In addition, the customer may require the seller to propose a method for verifying that reliability requirements have been met.

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

It should be the supplier's responsibility to select the tasks and other activities that will achieve these objectives and to describe the tasks and activities in the reliability program plan. When the customer mandates specific activities, previously referred to as tasks, the contractor is, to some extent, relieved of the responsibility to ensure each activity is value-added and preferable to other activities.

The following Template provides an outline for developing the reliability portion of a procurement package. The following conventions are used.

> Words within { } pertain only to new development efforts; words within [ ] pertain only to procurement of NDI or COTS. Procurement packages for programs involving both NDI/COTS and new development items should address each type of item separately but require that the reliability efforts be integrated.
>
> Blanks __ indicate where the user of the template must provide a value or other information.
>
> *Italicized words* are optional instructions that may or may not be used depending on the desires of the user and the needs of the procurement.
>
> Notes to the reader are in parentheses with NOTE printed all in caps.

The reader is reminded that when purchasing NDI or COTS, the best course of action may be to require only data that substantiates any claims for performance and to emphasize the role of manufacturing processes (for NDI not yet in production) in determining the reliability of the product. In some cases, even that data may not be needed if either the customer has already determined (through its own testing of samples, for example) that the product has the requisite performance or if use or independent testing of the product in actual applications has shown the product's performance to be satisfactory (for example, a personal computer in an office environment).

As previously discussed, in lieu of issuing a SOW with a specification, some customers now issue a SOO and require the offerors to include a SOW as part of their proposals. The SOO could include reliability objectives for the acquisition program, such as those listed in Section 3 of this Handbook. The best manner to respond to the solicitation would be left entirely to the bidders (for example, whether or not to have a reliability program plan).

A draft solicitation can be released by a customer for comment and suggestions for a statement of work by potential bidders. Based on the comments and suggestions received, a "negotiated" statement of work reflecting the bidders' best ideas on achieving the required level of reliability would be included in the formal solicitation (assuming a SOO is not being used instead).

TEMPLATE FOR DEVELOPING RELIABILITY PORTION OF A PROCUREMENT PACKAGE

SECTION L

(NOTE: Not all possible requirements are listed, and not all listed requirements are necessarily applicable to all procurements).

1. The bidder shall describe how the reliability requirements of the solicitation will be met. If a bidder elects to submit a reliability program plan, the plan may become part of the contract upon contract award. In any event, the bidders' responses will be evaluated using the following criteria.

1.1 The bidder shall describe all activities considered to {be necessary for ensuring the development of a} [have contributed to designing and manufacturing a] reliable product. For each activity, the bidder shall describe the objective, rationale for selection, method of implementation, methods of assessing results, and any associated documentation.

1.2 The bidder shall explicitly address how the included activities {will be} [were] integrated into the product and manufacturing design processes.

1.3 The bidder shall show how the results of the included activities {will be} [were] used to support other activities, such as logistics planning, safety analyses, etc.

1.4 The bidder shall explicitly show a clear understanding of:

a. the importance of designing in reliability and the relationship of reliability to other system performance characteristics.

b. reliability design techniques, methodologies, and concepts

c. the importance of integrating reliability activities into the overall systems engineering process

1.5 The bidder shall show how the following objectives {will be} [were] met:

a. thoroughly understand the design

b. validate the design and manufacturing processes

c. ensure proper parts application

d. address all portions of the product including those provided by suppliers and vendors

e. evaluate the achieved reliability

{f. determine feasibility}

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

TEMPLATE FOR DEVELOPING RELIABILITY PORTION OF A PROCUREMENT PACKAGE

STATEMENT OF WORK

1. The bidder shall identify all work activities {to be} conducted to meet the reliability performance requirements cited in _________ . In so doing, the bidder shall:

- *identify the specific objective of each work activity*
- *identify each work activity {is to be} [was] conducted*
- *identify specific product or outcome {expected} [achieved]*
- *explain how these work activities fit into the overall design effort*
- *identify any standards (commercial, military or company) that {will be} [were] used in performing the work activities*

1.1 The bidder will identify special reliability risks or issues associated with the chosen design approach and describe which work activities[ed] address these risks or issues and how.

1.2 {The bidder will identify work activities that are new to the company or are being used for the first time and explain what steps will be taken to minimize any risk associated with first use).

NOTE: Regarding the next section, the reader is reminded that mandating tasks, even for new development, is somewhat risky because it relieves the bidders of the responsibility for selecting the best means to accomplish the desired ends (in this case, meet the reliability performance requirements). Mandating tasks should be done only after careful consideration of the advantages and disadvantages of doing so. **Even then, bidders should not be told how to accomplish the required task. And, unless a waiver is obtained, processes may not be contractually mandated (reference OUSD (A&T) memorandum dated 18 September 1997, "Requiring Processes on Contract.")**

2. The following activities will be conducted by the bidder and reflected in the technical approach.

2.1 Implement a Failure Reporting and Corrective Action System (FRACAS).

2.2 Conduct either a Fault Tree Analysis (FTA) or Failure Modes and Effects Analysis (FMEA). Rationale for selecting one over the other will be given.

2.3 Institute an Environmental Stress Screening program. The bidder should indicate how the stresses and stress levels will be determined.

2.4 Develop a reliability model and make initial reliability predictions using that model. All predictions should be made at a stated level of confidence.

2.5 Implement a parts control program. Parts will be derated; the bidder will indicate how derating criteria will be developed.

TEMPLATE FOR DEVELOPING RELIABILITY PORTION OF A PROCUREMENT PACKAGE

STATEMENT OF WORK

2.6 Conduct thermal analyses to ensure parts and components are not subjected to thermal stresses that exceed design tolerances.

2.7 Conduct formal reliability growth testing for the purpose of uncovering design deficiencies and other failure mechanisms.

2.8 Conduct a reliability demonstration. The bidder shall explain how the demonstration will be implemented and the underlying statistical basis of the demonstration.

2.9 Conduct a ____________ (NOTE: Others as determined by buyer).}

(NOTE: All reports, data requirements, and deliverable documents should be identified in the Contract Deliverables Requirements List (CDRL). Data items can include FMEA results, results of trade studies, thermal analyses results, and so forth. Data items should be selected based on the nature of the development, the level of risk, intended use of the item [benefit], and cost. The CDRL should provide data format and content preparation instructions and data delivery requirements. Although the text of the SOW should not include these items, a data item description number listed in the CDRL may be cross-referenced in the SOW. This cross reference should usually be made in the paragraph describing the task that will lead to the development of the data or document).

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

TEMPLATE FOR DEVELOPING RELIABILITY PORTION OF A PROCUREMENT PACKAGE

THE SPECIFICATION

(NOTE: User should select the life units most appropriate for each product. For example, operating hours might be the best measure for an engine, miles traveled for a truck, cycles for a starter, and so forth).

1. The bidder shall carry out the activities described in the Statement of Work to achieve the following levels of reliability. Note: All values are the minimum acceptable values at a______ ___ confidence level.

   1.1 The product shall exceed ______ life units between any failure that requires a maintenance action

   1.2 The product shall exceed _______ life units between any failure that prevents the product from performing its mission

2. The service life of the product will be ________ life units. Service life is defined as the period over which the product can be operated and maintained in accordance with the contractor's prescribed maintenance and operating procedures before continued use becomes prohibitively expensive or risky without major structural repair or replacement, system modifications or replacement, or other actions not considered normal day-to-day maintenance and upkeep.

3. *The product will be designed so that its reliability and service life will not be reduced due to the effects of being shipped by land, sea, or air or by periods of storage up to ___________ life units.*

4. All reliability requirements apply to the product as it will be used in the environment defined in Section ____________ of the Specification and in accordance with the operating and support concepts defined in Section ____ of the _____________. (Customer must indicate where this information is provided in the solicitation).

5. Other. (Customer should indicate other requirements or information pertinent to the required level of reliability).

TEMPLATE FOR DEVELOPING RELIABILITY PORTION OF A PROCUREMENT PACKAGE

STATEMENT OF OBJECTIVES

1.0 Program Objectives

a. The program is: (here the customer defines the program as: (1) multi-phased, (2) single-phase, or (c) one program with multiple contractors).

b. The objective of the program is to design, test, and manufacture (*) to satisfy the performance requirements of the specification to meet a need date of [date].

2.0 Contract Objectives. The contractor shall meet the following objectives.

2.1 Design, Analysis, and Test

Design the [*] to satisfy performance requirements as defined in [cite applicable section of RFP]. Perform such analysis and tests necessary to design the [*], to reduce risk, and to verify that the product meets all performance requirements.

2.2 Configuration Management

Establish a product baseline to define the configuration of the [*] with a verified capability to satisfy all performance requirements. Establish and maintain a management process to thereafter control the product's configuration for the life of the contract. Document the design of the product baseline through the use of engineering data.

2.3 Quality Control

Institute a quality program to ensure the [*] is produced in accordance with engineering data, measuring and test equipment are properly maintained, and that appropriate actions are taken for nonconforming materials.

2.4 Logistics

Develop and deliver all data necessary to support the [*] (including provisioning, installation, and reprocurement data and operating and repair manuals) consistent with the maintenance concept as stated in [cite applicable section of RFP]. All data shall be in a form and format compatible with existing government data systems.

*Name of the product

TEMPLATE FOR DEVELOPING RELIABILITY PARTION OF A PROCUREMENT PACKAGE

STATEMENT OF OBJECTIVES

3.0 Scope of the Contract

The scope of this effort includes all activities and work associated with the:

- design, development, and manufacturing of the [*]
- design, development, and manufacturing of all new support equipment required for the [*] consistent with the maintenance concept
- development and delivery of all documentation required for the installation, operation and support of the [*] consistent with the maintenance concept

4.0 Work to be Accomplished under the Contract

Analyses, testing, documentation, modeling and simulation, process development, and management activities associated with developing the [*] and associated support equipment and documentation.

5.0 Program Control

The contractor shall institute those controls necessary to manage the program performance, schedule, and risk and to allow the government clear visibility into these factors.

6.0 Management Objectives

The management objective is to allow the offeror the maximum flexibility in managing the development and manufacture of the [*]. Also, it is the government's objective to encourage innovation in the management of the effort while maintaining clear government visibility into program performance, schedule, and risk.

*Name of the product

12.3.2. Guidance for Selecting Sources

The reliability portion of a bidder's proposal can be evaluated using the criteria in Figure 12.3-1. In addition to the criteria listed in the figure, the customer should encourage and look for innovative approaches that achieve the reliability performance requirements in the most effective way. Also, the proposal should emphasize the following objectives:

(1) Understand the Customer's Reliability Needs - if the customer has not explicitly done so, determine the required level of reliability as measured by the user during actual use of the product. No matter the source of the requirement, determine the feasibility of achieving the required reliability and track progress toward that achievement.

(2) Thoroughly Understand the Design - understand the reliability of the design and the manufacturing processes involved.

(3) Integrate Reliability with the Systems Engineering Process - make the reliability activities conducted during design and manufacturing an integral part of the product and processes design effort. Ensure all sources (i.e., suppliers, vendors, etc.) of components, materials, etc. used in the product, design and manufacture those items in accordance with the reliability requirements.

(4) Design for Desired Level of Reliability - use proven design approaches to make it safe, economical, and easy to maintain.

(5) Validate the Reliability Through Analysis and Development Test - conduct analyses, simulation, and testing to uncover reliability problems, revise the design, and validate the effectiveness of the redesign.

(6) Monitor and Analyze Operational Performance - assess the operational reliability of the product in actual use to uncover problems, identify needed improvements, and provide "Lessons Learned" for incorporation in handbooks and for refining modeling and analysis methods.

NOTE: The following list is not all-inclusive and not all items necessarily apply to every program

Understanding. Does the proposal show a clear understanding of:

- the importance of designing in reliability?
- reliability techniques, methodology and concepts?
- the importance of integrating reliability activities into the overall systems engineering process?

Approach

- **Management.** Does the proposal identify:
  - who is responsible for reliability and his/her experience and qualifications?
  - the number and experience of reliability personnel assigned to the program and the level of effort allocated to reliability activities?
  - how reliability personnel fit in the program's organizational framework?
  - an effective means of communication and sharing of information among reliability engineers and analysts, design engineers, manufacturing engineers, and higher management?
  - the suppliers' system for controlling the reliability of items from other suppliers and vendors?
  - how the supplier implements concurrent engineering practices and integrates reliability into the overall engineering and manufacturing effort?

- **Design.** Does the proposal explain:
  - if and how design standards, derating guidelines, and other criteria will be used?
  - if and how trade-off studies will be used for critical design areas?
  - the time-phasing of reliability activities in relation to key program milestones?
  - any areas of reliability risk?
  - if and how software reliability will be addressed?

- **Analysis/Test.** Does the proposal identify and describe:
  - methods of analysis and math models to be used?
  - reliability modeling, prediction, and allocation procedures?
  - the time phasing of any proposed reliability testing in relation to the overall program schedule?
  - the time available for the test type required (such as maximum time for reliability demonstration) and how that time was determined?
  - if and how the supplier will predict the reliability (in whatever parameters are specified) prior to the start of testing?
  - the resources (test chambers, special equipment, etc.) needed to perform all required testing, how they were determined, and their availability?
  - how the results of all testing will be used to evaluate reliability and identify reliability problems?

FIGURE 12.3-1: CHECKLIST FOR EVALUATING RELIABILITY PORTION OF A PROPOSAL

Compliance

- **Design.** The proposal should include:
  - evidence of compliance with military and commercial specifications and standards, when required, and good engineering practices for reliability.
  - justification (models, preliminary estimates, data sources, etc.) to back-up the claims of meeting reliability requirements

- **Analysis/Test.** The proposal should indicate:
  - an explicit commitment to perform all reliability analyses cited in the proposal or required by contract.
  - an explicit commitment to perform all reliability testing cited in the proposal or required by contract.
  - that the supplier complies with all product-level reliability test requirements and that the reliability figures-of-merit will be demonstrated by test using specified accept/reject criteria or by analysis.
  - if and how the contractor will perform verification and demonstration testing, the type of testing planned, and the specific purpose of the testing.

- **Data.** The proposal should show an explicit commitment to deliver all required reliability data items in the format specified.

FIGURE 12.3-1: CHECKLIST FOR EVALUATING RELIABILITY PORTION OF A PROPOSAL (CONT'D)

12.4 Reliability Program Elements

Once reliability has been quantitatively specified, a major challenge confronting Government and industry organizations is the selection of reliability engineering and management activities that can materially aid in attaining program reliability requirements. Each activity must be judicially selected to reflect funding and schedule constraints and tailored to the specific needs of the program.

Although the activities may vary among programs, some *elements* (i.e., categories of activities) have become common to most programs and companies. Consequently, they form a good basis from which to begin the selection of individual activities. Table 12.4-1 lists these common elements. For each element, the table lists some of the related activities. Note that many of the activities selected for a reliability program may be performed for other purposes (e.g., safety) or require inputs from other related activities (e.g., testability analysis). Accordingly, good communication and coordination among disciplines are essential. Integrated Product Teams is one way to ensure good communication and coordination.

TABLE 12.4-1: COMMON RELIABILITY PROGRAM ELEMENTS

ELEMENT	TYPICAL ACTIVITIES
Planning and Control	• Developing a reliability program plan • Monitoring and controlling subcontractors
Design	• Design Reviews • Developing design criteria • Parts selection • Derating • Identifying critical items • Robust design (fault tolerance, redundancy, graceful degradation)
Analysis	• Failure Modes, Effects and Criticality Analysis • Fault Tree Analysis • Sneak Circuit Analysis • Analysis of operating and environmental stresses • Modeling and allocations • Thermal Analysis
Testing	• Reliability growth testing • Reliability qualification testing • Environmental stress screening • Verification testing • Functional testing • Failure Reporting and Corrective Action System
Production	• Statistical Process Control • Inspection • Process Failure Modes and Effects Analysis
Other	• In-service reliability

12.5 Phasing of Reliability Program Activities

The previous section identified individual reliability program elements. The management, selection and control of activities within those elements must be based on recognition of the system's life cycle. Appropriate reliability management and engineering tasks must be performed during all life cycle phases.

The successful management of reliability during the system life cycle assumes the following premises:

(1) As defined by DoD Regulation 5000.2-R, there are four definable life cycle acquisition phases in the creation of any system, namely: Phase 0, Concept Exploration (CONCEPT); Phase I, Program Definition And Risk Reduction (DEF); Phase II, Engineering And Manufacturing Development (EMD); and Phase III, Production, Fielding/Deployment And Operational Support (PF/DOP). Each of these phases are defined as follows:

(a) <u>Concept Exploration Phase:</u> is the period where feasibility of alternative concepts are defined and evaluated and a basis for evaluating the relative merits of these alternatives at the next decision milestone are provided.

(b) <u>Program Definition and Risk Reduction Phase:</u> is the period when one or more candidate solutions are defined and design approaches and technologies are pursued as warranted. Risk reduction activities are also performed including prototyping, demonstrations, and early operational assessments.

(c) <u>Engineering and Manufacturing Development Phase:</u> is the period when for the most promising design approach, detailed design is developed, integration testing conducted, and a manufacturing capability established.

(d) <u>Production, Fielding/Deployment and Operational Support Phase:</u> is the period when systems are produced, operational tests and demonstrations are conducted, additional support provided, and modifications are incorporated as needed.

(2) There is a special role for reliability in each of these phases. To achieve the goals in the deployment phase requires planned actions in all previous phases. Each phase has specific objectives and the activities conducted during the phase must support these objectives. Milestone decision points mark the beginning and end of the acquisition phases. The milestone decision points are:

(a) Milestone 0 - Approval to Conduct Concept Studies (beginning of phase 0)

(b) Milestone I - Approval to Begin a New Acquisition Program (end of Phase 0, beginning of Phase I)

(c) Milestone II - Approval to Enter EMD (end of Phase I, beginning of Phase II)

(d) Milestone III - Production or Deployment Approval (end of Phase II, beginning of Phase III)

For some products, a phase may be "abbreviated" or even "skipped." For example, the research and development phase for a new product that is simply an updated or moderately improved version of an older, mature product will likely be very short, concentrating only on the differences between the two. Figure 12.5-1 shows the life cycle phases, milestone decision points, the objectives of each phase, and a summary of the activities associated with each phase.

12.5.1 Reliability Activities by Life Cycle Phase

The reliability activities conducted during each of the life cycle phases of a product must be consistent with and support the overall objectives for the phases. In the following discussion, reliability activities will be discussed in the context of the phase(s) in which they are most applicable. It is not practical to try and address all possible types of products, so the discussion assumes that a major product, such as an aircraft, tank, turbine engine, or similar item, is being developed. It should be obvious that the level of effort and types and scope of activities that would be necessary for a smaller system, such as a radar altimeter, will not be the same as for a new tactical fighter or other large system.

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

ACQUISITION PHASES

	0 Concept Exploration	I Program Definition and Risk Reduction	II Engineering and Manufacturing Development	III Production, Fielding/Deployment, and Operational Support
Objectives	1. Define and evaluate the feasibility of alternative concepts to meet deficiencies 2. Provide the basis for assessing relative merits of alternative concepts at next decision milestone	1. Define program as one or more concepts 2. Pursue design approaches and parallel technologies	1. Translate most promising design approach into stable, producible, supportable, and cost effective design 2. Validate manufacturing and production processes 3. Demonstrate product capabilities through testing	1. Produce and manufacture the product 2. Deploy the product 3. Operate and maintain the product
Key Activities	1. Validate assessment of need 2. Conduct concept studies 3. Develop product requirements in terms of effectiveness and performance measures	1. Conduct prototyping, demonstrations, and early operational assessments to reduce risk 2. Update product requirements 3. Identify accomplishments to be completed in Phase 2	1. Achieve design stability 2. Verify adequate resources have been programmed for production, deployment, and support 3. Conduct Initial Operational Test and Evaluation (IOT&E)	Manufacturing Use of the product in the intended environment, scheduled and unscheduled maintenance, improvements through updates and modifications, performance tracking
	▲ Milestone 0 Decision Point	▲ Milestone I Decision Point	▲ Milestone II Decision Point	▲ Milestone III Decision Point

MILESTONE DECISION POINTS

FIGURE 12.5-1: LIFE CYCLE PHASES OF A PRODUCT

12.5.1.1 Phase 0 - Concept Exploration

During the concept exploration phase, reliability activities are necessarily intended to prepare for Phase I. The reliability program plan begins with broadly stating the goals and objectives for the new product. Some analysis may be made of prior similar products to help establish ranges of realistic reliability goals. Very general modeling may also be used to complement the analysis in deriving ranges of goals. Also, new approaches and technologies related to reliability design, analysis, and validation can be identified during this phase. It is important to remember that the reliability program plan should be developed as part of the overall systems engineering effort. The interrelationship (inputs, outputs, etc.) between reliability tasking and other program tasks must be clearly stated and coordinated throughout the development life cycle.

The reliability activities to be considered in this phase are listed in Table 12.5-1. The specific activities implemented will depend on the specific requirements of any individual program. Therefore, Table 12.5-1 should be viewed as a guide only, and does not necessarily represent any "typical" program.

TABLE 12.5-1: RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE CONCEPT EXPLORATION PHASE

Reliability Program Planning	Reliability Modeling and Preliminary Allocations
Reliability Trade-off Studies	Test Strategy
Parts and Materials Programs	Benchmarking
Design Reviews	Quality Function Development
Supplier Control	Market Survey
Life Cycle Planning	Analysis of operational environment
Critical Item Control	

12.5.1.2 Phase I - Program Definition and Risk Reduction

For the alternative concepts that are carried into this phase, the reliability effort becomes more intense and focused and additional detail is added to the program plan. Additional analysis is required to begin refining requirements for the next phase of acquisition. Reliability engineers should be participating in and supporting trade studies in which the various alternatives are compared, different design approaches are evaluated, and overall system requirements are optimized. Some program and design reviews are usually held during this phase, and the issue of reliability must be considered during these reviews. The emphasis during these early reviews will be to choose among the alternative concepts. Preliminary modeling, using high-level reliability block diagrams of the various concepts for each design, may be needed. In addition, the reliability concept must be evaluated to ensure that the necessary and proper general design attributes are assigned to each product element. Data from whatever prototyping, proof-of-concept demonstrations, and similar "testing" is conducted should be analyzed and the results used in evaluating the relative reliability of each concept and in determining realistic reliability characteristics for the product. The reliability activities to be considered in this phase are presented in Table 12.5-2.

TABLE 12.5-2. RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE PROGRAM DEFINITION AND RISK REDUCTION PHASE

Reliability Program Planning	Reliability Predictions
Environmental Characterization	Thermal Analysis
Fault Tolerance	User Requirement Translation
Part and Material Program	Benchmarking
Reliability Modeling and Allocations	Market Survey
Durability Assessment	Software Reliability
Life Cycle Planning	

The emphasis on each chosen activity in this phase is as a means to evaluate one design concept/approach against a competing concept/approach. Information is still not as detailed as following phases, but will have more input than previous phases. For instance, reliability predictions may still be based on the parts count methods or on similar equipments/designs. Likewise, reliability modeling and allocations will be done at higher levels of abstraction. In choosing which specific tasks are to be performed, critical requirements or factors must be considered. If one approach favors more redundancy over another, for example, then a fault tree analysis may be a task chosen as a means to evaluate the reliability impact of this approach. If interfacing problems are a key issue, then some level of tolerance analysis may be warranted. In all cases, the reliability program manager must work in conjunction with all other members of the systems engineering team to clearly define the key issues such that reliability tasking can be judiciously and economically chosen. Further, close coordination with related functions such as maintainability, logistics and quality is necessary to avoid duplication of effort and to optimize resources.

12.5.1.3 Phase II - Engineering and Manufacturing Development

Usually only one concept is carried into Phase II. As indicated in Figure 12.5-1, the objectives of this phase are to:

(1) translate the most promising design approach into a stable, producible, supportable, and cost effective design

(2) validate the manufacturing and production processes to be implemented in Phase III

(3) demonstrate product capabilities through testing

During this phase, the design of the product is matured. The processes that will be used to manufacture and produce the product are developed. Development test and evaluation (DT&E) and some initial operational test and evaluation (IOT&E) is conducted to verify the design and demonstrate the product's performance. Consequently, it is during this phase that the reliability effort is most intense.

General and specific design criteria, standards, and policies for reliability are defined and implemented. Requirements for suppliers are developed based on allocations of product-level reliability requirements. Reliability analyses are conducted to evaluate the evolving design, identify problems, and develop solutions to those problems. Modeling and simulation are used as part of the evaluation effort. More detailed predictions and estimates of reliability are made. As testing and demonstrations occur, data from these events are collected and analyzed to refine the estimates of the design reliability. Data collection, analysis, and corrective action is an essential activity during this phase because it supports the analysis, design, and evaluation efforts. Table 12.5-3 presents the reliability program activities to be considered during this phase.

TABLE 12.5-3: RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE ENGINEERING AND MANUFACTURING DEVELOPMENT PHASE

Reliability Program Planning
Critical Item Identification
Derating Limits
Design Reviews
Fault Tolerance
Part Selection and Application
Thermal Design Limits

to better define the activities to be conducted during production, deployment, and operational support. Supplier control will continue during production and may continue into the operational support phase. During the operational support portion of this phase, efforts must continue to track and evaluate proposed design changes (due to modifications and upgrades), track and evaluate the operational reliability of the product, identify problems (ideally before they occur), and develop solutions for these problems. Reliability design, analysis, and test will be required to support modifications, whether those modifications are made to address reliability specifically or for any other reason (e.g., safety, upgrade functional performance, extend life, etc.). Those reliability activities to be considered during this phase are listed in Table 12.5-4.

TABLE 12.5-4: RELIABILITY PROGRAM ACTIVITIES TO BE CONSIDERED IN THE PRODUCTION, DEPLOYMENT, AND OPERATIONAL SUPPORT PHASE

Critical Item Control	Environmental Stress Screening (ESS)
Supplier Control	Production Reliability Acceptance Test (PRAT)
Failure Reporting Analysis and Corrective Action System (FRACAS)	Statistical Process Control (SPC)
Part Obsolescence	Inspection

The objective of many of the activities will be to monitor the effects of production testing on the reliability of the product. Data collected will be fed back into the detailed analyses performed in the previous phase to update results, allocations, predictions, etc. and to ensure that the impacts on reliability are not critical. Many of the elements will be continuations, therefore, of those elements chosen for the previous phase. The intensity of the effort will once again be related to criticality of requirements.

12.6 R&M Planning and Budgeting

The previous subsection provided information on reliability activities and in which program development phase these activities should be considered. Further guidance on program planning is provided in this subsection to assist in determining key issued to be considered in each phase.

The most basic of management functions is planning. Planning is deciding in advance what to do, how to do it, when to do it, and who is to do it. Budgeting, which goes hand in hand with planning, involves insuring that adequate resources, financial or otherwise, are available to carry out the plan to achieve the desired goal. This is the essence of the "Reliability Program Plan" task described in the previous subsection.

Reliability planning cannot be done in a vacuum; it is an effort that must be dovetailed into the overall system program development plan. In the concept exploration phase, for example, the choice of system design alternatives must include reliability estimates and projected costs in order to select the most cost effective system. In later development stages, reliability estimates are needed for system support planning for spare parts, repair and rework facilities, and personnel

training. Hence, reliability is a key element in overall program planning, and from this planning should emerge a set of realistic, cost effective, reliability objectives.
Of course, planning includes the budgeting process of allocating the necessary resources to implement the plan. Without proper budgeting, planning is an empty exercise. Accordingly, in the following discussions, planning assumes proper budgeting.

12.6.1 Conceptual Exploration Phase Planning

In this phase, system reliability estimates are necessary to identify the best possible system alternative and to provide a valid picture of the cost effectiveness of the proposed system for comparison with other system alternatives. Reliability estimates in the concept phase must be based on historical data.

Reliability planning in the concept phase should address:

(1) Definition and refinement of realistic reliability requirements to be finally demonstrated during EMD tests.

(2) Parts selection criteria using available Qualified Manufacturer’s Listing (QML), (Ref. [3], [4]) devices and devices procured to best commercial parts and processes (BCP) to the maximum extent possible. Guidance for BCP is provided in Reference [5]. Critical parts in terms of technology or reliability must be identified so that the program provides for the procurement of these special parts in a timely manner.

(3) Planning for tracking reliability progress through the development life cycle to provide a continual measure of achieved versus required values.

(4) Identification of program review milestones for assessing reliability progress.

(5) Adequate manning and budgeting to ensure competent reliability planning and surveillance of the contractor's efforts, and the possible need to use outside activities for reliability support.

(6) Interfacing with eventual user and support commands on reliability plans and requirements.

12.6.2 Program Definition and Risk Reduction

In this phase, hardware will be developed, perhaps by competing contractors, and reliability planning will focus on contractual requirements. Under acquisition reform policies, only performance specifications are now developed by the acquisition activity. Previous to acquisition reform, solicitations would have included a statement of work (SOW) as well; however this now may be the contractor’s responsibility. The key reliability issues when developing the performance specification are:

(1) Quantitative reliability requirements must be specified and defined and the hardware must be inherently capable of achieving the required reliability. This requirement must be based on a translation of the user's reliability needs into performance requirements such as a probability of success or a mean time between failure (MTBF) value.

(2) Demonstration testing can still be called out as a requirement, but it is up to the contractor to select the best means of implementation. For newer and unproven designs, demonstration testing is recommended.

(3) Parts selection must be controlled. All substitute parts should be identical in form, fit, and function to the preferred parts to preclude difficulty with including preferred parts in later systems.

(4) Reliability design trade-off studies should be performed to include design for reliability, redundancy options, optimum repair level analysis, failure mode analysis, and any other analyses required to optimize the design.

(5) Reliability predictions should be performed and continually refined as the design matures to provide an indication of potential reliability of the system.

(6) Program and design reviews are essential for control and motivation of the entire reliability program and to ensure that the detailed reliability design effort is progressing according to plan.

12.6.3 Engineering and Manufacturing Development (EMD) Phase Planning

Essential differences between the previous phase and the EMD phase are:

(1) During validation, the realism of reliability requirements must be established, system trade-offs made, and systemic reliability problems identified and eliminated.

(2) During EMD, the requirements are firm, and the program geared toward implementing final design decisions and providing adequate demonstration tests to ensure that reliability requirements will be met.

During the early part of EMD, the contractor should prepare reliability test plans which are important key program documents. Such plans provide the execution details of any planned reliability demonstration tests. Together with unambiguous requirements, carefully planned reliability tests are essential elements during the EMD phase.

It is essential that, during EMD, adequate budgeting be provided for both the government and contractor to perform the necessary reliability program functions. All too often, budgeting for these activities is not given proper priority in the total program budget estimates.

12.6.4 Production, Deployment, and Operational Support Phase Planning

In the production phase, reliability activities will be concerned with the following:

(1) Finding and fixing problems arising during production. These will be primarily process problems, workmanship problems, and parts defects, since most design problems will have been resolved in previous phases.

(2) Performing stress screening, and periodic verification tests to identify and correct reliability degradation during production runs.

(3) Ensuring that quality/reliability control procedures are given required attention.

(4) Evaluating engineering change proposals (ECPs) for their effects on reliability.

During deployment and operation, reliability activities will be concerned with:

(5) Data collection to track field reliability performance.

(6) Establishment of test criteria and controls (and analyses of storage data) to ensure the readiness of equipment and material items during storage.

(7) Analyses of field data to determine significant areas for reliability improvement.

12.7 Trade-offs

Throughout the system acquisition process, system engineers, designers, and acquisition managers are confronted with decision problems concerning the selection of one solution from among many alternatives. The term "trade-off" as it applies to decision-making is defined as the procedure by which several alternatives are evaluated to provide a solid basis for choosing only one. It is essentially a system optimization problem under a series of constraints. This was discussed in Section 4.

Trade-off studies are an inherent part of the design process, and are performed in sequence beginning at the highest system level parameters and proceeding downward to equipment design details. For example, as was shown in the previous section, in the early phases of system design, trade-off studies are performed at the broad system level, e.g., trade-off of the performance, cost, schedule and risk parameters to arrive at the “best” alternative solution.

As design proceeds and requirements become firmer, trade-off studies are performed involving detailed system parameters, e.g., reliability, maintainability, availability, safety, logistics supportability, and life cycle costs. As these parameters become fixed, trade-offs are performed within each parameter. For example, in reliability trade-off studies, one might study the following options: more reliable parts, design simplification, component derating, reliability growth, or redundancy.   Even within each of these parameters further trade-off studies may be needed; for example, active versus standby redundancy might be considered. Consequently, many of the reliability elements described in Section 12.4 are included for the purpose of trading off one design approach versus another.

Previous sections have described trade-off procedures for the design engineer. Following is some guidance on the types of trade-off studies which should be performed, and when.

12.7.1 Concept Exploration Phase Trade-off Studies

Trade-off studies should be performed among reliability, maintainability, safety, performance, physical configuration, environmental use conditions, and other system requirements and design constraints to provide the basis for design optimization by the system activities in the concept phase. The analyses must be kept current with each design iteration of each alternative consideration. Dynamic feedback of analytical results should be provided to system engineering and concept design activities for guidance in performing design iterations. These studies should typically include the following:

(1) Performance Analysis: Evaluate reliability as a function of mission performance characteristics. Plot reliability functions for each of several possible alternative definitions of “acceptable” performance.

(2) Maintainability Analysis: Evaluate reliability vs. maintainability under alternative design concepts and life cycle cost objectives for specified levels of availability.

(3) Availability Analysis: Evaluate reliability and maintainability trade-offs for several “acceptable” levels of availability and for several alternative approachs to availability assurance, e.g., design redundancy, pre-mission system operability testing, preventive maintenance, etc.

(4) Life-Cycle Cost Analysis: Evaluate the cost of reliability and maintainability acquisition (for several levels of performance) vs. the cost of maintenance and support in the deployment phase.

(5) Schedule/Risk Analysis: Evaluate the technical risks and schedule requirements associated with the reliability and maintainability acquisition objectives.

(6) Operational Suitability: Combine the results of the preceding analyses to produce a family of design configurations which would satisfy the operational requirements with a quantitative assessment of operation suitability, logistics supportability, life-cycle costs, and acquisition schedule projected for each configuration.

(7) Select and Verify Configuration: Select the best all-around configuration from those described in (6), and reassess the feasibility of achieving the reliability and maintainability requirements and the potential for the selected design configuration.

12.7.2 Program Definition and Risk Reduction Phase Trade-off Studies

During this phase, the contractor's system analysis involving reliability trade-offs against other design parameters is reviewed to verify realism, completeness and objectivity in predictions, allocations, and simulation analyses made on each design configuration considered. Contractor reliability data required for in-process review of this task include a current, updated version of earlier analysis, to verify that the contractor's proposed allocations are consistent with the mission models for the design, considering relative importance and duty cycle of constituent end items. Procedures (and data) by which requirements are allocated to equipment and lower end-item levels must be revalidated. Reliability and maintainability requirements must be defined in quantitative terms for integration into the allocated baseline specifications for constituent end items of the system.

The trade-off and system analysis should typically include the following:

(1) System Description: Verify system description in terms of functional and physical configuration, performance limits associated with primary and alternate modes of operation, maintenance concept applicable to the design, equipment utilization factors, and mission profiles for the defined missions.

(2) Reliability Modeling: Validate block diagrams, taking into consideration redundancy possibilities, alternate modes, and back-up system capabilities.

(3) Data Validity: Validate equipment failure rates and repair rates, etc., used in the simulation study.

(4) Reliability Allocations: Verify consistency of allocated design requirements for each constituent subsystem, equipment, and separately procured end item of the system; and verify that minimum acceptable reliability and maintainability requirements to be demonstrated by test correspond to the allocated design requirements.

(5) Test Requirements: Verify adequacy and applicability of any reliability

demonstration test requirements, conditions, and acceptance criteria for each allocated requirement.

(6) Feasibility Study: Validate feasibility estimates for each of the allocated values, based on current design configuration; evaluate differences between specified, predicted, and allocated reliability for each subsystem; evaluate alternative approaches under consideration by system engineering, to achieve the specified requirements.

(7) Problems: Review problems identified within each subsystem/equipment; verify criticality ranking, corrective action requirements, and estimated growth potential available through problem correction. Identify areas where further system design and operational analyses are required to determine equipment essentiality, back-up capabilities, etc. Approval of design analyses and reliability trade-off study results at this point are contingent on satisfying the following criteria:

(a) Conformance: Allocated reliability requirements, when recombined at the system level, must satisfy system reliability requirements defined in the functional baseline specification.

(b) Validity: Analytical procedures and data used in the trade-off studies must be proven valid by independent assessment.

12.7.3 Trade-offs During Engineering Manufacturing Development (EMD), Production, Deployment and Operational Support Phases

During EMD, the contractor is involved in detailed design trade-off studies concerned with aspects of design philosophy such as level and allocation of redundancy, reliability and maintainability test methods and procedures, built-in versus external test equipment philosophy, maintenance concepts, etc. This is the phase in which the "paper design" of the preceding phases is transformed into working hardware for test and evaluation. Hence, during this phase, the role of the acquisition manager and his or her staff is primarily one of acting as reviewers - reviewing designs, program plans, and test plans to ensure that they are in consonance with the specification requirements, and that the desired results will be achieved.

This would involve evaluating the results of design analysis and reliability engineering trade-off studies involving considerations of safety, redundancy, failure mode/effects, critical reliability factors, degrading interface tolerances, power levels and regulation, physical dimensions, packaging and environment control features and requirements, etc., underlying the configuration selected for production.

Subsequently, within the framework of the previous system studies, contractors and their subcontractors will carry out trade-off studies at progressively greater levels of detail. These studies will address such factors as testability (test equipment needed and schedules) optimum thermal design, power supply requirements, component choices, and circuit layouts.

The preceding sequence of trade-off studies, starting with broad issues and converging into equipment details, will, in general, be concluded by the end of the EMD and Production Phases. However, additional involvement of all parties will be required, even during the deployment and operational support, to assess the desirability of proposed modifications arising from field experience and use.

For example, during field operation, it will be necessary to verify that any negative effect of individual Engineering Change Proposals (ECPs) on system reliability and maintainability is acceptable from the overall mission effectiveness viewpoint, as determined from a trade-off study with the other system parameter changes for which the change was designed.

12.8 Other Considerations

In addition to selecting and implementing reliability program tasking in support of meeting system requirements and performance of trade-off analyses, as described above, there are other concerns that will need to be addressed which can affect the reliability program. In this section, the following three areas will be discussed: Software Reliability, Life Cycle Costs and Warranties, and Reliability Program Evaluation and Surveillance.

12.8.1 Software Reliability

Despite the fact that software reliability has not reached the sophisticated stage of evaluation as hardware reliability, there are some procedures available which a manager can use to help achieve the desired level of software quality and reliability. Admittedly, these procedures are not geared solely to reliability achievement; however, their proper and timely application has been shown to enhance the reliability of the developed software.

Software is frequently only one part of a total system, but there is an ever increasing use of software to provide system functionality that was previously provided by non-computing hardware. In military systems especially, software now controls a major portion of the overall system functionality. Thus, there are many instances (e.g., *concurrent engineering, integrated product development)* where software must be developed in parallel with the hardware to create system functionality. It is essential that the approach to reliability and quality takes a total system view both for the product and for the development process. Figure 12.8-1 illustrates the degree of coordination that must occur for the hardware and software efforts to reach an effective system development process.

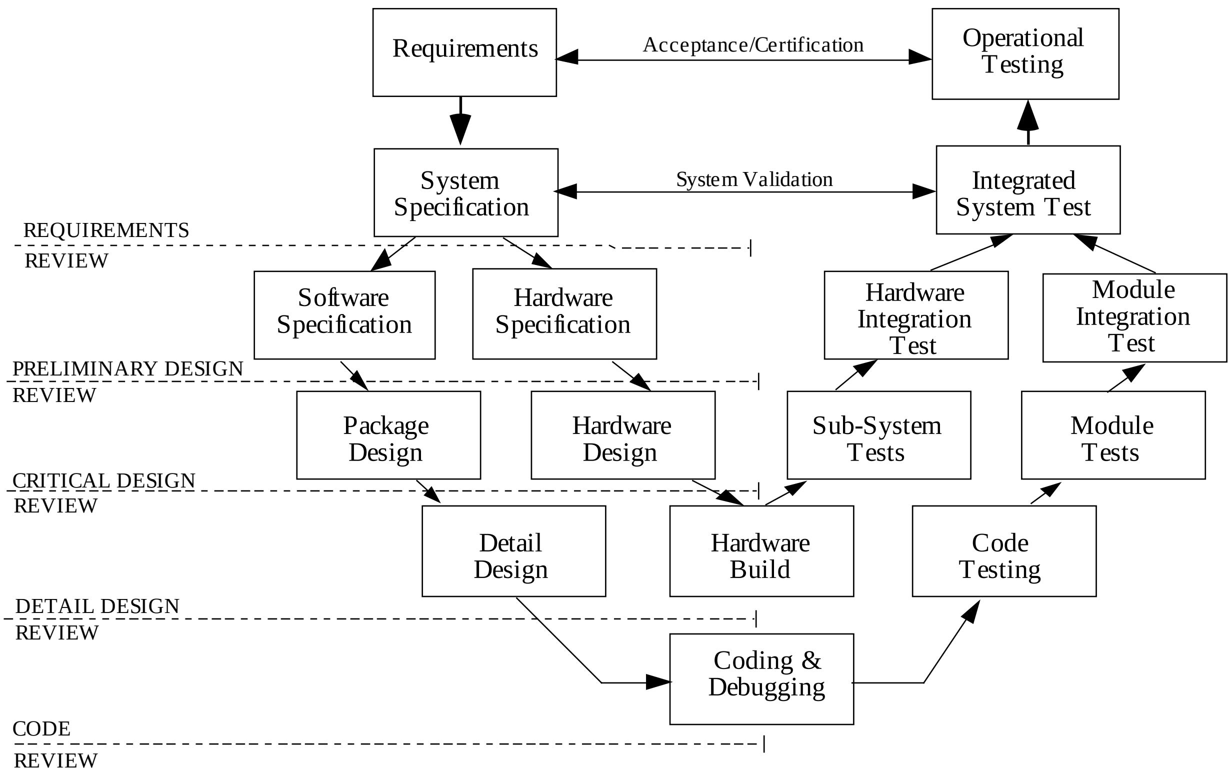


FIGURE 12.8-1: CONCURRENT SYSTEM DEVELOPMENT PROCESS FOR BOTH HARDWARE AND SOFTWARE (REF. [6])

Each of the phases shown in Figure 12.8-1, pertaining to software development is summarized in the following subsections.

(1) The requirement phase involves performing preliminary hardware/software trade-offs to produce a statement of system requirements. The statement will provide specific system functional specifications/requirements as well as the constraints (design, cost, etc.) that the system must meet.

(2) In the preliminary design phase, the requirements are translated into well defined functional specifications. Detailed hardware/ software trade-offs are performed, and a design approach is selected among the various alternatives. The computer program design specification is prepared during this phase, and a preliminary design review is normally held at the end of this phase to assess the adequacy of the selected approach.

(3) During the critical and detailed design phases, the software component definition, interface structure, modularity, and data definition are developed and verified against the requirements. Functional flow charts and detailed flow charts are prepared. Detailed flow charts are used to define the

information processing in terms of logical flow and operations to be performed by the computer program. The relationship between the computer program and the interfaces between the software, the computer(s), and other peripheral devices are also defined at this time. A preliminary computer program product specification is prepared at the completion of this phase. At the end of the design phase, and prior to the coding and testing phase, a design review is usually held to establish the integrity of the flow charts and the preliminary computer program specification.

(4) During the coding and debug phase the detailed design is translated into actual program code, and the initial testing of the code is performed. This initial testing normally is designed to check for correct outputs using predefined inputs.

(5) In the integration and test phase, the computer program modules are tested against the requirements as stated in the preliminary program specifications, and, once tested, the software package is integrated with other modules and tested. The computer program product specification is finalized during this phase.

(6) During the integrated system tests, the computer programs are loaded and run to ensure that the system performance meets requirements. The system is completely documented during this phase, and all changes resulting from the previous phases are incorporated into the supporting documentation, including the flow charts and final product specification.

Overall management must begin with the development of system requirements and continue through preparation of specifications during system analysis, interact with design and development efforts and extend through control of changes. Reliability analysis must be performed as part of early system analyses (trade-offs) to establish the optimum levels of reliability to be achieved in both hardware and software design. These analyses must extend through design and development to further define reliability requirements to establish the basis for meaningful integration tests. The test program must include module and system testing during development to force out design errors, and system integration and acceptance testing prior to delivery, to assure that the requirements are met.

Figure 12.8-2 (Ref. [7]) lists the principal software reliability elements, and shows the importance of each element during each of the system's life cycle phases. Also shown is the percentage distribution (column 2) of contractor man-hour effort for the various elements for an "average" program. Each of the elements of Figure 12.8-2 is addressed in the following paragraphs.

12.8.1.1 Requirements Definition

Software requirements define the overall mission problem to be solved by the software, the operational constraints, and any fixed interfaces with system hardware and people. Requirements must cover the following:

(1) Problems to be solved by the software system

(2) Software-related system hardware design decisions

(3) Software design constraints imposed on the system

(4) Input data sources, rates and formats (if established)

(5) Output data destinations, rates, and formats (if established)

(6) Software-dependent maintenance concepts and plans

(7) Security needs

(8) Operational hazards and environment

(9) Reliability and maintainability needs

12.8.1.2 System Analysis

System analysis proceeds in parallel with requirements definition, and evaluates the system design trade-offs between hardware and software. It considers computer hardware options, maintenance options, and in general, all of the software-related hardware alternatives. The objective is to design the hardware/software system so as to maximize the chances of success at the lowest life cycle cost. These chosen design options are documented in system and interface specifications used by the software designers. The first set of “A”s on Figure 12.8-2 refers to delivery of these hardware parameters to the specification writers.

Another important area of system analysis which continues through the middle of engineering manufacturing development, is the development of schemes for system testing and acceptance. The thoroughness of these schemes directly effects the verification of software reliability. Test schemes are documented in the system test plans and acceptance specifications. The second set of “A”s on Figure 12.8-2 refers to delivery of this test planning information to the test plan writers.

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

Element	% Effort	Life Cycle Phase: Concept Exploration / Program Definition and Risk Reduction / EMD / Production, Fielding/Deployment and Operational Support
Requirements Definition	5	xxxxxxxxxxAAAAA...
System Analysis	10	xxxxxAAAxxxxxxxxxxxxxxAAA.............
Package Design	10	---ooooxxxxxxxxx..
Unit Design, Code & Debug	15	---ooooooooooxxxx..
Module Integration & Test	30	----oooxxxxxAAA...
System Integration & Test		xxAAAAxx...
Acceptance Test		xxxAAA.....................
Program Plan	30	---ooooooxxxxxxxxxxxxxxxxxxxxxxxxx..............................
Specifications		xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxAAAA.......................
Data System		oooooooooooooooooooooooooxxxxxxxx...
Program Review		ooooooxxxxxxxxxxxxxxxxxxxxxxxxxxAAAA.......................
Test Plan		----ooooxxxxxxxxxxxxxAAAA.......
Technical Manuals		ooooooxxxxxxxxxxxAA...

First contract

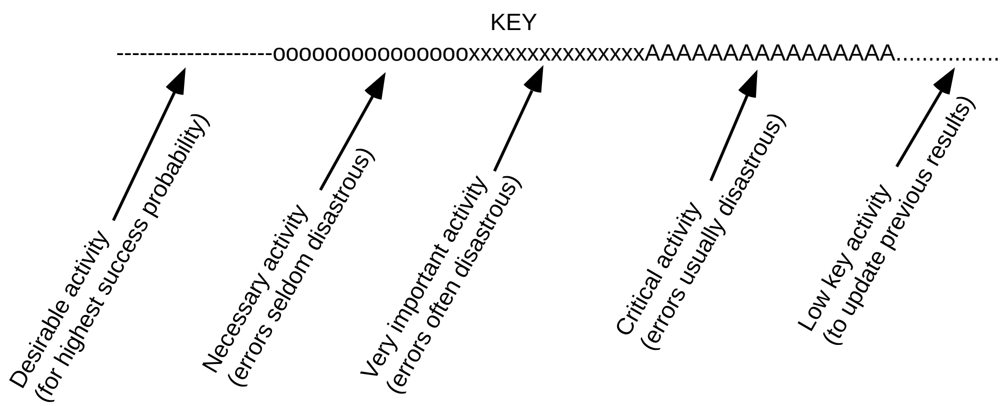


FIGURE 12.8-2: SOFTWARE RELIABILITY PROGRAM ELEMENTS BY PROGRAM PHASE

12.8.1.3 Package Design

Package design refers to the development of the complete software system functional organization. That is, the programming hierarchy of tasks of the software system are defined in terms of categories and subcategories, all the way down to the unit level. (The process is analogous to organizing a large group of people with diverse skills to carry out a project). To enhance reliability of a large software system, this software functional organization must be thorough, well documented, and all interface rules between functional elements must be precisely defined and their application carefully controlled.

A “chief programmer” or a senior software system engineer is usually assigned to oversee and manage this whole process. Subordinate programmers responsible for the separate programs in the functional categories are assigned to this individual. The subordinate managers will plan, organize, direct and control the detailed coding, testing, and documentation of programming within their domains using the ground rules laid down by the chief programmer. In addition to organizing the whole operation, the chief programmer must identify the source program languages to be used (from system analyses documented in the system and interface specifications) and the general rules for program structure and progress documentation throughout the software development organization. The programming rules should be documented in one of the computer program design specifications.

To enhance the readability and testability of the computer programs, “structured programming” techniques should be employed. In part, this means that the programmer is restricted to a small set of standard language constructs which prevent skipping to some remote segment of the computational sequence. This approach reduces the possibility of logical traps or “dead-ends.”

12.8.1.4 Unit Design, Code and Debug

Another attribute of “structured programming” is the size restriction on program units or models. The unit is typically defined to be about 50 lines of program code which will fit on one listing page. Furthermore, the unit will have only one link from the preceding unit and one link to the following unit. These rules enhance readability, comprehension, and independent testability of each unit. Each manager will supervise the design, code, debug, and test of his or her group's output.

12.8.1.5 Module Integration and Test

Module integration and test means that programs are assembled and tested in groups of increasing size until the entire software package is put together. This assembly and testing is usually done with the aid of general purpose computers, since the operational hardware computer may not be available. A test plan is used throughout this process, and results are documented in the data system.

12.8.1.6 System Integration and Test

System integration and test means that the integrated software program is inserted into the operational hardware, and complete system tests are run to ensure that hardware and software are compatible and that operational requirements can be fulfilled. This element is also critical to verification of operational suitability. It occurs in the final phases of development. A test plan is used to conduct these tests.

12.8.1.7 Acceptance Test

The software acceptance test is defined in a test plan, and possibly in an overall system acceptance specification. This test is the final test which formally establishes acceptability of software products for delivery under the development or production contract.

12.8.1.8 Program Plan

The program plan outlines and explains all elements of the software development effort. It shows requirements, interfaces, organization, task breakdown, responsibilities, schedules, and the approach to solving all the software development problems so as to fulfill the requirements on schedule and within projected cost. This plan is developed during concept and early program definition phases, but must be continuously updated.

12.8.1.9 Specifications

Specifications formally and precisely document all requirements and design decisions. They may be grouped into several categories:

(1) System Specification: Defines the system requirements and the overall hardware/software system design in top level detail.

(2) Software Performance Specification: Defines the software requirements, software design ground rules, selected software-dependent hardware parameters, interface identification, and overall structure of the software system. This specification goes into a level of detail below the System Specification.

(3) Interface Specifications: Defines the interface design details between software and hardware elements and between software subdivisions. It goes into a level of detail below the preceding Software Performance Specification.

(4) Software Design Specification: Defines and describes the computer programs that will meet the Software Performance Specifications in functional flow diagram detail. It also defines the programming scheme and rules which will be used by programmers to implement the functional elements in computer code.

(5) Subprogram Design Document: Gives a detailed technical description of each subprogram including input, output, functional flow, narrative description, limitations, interfaces, and mathematical equations solved or operations performed. It also describes the tests used to check it out.

(6) Common Data Base Design Document: Gives a detailed technical description of all data items used by the software system. This includes constants, variables, and tables. Details include data name, table index, purpose, dimensions, units, initial values, range of values, exact format, etc.

(7) Acceptance Specification: Defines the criteria to be used in judging formal acceptability of software products under contract.

12.8.1.10 Data System

The data system, also called the program support library, is designed to provide management control information and documentation discipline. It will consist of some kind of periodic reporting procedure where every programmer will be required to submit at least a weekly report on his effort. The reports might include estimates of coding, completion of assigned units, numbers and classifications of errors found in debugging and testing, information shortages which hamper coding progress, specification errors discovered, man-hours spent on separate units, documentation contributions, etc. Listings of each run are also collected and stored in this system. The chief programmer will have an administrative staff to compile the reports into composite summary charts, graphs and narratives for use in management reviews. The data system must also cover status of the documentation, and some very disciplined scheme must be devised to ensure that documentation keeps up with changes in requirements, system design and software design.

12.8.1.11 Program Review

In the Government program reviews, overall program progress is reviewed and compared with the Computer Program Development Plan. Also, a technical review of the software is performed by software specialists from Government laboratories or specialists from some other advisory organization. These reviews are documented with action items assigned to the Government or contractor for resolution by specified dates.

Typical formal reviews include: the systems requirements review (SRR), the system design review (SDR), the preliminary design review (PDR), and the critical design review (CDR). The

SRR is conducted after a significant portion of the system functional requirements have been established, and is used to evaluate responsiveness to the statement of work and the interpretation of the system requirements. The SDR is conducted prior to the beginning of preliminary design, and is used to review system documentation and assess the degree of accomplishment of the engineering management activities.

12.8.1.12 Test Plan

Several test plans are prepared during the software development cycle to define procedures for package integration and test, and system integration and test. These plans explain who does what and when. They may also specify test requirements down to the unit level. These test plans are developed from data provided by requirements, system analysis, package design, and unit design. These test plans are used to define the test problems to be solved by the software along with acceptable solutions. Reliability test criteria are, of course, included.

12.8.1.13 Technical Manuals

While the various specifications and design documents described previously document the exact structure of the software, they are not necessarily suitable for field use in training and operations. The technical manuals are written using those specifications and documents, but are written by people who know how to convey that information to operating personnel in the most effective way. The manuals normally include the following types:

(1) User's Manual

(2) Computer Operator's Manual

(3) Software Maintenance Manual

(4) System Maintenance Manual

12.8.2 Cost Factors and Guidelines

To this point, cost and budgeting has been alluded to without providing any detailed guidance. This section will present information on controlling costs and planning options that are available to support cost control throughout the entire life cycle of the system.

Most military equipment/system acquisition managers and their counterparts in the commercial world must cope with four basic, usually conflicting, criteria, which are:

(1) Performance

(2) Cost

(3) Schedule

(4) Risk

The goal is to balance these criteria so as to obtain the "best" system. With the increasingly high costs of buying, operating, and maintaining weapon systems, further exacerbated by the reduction in defense spending over recent years, the term "best" has come to mean developing a system with minimum life cycle costs (LCC) consistent with required performance.
This balanced design approach is shown in Figure 12.8-3 in which design engineers and acquisition managers must balance performance, reliability, and unit production costs equally against the overall objective of minimizing the cost of ownership or LCC.

An important fact that the manager must keep in mind is that early design decisions lock-in a major portion of the life cycle costs. This is shown graphically in Figures 12.8-4 and 12.8-5 (Ref. [8]).

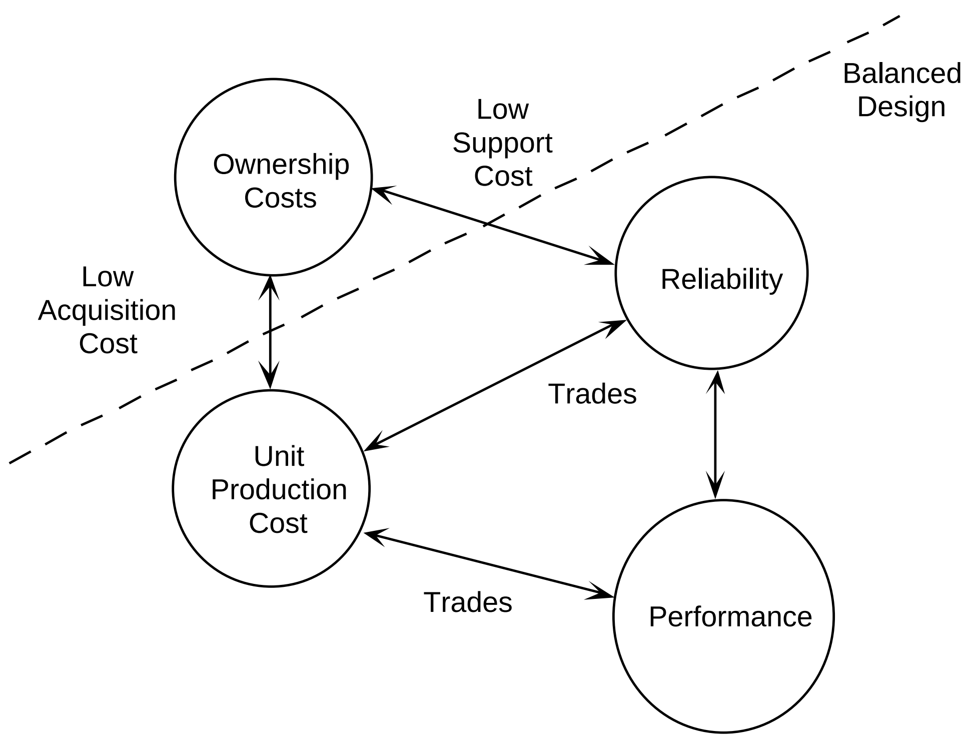


FIGURE 12.8-3: BALANCED DESIGN APPROACH

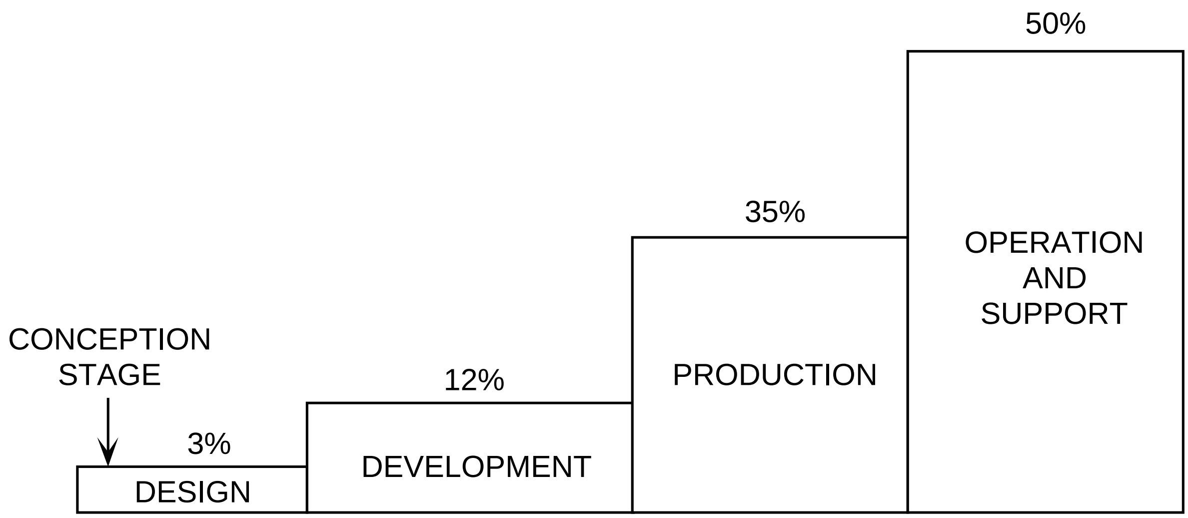


FIGURE 12.8-4: EXPENDITURES DURING LIFE CYCLE

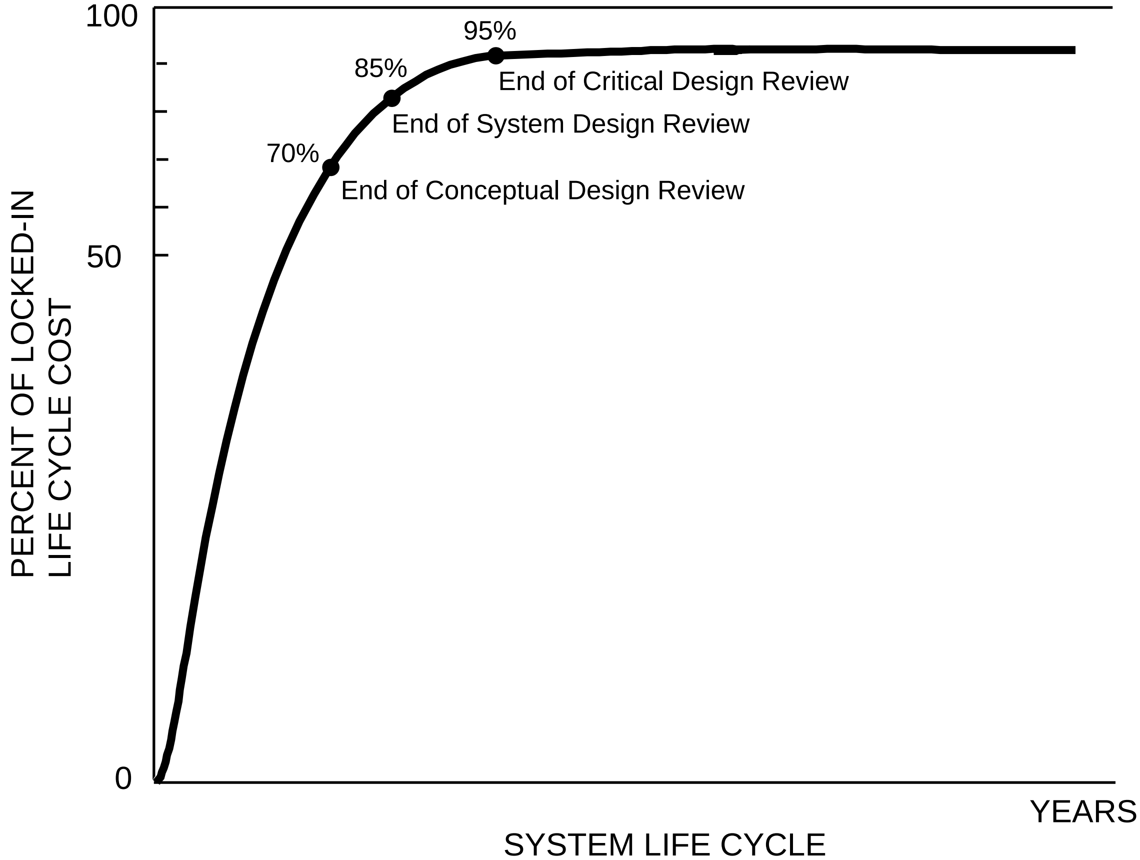


FIGURE 12.8-5: EFFECT OF EARLY DECISION ON LIFE CYCLE COST

These figures relate dollar expenditures and percent of locked-in life cycle costs to the life cycle of a project. These figures are held as being representative for the US Department of Defense. Figure 12.8-4 shows that the design and development phase of a project consumed only 15% of the cost of a typical project, as opposed to 35% for the production phase and 50% for the in-service phase. Although only 15% of the expenditures were made prior to production, Figure 12.8-5 shows that about 90-95% of the life cycle costs were determined. The design specifications that were approved prior to production determined how it would proceed and, therefore, determined the costs to be incurred in that phase. Similarly, the detailed specifications were produced based upon a certain operational, maintenance and supply support policy. These policies and the design dictate such in-service variables as manpower, consumables and spares levels.

The significance of these figures should be kept in mind by the acquisition manager. Prior to the conceptual design review, 100% of the design can be altered and 100% of the life cycle cost can be affected. Completion of the conceptual design review gives approval for the basic framework of the design. The concepts approved, although not a written set of specifications, place constraints on the design team, narrow their decision horizon and fix a certain level of the life cycle cost on the project. As time progresses, the decision horizon narrows and a greater percentage of life cycle costs become determined. It has been estimated that by the time 15% of a typical system's life cycle has expired, 90% of the life cycle costs have been determined. Thus, a manager needs to be familiar with the available tools to enable him to make timely decisions to minimize LCC.

Reliability as well as maintainability decisions have a great impact on LCC. The frequency of failures and the time to repair them determine the resources, manpower, and materials needed to maintain the system in the field. The principal difficulty which confronts the acquisition manager in making R&M decisions is the complexity of the problem. The equipment R&M requirements defined in the development specification establish the objectives of the design. However, these must be considered in conjunction with numerous other requirements and constraints, all of which influence operations and support costs.

12.8.2.1 Design-To-Cost Procedures

Design-to-cost goals are used in contracts to seek the best balance between performance and acquisition cost in a system development program. The original intent of the use of design-to-cost procedures was to slow the trend of continually increasing acquisition costs due to emphasis on achieving the ultimate in system performance.

Design-to-cost (DTC) can take different emphasis dependent on the type of development program. Four programs with varying design-to-cost emphasis are shown in Table 12.8-1. As seen in the table, "Design-to-Unit-Production-Cost" (DTUPC) has been emphasized in past major military programs. DTUPC can determine the number of aircraft or equipment that the DoD can "afford." DTC policies and objectives are delineated in MIL-HDBK-766 (Ref. [9]).

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

As is shown in Table 12.8-1, most DTC contractual requirements have emphasized unit production cost. However, the ultimate goal is to minimize the cost of ownership or LCC. Minimizing unit production cost is only one step toward achieving the ultimate goal. Emphasizing this step and ignoring the ultimate goal could conceivably result in compromising reliability requirements, thus resulting in increased support costs. This means that one should strive for a design which will

(1) Maximize performance within unit cost goals

(2) Minimize support cost to minimize LCC

In other words, DTC and LCC must be jointly considered.

TABLE 12.8-1: TYPES OF DESIGN-TO-COST PROGRAMS

Design-to-Cost Programs	Program Characteristics	Program Examples
Production Unit Price	• Large Quantity Procurements	• Close Support Aircraft A-10 • Lightweight Fighter
Total Program Costs	• Complex Equipment • Small Buys • High Development Cost	• AWACS • Advanced Airborne Command Post
Production Unit Cost and Installation Cost	• Large Quantity Procurement of Subsystems	• Airborne Radar • Avionics Equipment • TACAN • Gyroscope
Development and Operating Costs	• Facilities and Construction Programs	• Ground Radar Installations

12.8.2.2 Life Cycle Cost (LCC) Concepts

LCC is defined as the total cost to the government of acquisition and ownership of a system over its full life. It includes the cost of development, acquisition, operation, support, and eventual disposal. Figure 12.8-6 is provided as a guide for the acquisition manager in terms of the activities that should be performed at each phase of a system's life cycle in order to minimize LCC. They are quite self explanatory, and, for the interested reader, are treated in greater detail in Section 10.

From a managerial point of view, for LCC to be successful it must be an explicit part of the original contract competition. Competition in system development and production serves to place a "downward pressure" on the estimates of equipment production costs proposed by competing suppliers. Recognizing that competition will almost certainly cease to exist at entry into the production phase of a program, the objective of LCC competition is to obtain as much assurance as possible prior to production that the selected equipment will satisfy the requirement for lowest practical life-cycle cost. To accomplish this, the competitive phases of an LCC program are structured with emphasis on identifying and reducing the life-cycle cost drivers. In addition, in a properly planned development program where the participating contractors are thoroughly briefed on the importance of LCC and where provisions exist for extensive development testing to validate cost-related parameters (e.g., reliability), competition serves to induce each contractor to address cost-risk design problems which would otherwise not be encountered until after production was underway.

12.8.3 Product Performance Agreements

A means that may be used to reduce life cycle costs while improving the performance of an item in the field is the use of warranties and guarantees (one form of a warranty). The following definitions are provided for warranty and guarantee:

(1) Warranty: a contractual obligation that provides incentives for the contractor to satisfy system field operational objectives of the user. The contractor is given an incentive, through a fixed price commitment, to repair or replace equipment found to be defective during the period of warranty coverage.

(2) Guarantee: a commitment embodying contractual incentives, both positive and negative, for the achievement of specified field operational goals.

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

Concept Exploration Phase	Program Definition and Risk Reduction Phase	Engineering and Manufacturing Development Phase	Production, Deployment, and Operational Support Phase
a. Identify LCC cost drivers. b. Perform sensitivity performance level cost trade studies. c. Identify optimum R&M levels. d. Develop LCC procurement approaches e. Define design to cost (DTC) requirements. f. Review LCC Provisions including DTC/RIW. g. Perform cost of ownership assessment. - operation - support h. Conduct comparison cost studies of current systems/equipments i. Review LCC provisions - DTC - RIW	a. Refine R&M cost trade-off analysis. b. Perform cost of ownership assessment. - operation - support c. Review LCC procurement requests. d. Conduct source selection. e. Develop LC provisions (DTC, RIW). f. Initiate development cost monitoring plan. g. Initiate LCC guidelines for RFP and source selection.	a. Review contractor LCC provision/goals. b. Define LCC procurement requirements. c. Conduct detailed cost of ownership analysis. d. Prepare LC achievement test plan/procedure. e. Monitor contractor LCC plans and achievement of goals. f. Establish "cost of change" review guidelines.	a. Perform Engr Change Proposal (ECP) reviews. b. Develop warranty/guarantee selection guidelines. c. Conduct LCC achievement test. d. Collect field cost data. - operation - support e. Perform cost comparison studies. f. Perform up-date studies, modify and retrofit to reduce cost.

FIGURE 12.8-6: LIFE CYCLE COST ACTIVITIES

The use of warranties in the commercial marketplace should be familiar to most readers. Such warranties provide protection to the consumer against a defective product for a specified time period. This "protection" comes in various forms, depending on the product, from total replacement of the defective product with another one, to coverage of all parts and labor costs needed to correct a deficiency.

In the DoD, the use of failure-free warranties and reliability improvement warranties (RIW) were used on a limited basis in the late 1960s and early 1970s for the purpose of studying the success such programs have in improving performance and reducing LCC. Success of the early programs prompted more detailed variations, particularly for RIW and MTBF guarantees during the mid-1970s.

In 1980, the Air Force published a Product Performance Agreement Guide (PPA), which expanded the warranty concept to areas such as software, repair/exchange agreements, logistics support, etc. In 1982, the Product Performance Agreement Center (PPAC) was established by the Air Force as a focal point for use of product performance agreements/warranties. The PPAC revised the 1980 PPA guide in 1985 (Ref. [10]). The use of warranties in the acquisition of military systems as a common option led to the passage, in 1984, of the Defense Procurement Reform Act of 1984, PL 98-525, Title 10 US Code, Paragraph 2403 (referred to hereafter as 10 USC 2403) (Ref. [11]).

The central theme of 10 USC 2403 is the mandate to warrant essential performance requirements (EPRs). As defined by 10 USC 2403, EPRs are "operating capabilities and/or reliability and maintenance characteristics of a weapon system that are determined by the Secretary of Defense (or delegated authority) as necessary for the system to fulfill the military requirement for which it is designed." Guidance on the selection and use of warranties can be found in references [10] and [11]. The remainder of this subsection will briefly define the types of warranties that can be considered, and a discussion of the more commonly used warranties oriented toward reliability characteristics.

An understated part of 10 USC 2403 is that the procuring activity must conduct analyses and studies sufficient to determine if the use of a warranty is appropriate and in the best interests of the government.

12.8.3.1 Types of Product Performance Agreements

This subsection, and all remaining material, will provide information on warranties considered for the procurement of DoD systems. where appropriate, these warranties could be tailored for similar commercial systems. The following are brief descriptions of a number of warranties presented in more detail in references [10] and [11].

Maintainability Guarantee. Applicable to critical, potentially high MTTR end items/components on fixed price contracts. The objective is to reduce the MTTR during maintenance and/or overhaul.

SECTION 12: RELIABILITY MANAGEMENT CONSIDERATIONS

Reliability and Maintainability Improvement Warranty. Applicable to fixed price contracts, normally for avionics equipment at the LRU level. Applicable to SRUs if fault capability isolation at the user level is available. Purpose is to motivate producer to increase equipment R&M and reduce repair costs. Applies to preventive and corrective maintenance.

Warranty of Supplies. This agreement is applicable to fixed price contracts for stable design items or equipment. It extends the contractor's responsibility for materials, workmanship, and specification conformance beyond the period of acceptance of supplies.

Warranty of Technical Data. This agreement is applicable to either cost reimbursement or fixed price contracts. It provides for correction or replacement of deficient data for a specified time after delivery and inspection.

Rewarranty of Repaired/Overhauled Equipment. This warranty is applicable to fixed price contracts and provides warranty coverage for items which have been overhauled, repaired, or furnished as replacements by a contractor. The contractor agrees to warrant that the repaired or replacement parts and/or materials are free from any further defect in material or workmanship for a specified period.

Repair/Exchange Agreements. When the volume of repair activity for an item being introduced into the DoD inventory is expected to be too low to justify organic support, the repair/exchange agreement can provide an alternate approach. The contractor must establish the capability to exchange complete items or to repair parts returned to his facility within agreed-upon turnaround times.

Reliability Warranty. The contractor agrees to maintenance/overhaul intervals for components and/or subsystems. When specific types of failures occur between overhaul intervals in covered items, the contractor is responsible for the supply of a specified combination of labor, material, or replacement items.

Reliability Improvement Warranty (RIW). Applicable to fixed priced contracts. Normally applied to avionics equipment at the LRU level, or at the SRU level if fault capability isolation at the user level is available. Under RIW, the contractor agrees to repair all covered failures for a specified period at no additional expense to the Government. This warranty is designed to increase equipment reliability and reduce repair costs.

Mean-Time-Between-Failure Verification Test (MTBF-VT). The MTBF-VT can be used to achieve improvement in operational reliability. The MTBF-VT can be applied at the "black-box" or subsystem level or components from several subsystems can be aggregated to system level commitment. The test of compliance would normally be scheduled for the first deployed unit. Deviations between MTBF targets and measured performance form the basis for rewards or corrective action.

Reliability Improvement Warranty (RIW) with an MTBF Verification Test (MTBF-VT. Applicable to fixed price contract as a means to motivate producer to increase equipment reliability and decrease logistics support costs. Applicable to avionics equipment LRUs or black boxes that can be tested without entry into the unit.

R&MIW with MTBF/VT. Applicable to fixed price contracts for avionics equipment LRUs, or SRUs if the fault isolation capability is at an organic level. Provides producer with a strong incentive to achieve specified R&M in the field.

Component Reliability Warranty. Corrects deficiencies and improves product performance for selected system components. Designed for use on components for which certain minimum levels of operational performance are critical to overall satisfactory operational performance. Oriented toward “fleet-wide” reliability minimums, specified at the component or black box level, over a specified time period.

Spare Parts Level Warranty. Applicable to fixed price contracts for equipment or items which are prime mission essential or safety of operationally essential, and designed for organic Government maintenance. Objective is to maintain the original system capability with lowered mean time between (LRU or SRU) removals (MTBR).

Availability Guarantee. The Availability Guarantee can be used to reduce downtime for systems or equipments which operate in a continuous mode or with dormant systems where readiness upon random demand is a critical requirement. The equipment should provide a positive indication of operability either through continuous performance checks or, in the case of dormant systems, through go/no-go checks.

Logistics Support Cost Guarantee (LSCG). The LSCG is used to control and reduce selected aspects of life cycle cost and to improve equipment supportability in operational use. The LSCG uses a cost model which describes the effect that system design, operating, and logistics characteristics have on potential support costs. The model addresses those features of the equipment which impact support investment and recurring operations and support costs. Deviations between target logistics parameters and measured performance form the basis for rewards or corrective action.

Maximum Parts Cost Guarantee. The Maximum Parts Cost Guarantee can be applied to equipments when repair costs are critical. The contract specifies an average repair cost (which can include parts and labor) for the system or critical portions thereof. “Actual” average repair cost is then compared with the specified average to determine what remedy or consideration is applicable.

Utility Functions Guarantee. Utility function agreements can be applied to consumable items. The DoD establishes a utility function for the item being procured (e.g., landings per tire or set of brakes or starts per battery). The contractor specifies a value for this function. A demonstration is performed to develop an “actual” value of the utility function. The “actual” and “specified” values are compared to determine what remedy or consideration is applicable. This type agreement is often incorporated as the basis of life cycle cost procurement actions for consumable items.

Reliability Improvement Warranty (RIW) With a Mean-Time-Between-Failure (MTBF) Guarantee. The MTBF guarantee can be applied to systems where the objective is to achieve substantial reliability growth. An increasing series of target values is specified over consecutive time periods of the guarantee. Failure to meet target values results in contractor corrective action. Exceeding targets could result in incentive awards.

Chronic Line-Replaceable-Unit (LRU) Guarantee. A Chronic LRU Guarantee can be applied to LRUs where mean time between removals (MTBR), mean time between failure (MTBF) or similar reliability criteria are an important consideration. During the period of the guarantee, any LRU which experiences an extraordinary number of consecutive removals is designated a “chronic LRU.” The contractor is required to replace chronic LRUs and chronic LRUs are not counted in calculating actual MTBR or MTBF results.

Mean-Time-to-Repair (MTTR) and Mean-Time-Between-Unscheduled-Removals (MTBUR) Guarantees. MTTR and MTBUR Guarantees can be used on systems and subsystems where downtime or frequency of maintenance are critical to equipment performance. Measurements of achievement under operational conditions will be made over a series of specified intervals. When measured achievements fall outside of acceptable limits, a specified remedy is required.

Ultimate Life Warranty. The Ultimate Life Guarantee can be applied to basic elements of a system such as aircraft structure, engines, and landing gear. A value is established for the life of the item and a remedy identified if failures occur.

Commercial Service Life Warranty. A Commercial Service Life Guarantee can be used to extend limited term warranty coverage to the service life of the item.

Software Design Commitment Guarantee. This guarantee may be applied prior to a production contract award since its purpose is to provide contractors with an incentive to develop software packages with inherently high quality, low maintenance and update costs. As part of a development contract, quantitative targets for parameters such as modularization, documentation, testability, and transportability are established for software and demonstration requirements.

LRU Software Configuration Control and Support Agreement. This guarantee may be applied to software packages associated with a system or other specific set of LRUs. The contractor agrees to be responsible, at no additional cost, for software changes due to associated changes that are the contractor's responsibility. The contractor is also responsible to maintain software configuration and documentation.

Fault Detection, Isolation, and Repair Warranty. Applicable to fixed price contracts for operational systems that are intended for organic Government support. Intent is to reduce Mean

Troubleshooting Time (MTT) for the guaranteed system/item to a guaranteed level, and maintain that reduced MTT for a specified period of time. Producer will correct deficiencies in the FMEA, Optimum Repair Level Analysis (ORLA), TOs, Test System Hardware, or Test Software when they impact the operational availability of the end item.

Test and Repair Improvement Guarantee. This guarantee may be applied to the test equipment and test procedures that are developed for a system. The contractor guarantees that his test equipment and procedures, when applied in accordance with applicable documentation, will demonstrate MTBR (or MTBF) characteristic of systems in field operation. When comparisons between operational and test results fall outside specified boundaries, the contractor is responsible to make changes to the test equipment or procedures.

Method of Test Guarantee. This guarantee is intended to ensure that the unique test equipment and test methods used for specified LRUs will accurately verify the performance of the LRUs during an agreed-upon period of time. The contractor, at no additional cost, will replace, modify, or repair test equipment and methods when deficiencies occur. A demonstration will be conducted to determine compliance with this guarantee.

Quality of Training Warranty. Applicable to fixed price contracts for Items intended for organic Government maintenance. Intent is to ensure the level of skill and knowledge available in the repair shops at all levels of maintenance. Contractor corrects voids in system training brought about by configuration changes and oversights.

12.8.3.2 Warranty/Guarantee Plans

Three of the most commonly used plans to improve reliability and reduce support costs have been the RIW, MTBF guarantee and the LSC guarantee.

Table 12.8-2 highlights the principal features of these three basic types of warranty-guarantee plans that have been used in DoD procurements. The following paragraphs briefly describe the plans; more details are provided in the cited references [10] and [11].

Reliability-Improvement Warranty (RIW). The RIW plan commits the contractor to perform stipulated depot type repair services for a fixed operating time, calendar time, or both, at a fixed price. While the major expenditures of a warranty procurement are for the repair services involved, the primary objectives are to secure reliability improvement and reduce support costs. The question of whether the contractor can provide depot repair services at a cost lower than that of military repair is secondary to the objective of reliability achievement.

TABLE 12.8-2: FEATURES OF CURRENT WARRANTY-GUARANTEE PLANS

Features	RIW	RIW/MTBF	LSG
Objective	Secure reliability improvement/reduce support costs	Achieve stated reliability requirements/reduce support costs	Achieve stated logistic-cost goal
Method	Contractor repairs or replaces all applicable items that fail during coverage period; implements no-cost ECPs to improve reliability	Same as RIW; in addition, contractor provides additional spare units to maintain logistic pipeline when MTBF goals are not met	Normal Air Force maintenance; operational test performed to assess LSC: penalty or corrective action required if goals are not achieved
Pricing	Fixed price	Fixed price	Fixed price or limited cost sharing for correction of deficiencies
Incentive	Contractor profits if repair costs are lower than expected because of improved R&M	Similar to RIW, plus possible severe penalty for low MTBF	Award fee if goal is bettered; penalties for poor cost performance

Under the RIW, the producer typically agrees, prior to production, that the delivered equipment will achieve a specific reliability level (MTBF) before expiration of the warranty period. In return, the producer is paid a fixed warranty price for each warranted item as part of the procurement contract. Typical warranty periods range from two to five years. While the warranty agreement is in effect, the producer will perform all necessary repairs to failed equipment. The agreement may also contain settlement provisions which delineate the producer's liability in the event the reliability goal is not achieved. During the warranty period, the incentive for the producer is to minimize his outlay for repair and potential settlement liability by closely monitoring the actual reliability, and implementing improvements which promote reliability growth.

MTBF Guarantee. The MTBF guarantee requires the contractor to guarantee that a stated MTBF will be experienced by the equipment in the operating environment. If the guaranteed level is not met, the contractor is typically required to institute corrective action and to provide consignment spares until the MTBF improves.

The MTBF guarantee is normally procured in association with an RIW. The RIW plan provides a solid incentive for MTBF achievement through the contractor maintenance support commitment. The MTBF guarantee provides an even stronger incentive because the contractor is obligated to provide consignment spares to relieve pipeline shortages that may result from low MTBF. The MTBF plan also includes requirements for improving the MTBF to stated values. The added risk the contractor takes in providing this guarantee will be reflected in his bid price. The procurement organization must then determine if the protection provided is cost effective in relation to the price.

Logistic Support Cost Commitment. The logistic support cost (LSC) commitment is another means of controlling an equipment's operational effectiveness. Under this plan the contractor makes a contractual commitment regarding a specified LSC parameter, which is quantified through an LSC model. A controlled operational field test is subsequently performed to acquire data for the key variables in the LSC model. The measured LSC parameter is then compared with the contractually specified or target value. There is considerable variation among LSC commitment plans regarding the action taken as a result of the operational test. Most plans, in the event of achieving a lower measured LSC, provide for an award fee predicated on the amount by which the goal is underrun. In the event of an overrun, the plans provide for reducing or eliminating the award fee. In addition, some plans have required the contractor to take corrective action to achieve the stated goals or be penalized monetarily. In recognition of the risk inherent in this concept, the contractor bids a fixed price for undertaking a commitment where corrective action may be required. These types of plans are considered to fall under, or are an adjunct to, correction-of-deficiencies (COD) clauses. In the event the cost of correcting deficiencies exceeds the contractor's bid amount, provision may be made for Government and contractor cost sharing the overrun up to some specified ceiling. Costs beyond the ceiling must be borne solely by the contractor.

12.8.4 Reliability Program Requirements, Evaluation and Surveillance

This subsection will provide guidance on both specifying reliability requirements based on the type of procurement, followed by guidance for the procuring activity on issues to be considered when evaluating contractor’s proposed reliability effort as well as surveillance of the contractor effort after contract award.

12.8.4.1 Reliability Program Requirements Based Upon the Type of Procurement

This section of the handbook discusses basic program requirements within the framework of the guidance provided in section 12.4, which would form reliability programs considered applicable to the procurement of military systems. There are three major categories of procurements that exist to meet a specified need and are:

Existing Commercial. Commercial procurements provide for the purchase of existing hardware systems in order to obtain a low cost, quick response capability for certain requirements. Advantages of this type of procurement include use of a proven design, reduced lead-times and minimal development expense. Possible disadvantages associated with commercial procurements include inability to meet reliability requirements, limited performance, parts availability, reduced control of model changes, and increased logistic support requirements.

Commercial procurements seldom require analysis to specific reliability levels. Criticality in terms of mission requirements is normally low and the cost of acquisition may be optimal if the equipment is an off-the-shelf or commercial type item and no new development is required. Procurement of commercial equipment requires effort to select items with "as is" suitability and demonstrated acceptability to meet project needs. Specification efforts should be restricted to describing only those requirements in functional terms necessary to assure hardware acceptability. Design requirements are to be specified only to the extent necessary and essential to satisfy procurement requirements. The description and specification of additional reliability and quality controls should be avoided. Validated commercial tests should not be repeated. Procurement emphasis is in selection, not specification. Among the factors to be considered when selecting commercial products are

(1) Identification of one or more established products that appear suitable

(2) Analysis of all available data

(3) Consideration of industrial standards

(4) Reliability, maintainability, service life and spare parts availability

(5) Estimate of the extent to which reliance can be placed on warranties

Modified Commercial. Modified commercial procurement provides for use of the basic commercial configuration with modifications to meet certain specifications. Possible advantages to this form of procurement are quicker availability and lower development cost than a new military design item. Possible disadvantages include the loss of integrity of the commercial product, the addition of unproven components, and the compromise of mission capability.

Military Requirements. The procurement of systems to meet military requirements present the greatest challenge. Included are two subcategories:

(1) Existing development, sometimes called non-developmental items (NDIs) (production or build-to-print contracts). In this subcategory, the establishment of reliability levels is aided by the existence of previous demonstration or field data, prior reliability estimates, and judgment factors arising from the consideration of these data.

(2) New development which involves a completely new design or changes to major components and major redesign of existing system. New system development is characterized by the establishment of a program office.

The possible advantages of procuring a newly designed item are that the item can fully meet military requirements, that the design and configuration can be government controlled, and that the logistic support can be assured. Possible advantages of procurement of an existing design are the shorter lead times involved, the use of less costly changes to reach required performance objectives and the utilization of existing technology.

The determination of appropriate reliability specification levels as well as program task activities involves reviewing the type of contract in view of the reliability design requirements. The nature of the procurement (for example, commercial, military, etc.) will, to a large extent, dictate the quantitative reliability requirements developed by the government. The type of procurement will also dictate which reliability tasks are most applicable both for the prime contractor and their applicable subcontractors. If the hardware to be procured is an off-the-shelf, commercial product, performance of reliability prediction or reliability growth and demonstration testing is less likely. However, depending upon the design reliability level, acceptance and screening tests may be required. The relationship of the specified reliability levels to the state-of-the-art will also dictate the extent of the reliability program activities. If the specified reliability, for example, is close to the maximum that can be achieved within the state-of-the-art (i.e., if there is little room for reliability improvement) then, possibly, a very vigorous and intensive reliability program should be structured and implemented. The program in that case would then include reliability predictions, FMECA, reliability growth tests, demonstration tests, screening tests, and production acceptance tests to ensure compliance to specified requirements with high confidence. However, if the specified value is not stringent and there is ample room for reliability improvement, then the program would not have to be as extensive.

12.8.4.2 Reliability Program Evaluation and Surveillance

The procuring activity, in addition to preparing reliability requirements that are integrated into system specifications, the statement of work (in primarily commercial acquisition programs), and other contractual documentation, also evaluates proposals, reviews data deliverables (e.g., reliability program plans, predictions, analyses, etc.) participates in design reviews, prepares reliability responses and, in general, continually evaluates and monitors reliability program outputs throughout development and production. Specifically the contractor's reliability programs are evaluated and monitored to:

(1) Determine the effectiveness of specific programs

(2) Rate and compare different programs

(3) Track the implementation of reliability programs by surveying contractor's facilities, participating in design reviews and evaluating test plans, procedures and results

Several military organizations have developed checklists for evaluating and monitoring R&M programs. Examples of these are provided in Appendices to Section 7. These checklists can be directly applied or at least provide a basis, to formulate or tailor more specific criteria to evaluate and monitor R&M development and production programs in general. They should be used in conjunction with Section 12.5 which lists the reliability tasks to be performed during each life-cycle phase. Included in these checklists are evaluation considerations and monitoring criteria with respect to individual R&M tasks and control elements. It should be noted that in addition to the technical criteria associated with each task, certain aspects associated with management and control are covered. The intent is that each activity is evaluated and monitored with respect to management including their interaction with other activities within the framework of the overall R&M plan, as well as how each task impacts design activities. The guidelines cover overall R&M organization and control stress factors within the areas of organization, methods of control, planning, and reporting activities.

12.9 References for Section 12

1. Guide to Performance Specifications. Publication SD-15, April 1996.

2. Buying NDI, Publication SD-2, April 1996.

3. Hybrid Microcircuits, General Specification For, MIL-PRF-38534, 23 August 1995.

4. Integrated Circuits (Microcircuits) Manufacturing, General Specification For, MIL-PRF-38535, 31 October 1995.

5. Commercial Parts and Processes for Military Applications, Reliability Analysis Center, September 1996.

6. Wingrove, A.A., The Problems of Managing Software Projects, IEE/BCS Software Engineering Document, 1986.

7. Coppola, A., and A. Sukert, Reliability and Maintainability Management Manual. RADC-TR-79-200, AD-A073299, July 1979.

8. Arsenault, J.E., and J.A., Roberts, "Reliability and Maintainability of Electronic Systems," Computer Science Press, Potomac, MD, 1980.

9. Design to Cost, MIL-HDBK-766, 25 August 1989.

10. Product Performance Agreement Guide, The Air Force Product Performance Agreement Center, ASD/ALW - AFALC/XRCP, 1 November 1985.

11. Warranty Guidebook, A Reference for Use by DoD Managers in Developing, Applying and Administering Warranties, Defense Systems Management College, Fort Belvoir, Virginia, October 1992

CONCLUDING MATERIAL

Custodians:	Preparing Activity:
Army - SY	Air Force - 17
Navy - EC	
Air Force - 17	(Project RELI-0090)
DLA - DH	

Review Activities:
Army - CR, IE, PT, TM2
Navy - AS, CG, CH, MC, NW, SA, TD
Air Force - 02, 08, 10, 13, 19, 21, 33
DLA - CC
OSD - HS, MA

STANDARDIZATION DOCUMENT IMPROVEMENT PROPOSAL

INSTRUCTIONS

1. The preparing activity must complete blocks 1, 2, 3, and 8. In block 1, both the document number and revision letter should be given.

2. The submitter of this form must complete blocks 4, 5, 6, and 7.

3. The preparing activity must provide a reply within 30 days from receipt of the form.

NOTE: This form may not be used to request copies of documents, nor to request waivers, or clarification of requirements on current contracts. Comments submitted on this form do not constitute or imply authorization to waive any portion of the referenced document(s) or to amend contractual requirements.

I RECOMMEND A CHANGE:	**1. DOCUMENT NUMBER** MIL-HDBK-338B	**2. DOCUMENT DATE (YYMMDD)** 991001

ELECTRONIC RELIABILITY DESIGN

4. NATURE OF CHANGE *(Identify paragraph number and include proposed rewrite, if possible. Attach extra sheets as needed.)*

5. REASON FOR RECOMMENDATION

6. SUBMITTER

a. NAME *(Last, First, Middle Initial)*	b. ORGANIZATION	
c. ADDRESS *(Include Zip Code)*	d. TELEPHONE (Include Area Code) (1) Commercial (2) AUTOVON *(if applicable)*	**7.DATE SUBMITTED** (YYMMDD)

8. PREPARING ACTIVITY

a. NAME AIR FORCE RESEARCH LABORATORY INFORMATION	b. TELEPHONE *Include Area Code)* (1) Commercial (315) 330-4205 (2) AUTOVON 587-4205
DIRECTORATE AFRL/IFTB 525 BROOKS ROAD ROME, NY 13441-4505	Defense Quality and Standardization Office (DLSC-LM) 8725 John J. Kingman Road, Suite 2533 Fort Belvoir, Virginia 22060-6221 (703) 767-6888 DSN: 427-6888