Springer
Berlin
Heidelberg
New York
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

A. Birolini

Reliability Engineering

Theory and Practice

Third Edition
with 120 Figures, 60 Tables, and 100 Examples

Springer

Prof. Dr. Alessandro Birolini
emerit. o. Prof. ETH Zurich
Via Generoso 5
6900 Lugano
Switzerland

First and second edition printed under the title "Quality and Reliability of Technical Systems"

ISBN 3-540-66385-1 3rd ed. Springer-Verlag Berlin Heidelberg New York
ISBN 3-540-63310-3 2nd ed. Springer-Verlag Berlin Heidelberg New York

Cip data applied for
Die Deutsche Bibliothek - CIP-Einheitsaufnahme
Birolini, Alessandro: Reliability engineering : theory and practice ; 60 tables /
A. Birolini. - 3. ed. - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; Milan ; Paris ;
Singapore ; Tokyo : Springer, 1999
Bis 2. Aufl. u.d.T.: Birolini, Alessandro: Quality and reliability of technical systems
ISBN 3-540-66385-1

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in other ways, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution act under German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1994, 1997 and 1999
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author; Cover design: medio, Berlin
Printing: Mercedesdruck, Berlin; Binding: Lüderitz & Bauer, Berlin
SPIN:10737463 62/3020-5 4 3 2 1 0 - Printed on acid-free paper

Preface

Reliability engineering is a rapidly evolving discipline, whose purpose is *to develop methods and tools* to predict, evaluate, and demonstrate reliability, maintainability, and availability of components, equipment, and systems, as well as *to support development and production engineers in building in* reliability and maintainability. To be cost and time effective, reliability engineering has to be coordinated with quality assurance activities, in agreement with Total Quality Management (TQM) and Concurrent Engineering efforts. To build in reliability and maintainability into complex equipment or systems, *failure rate* and *failure mode* analyses have to be performed early in the development phase and be supported by *design guidelines* for reliability, maintainability, and software quality as well as by extensive *design reviews*. Before production, *qualification tests* on prototypes are necessary to ensure that quality and reliability targets have been met. In the production phase, *processes* need to be selected and monitored to assure the required quality level. For many systems, *availability* requirements have also to be satisfied. In these cases, stochastic processes can be used to investigate and optimize availability, including logistical support as well. Software often plays a dominant role, requiring specific *quality assurance* activities.

This book presents the state-of-the-art of reliability engineering, *both in theory and practice*. It is based on over 25 years experience of the author in this field, half of which was in industry and half as Professor for reliability engineering at the ETH (Swiss Federal Institute of Technology Zurich). Following Chapter 1, in which basic concepts as well as the main tasks and organizational requirements for a cost and time effective quality and reliability assurance/management of complex equipment and systems are introduced, the book is structured in the following three parts:

1. Chapters 2 to 8 deal with reliability, maintainability, and availability *analyses* and *tests*, with emphasis on practical aspects in Chapter 3 (selection and qualification of components), Chapter 5 (design guidelines for reliability, maintainability, and software quality), and Chapter 8 (quality and reliability assurance in the production phase). This part answers the question *how to build in, predict, evaluate, and demonstrate reliability, maintainability, and availability.*
2. Appendices A1 to A5 deal with definitions, standards, and program plans for quality and reliability assurance of complex equipment and systems. This part addresses the needs of *project and quality assurance managers* and answers the question *how to specify and realize reliability targets.*

3. Appendices A6 to A8 give a sound introduction to probability theory, stochastic processes, and statistics. This part gives the mathematical foundations necessary for Chapters 2, 6, and 7, respectively, and addresses (together with Chapter 6) the needs of *system oriented engineers.*

Methods and tools are presented in such a way that they can be *tailored* to cover a range of reliability, maintainability, or availability requirements from low up to stringent. The investigation of repairable systems is performed systematically, starting from constant failure and repair rates between consecutive states (Markov processes) and generalizing step by step up to the case in which the process involved is regenerative with a minimum number of regeneration states (Chapter 6). The convergence of the point availability to its steady-state value is analyzed in detail. A correct explanation of the *waiting time paradox* is given (Appendix A7). Approximate expressions for the reliability and availability of complex repairable systems and for spare parts reservation are developed in depth (Chapters 6 and 4). The estimation and demonstration of a constant failure rate λ (or $MTBF = 1/\lambda$) and the empirical evaluation of field data are considered in detail (Appendix A8, Chapter 7). A new derivation of the confidence interval for an unknown probability p is given (Appendix A8). Methods and tools for the selection and qualification of (electronic) components and assemblies are presented in depth (Chapters 3 and 8). Design guidelines for reliability, maintainability, and software quality as well as checklists for design reviews are discussed extensively (Chapter 5, Appendix A4). Cost optimization is considered for several applications (Chapters 1, 4, and 8). Aspects of safety and risk management as well as trends in quality management systems are outlined (Chapters 1 and 2, Appendices A2 and A3). Many results are presented in tables or graphs. More than 100 practice oriented examples illustrate methods and tools.

Stochastic processes and tools introduced in Appendix A7 and Chapter 6 can also be used to investigate the reliability and availability of *fault tolerant systems* for cases in which a reliability block diagram does not exist, on the basis of an extended reliability *state transition diagram* (a publication on this subject is in preparation).

The book has served for many years (4th German ed. 1997, Springer) as a text book for three semesters teaching at the ETH Zurich and for courses aimed at engineers in industry. The basic course (Chapters 1, 2, 5, with an introduction to Chapters 3, 4, 6 to 8) should belong to the curriculum of every engineering degree.

This book is a careful update and revision of the book *Quality and Reliability of Technical Systems* (2nd ed. 1997, Springer), it aims to be a contribution to a sustainable development/world. The comments of many friends and the agreable cooperation with Springer-Verlag are gratefully aknowledged here.

Lugano and Zurich, June 1999 *Alessandro Birolini*

Contents

1 Basic Concepts, Quality and Reliability Assurance of Complex Equipment and Systems

Reliability engineering is a rapidly evolving discipline, whose purpose is *to develop methods and tools* to predict, evaluate, and demonstrate reliability, maintainability, and availability of components, equipment, and systems, as well as *to support development and production engineers in building in* reliability and maintainability. To be cost and time effective, reliability engineering has to be coordinated with quality assurance activities, in agreement with Total Quality Management (TQM) and Concurrent Engineering efforts. This chapter introduces basic concepts, shows their relationship, and discusses the main tasks and organizational requirements necessary to assure cost and time effectively the quality and reliability of complex equipment and systems. Refinements of these *management aspects* are given in Appendices A1 to A5. *Quality assurance* is used here in the sense also of *quality management* as per *TQM*.

1.1 Introduction

Until the ninteen-sixties, quality targets were deemed to have been reached when the item considered was found to be free of *defects* or *systematic failures* at the time it left the manufacturer. The growing complexity of equipment and systems, as well as the rapidly increasing cost incurred by loss of operation and for maintenance, have brought to the fore the aspects of *reliability, maintainability, availability*, and *safety*. The expectation today is that complex equipment and systems are not only free from defects and systematic failures at time $t = 0$ (when they are put into operation), but also perform the required function *failure free* for a stated time interval. However, the question of whether a given item will operate without failures during a stated period of time cannot be simply answered by *yes* or *no* on the basis of a compliance test. Experience shows that only a *probability* for this occurrence can be given. This probability is a measure of the item's *reliability* and can be interpreted as follows:

> *If n statistically identical items are put into operation at time $t = 0$ to perform a given mission and $\bar{\nu} \le n$ of them accomplish it successfully, then the ratio $\bar{\nu}/n$ is a random variable which converges for increasing n to the true value of the reliability (Appendix A6.11).*

Performance parameters as well as *reliability, maintainability, availability,* and *safety* have to be *built into an item*. To do this, a set of specific engineering and management activities must be performed during *all life-cycle phases* of the item considered. For complex equipment and systems, these activities have to be coordinated by a *quality and reliability assurance program plan*, realized in the context of *Total Quality Management* (TQM), to be time and cost effective.

1.2 Basic Concepts

This section introduces the most important concepts used in reliability engineering and shows their relationship to quality assurance and TQM (see Appendix A1 for a more complete list of definitions).

1.2.1 Reliability

Reliability is a *characteristic* of an item, expressed by the *probability* that the item will perform its *required function* under given *conditions* for a stated *time interval*. It is generally designated by R. From a qualitative point of view, reliability can be defined as the *ability of an item to remain functional*. Reliability specifies thus the *probability that no operational interruptions* will occur during a stated time interval. This does not mean that *redundant* parts may not fail, such parts can fail and be repaired. The concept of reliability thus applies to *nonrepairable* as well as to *repairable* items (see Chapters 2 and 6, respectively). To make sense, a numerical statement of reliability (e.g., $R = 0.9$) must always be accompanied by the definition of the *required function*, the *operating conditions*, and the *mission duration*, say T. In general, it is also important to know whether or not the item can be considered new when the mission starts.

An *item* is a functional or structural *unit* of arbitrary complexity (component, device, assembly, equipment, subsystem, or system) that can be considered as an *entity* for investigations. It may consist of hardware, software, or both and may also include human resources. Often, *ideal* human aspects and logistical support are assumed, even if for simplicity in the notation the term *system* is used instead of *technical system.*

The *required function* specifies the item's task. For example, for given inputs, the item outputs have to be constrained within specified tolerance bands (performance parameters should still be given with tolerances and not merely as fixed values). The definition of the required function is the starting point for *any* reliability analysis, as it defines *failures.*

Operating conditions have an important influence upon reliability, and must therefore be specified with care. Experience shows for example, that the failure rate of semiconductor components will double for an operating temperature increase of 10 to 20°C.

The required function and/or operating conditions can also be *time dependent.* In these cases a *mission profile* has to be defined and all reliability statements will be related to it. A representative mission profile and the corresponding reliability targets should be given in the item *specifications.*

Often the mission duration T is taken as parameter t, the *reliability function* is then defined by $R(t)$. $R(t)$ is the probability that no failure at item level will occur in the interval $(0, t]$, generally with the assumption $R(0) = 1$.

To avoid confusion, a distinction between *predicted* and *estimated* (or *assessed*) reliability should be made. The first one is computed on the basis of the item's reliability structure and the failure rate of its components (Section 2.2), the second is obtained from a statistical evaluation of reliability tests (Section 7.2) or from field data if environmental and operating conditions are known.

1.2.2 Failure

A *failure* occurs when an item stops performing its required function. The failure-free operating time is generally a *random variable*. It is often reasonably long, but it can also be very short, for instance because of a systematic failure or an early failure caused by a transient event at turn-on. A general assumption in investigating failure-free operating times is that at $t = 0$ the item is free of *defects* and *systematic failures*. Besides their *relative frequency*, failures are classified according to the mode, cause, effect, and mechanism:

1. *Mode*: The mode of a failure is the *symptom* (local effect) by which a failure is observed; for example, opens, shorts, or drift for electronic components; brittle rupture, creep, cracking, seizure, or fatigue for mechanical components.
2. *Cause*: The cause of a failure can be *intrinsic*, due to weaknesses in the item and/or wearout, or *extrinsic*, due to misuse or mishandling during the design,

production, or use. Extrinsic causes often lead to *systematic failures* which are *deterministic* and should be considered like *defects* (dynamic defects in software quality). Defects are present at $t = 0$. Failures appear *always in time* (even if the time to failure is short as it can be with systematic failures or early failures).

3. *Effect*: The effect (consequence) of a failure is generally different if considered on the item itself or at a higher level. A usual classification is: *nonrelevant*, *partial*, *complete*, and *critical failure*. Since a failure can also cause further failures in an item, a distinction between *primary failure* and *secondary failure* is important.
4. *Mechanism*: Failure mechanism is the physical, chemical, or other process resulting in a failure (see Table 3.6 for some examples).

Failures are also classified as *sudden* and *gradual*. Sudden and complete failures are termed *cataleptic (or catastrophic) failures*, gradual and partial failures are termed *degradation failures*. As failure is not the only cause for an item being down, the general term used to define the down state of an item which is not caused by a preventive maintenance, other planned actions, or lack of external resources is *fault*. Fault is thus a state of an item and can be due to a *defect* or a *failure*.

1.2.3 Failure Rate

The *failure rate* plays an important role in reliability analyses. This Section introduces it heuristically (see Sections 2.2.5 and A6.5 for an analytical derivation).

Let us assume that n *statistically identical*, independent items are put into operation at time $t = 0$ under the same conditions, and at the time t a subset $\bar{\nu}(t)$ of these items have not yet failed. $\bar{\nu}(t)$ is a right continuous decreasing step function, shown in Fig. 1.1. $t_1, ..., t_n$ are the *observed* failure-free operating times of the n items. As stated above, they are independent realizations of a *random variable* τ, considerd here as the item's *failure-free operating time*. The expression

$$\hat{\mathrm{E}}[\tau] = \frac{t_1 + \ldots + t_n}{n} \tag{1.1}$$

is the *empirical* expected value or *empirical mean* of τ (empirical quantities are statistical estimates and marked with ^). For $n \to \infty$, $\hat{\mathrm{E}}[\tau]$ converges to the true value of the mean failure-free operating time $\mathrm{E}[\tau]$ (Eq. A6.147). The function

$$\hat{\mathrm{R}}(t) = \frac{\bar{\nu}(t)}{n} \tag{1.2}$$

is the *empirical reliability function*. As shown in Appendix A8.1.1, $\hat{\mathrm{R}}(t)$ converges to the reliability function $\mathrm{R}(t)$ for $n \to \infty$.

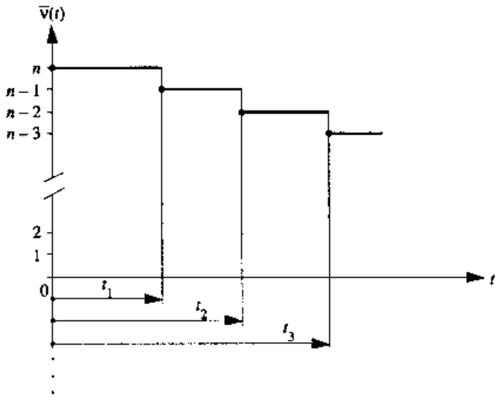


Figure 1.1 Number $\bar{\nu}(t)$ of items not failed at the time t

The *empirical failure rate* is defined as

$$\hat{\lambda}(t) = \frac{\bar{\nu}(t) - \bar{\nu}(t + \delta t)}{\bar{\nu}(t)\delta t}. \tag{1.3}$$

$\hat{\lambda}(t)\delta t$ is the ratio of the items failed in the interval $(t, t + \delta t]$ to the number of items that have *not yet failed at the time t*. Applying Eq. (1.2) to Eq. (1.3) yields

$$\hat{\lambda}(t) = \frac{\hat{\mathrm{R}}(t) - \hat{\mathrm{R}}(t + \delta t)}{\delta t\, \hat{\mathrm{R}}(t)}. \tag{1.4}$$

For $n \to \infty$ and $\delta t \to 0$, in such a way that $n\delta t \to 0$, $\hat{\lambda}(t)$ converges to the failure rate

$$\lambda(t) = \frac{-d\,\mathrm{R}(t)/dt}{\mathrm{R}(t)}. \tag{1.5}$$

Equation (1.5) implies (tacitly) that $\mathrm{R}(t)$ is derivable. It shows that the failure rate $\lambda(t)$ *fully determines* the reliability function $\mathrm{R}(t)$. With $\mathrm{R}(0) = 1$, Eq. (1.5) yields

$$\mathrm{R}(t) = e^{-\int_0^t \lambda(x)\,dx}. \tag{1.6}$$

In many practical applications, the failure rate can be assumed to be nearly constant (time independent) for all $t \ge 0$

$$\lambda(t) = \lambda.$$

From Eq. (1.6) then follows

$$R(t) = e^{-\lambda t}. \tag{1.7}$$

The failure-free operating time τ is in this case *exponentially distributed.* For this case (and *only* in this case), the failure rate λ can be estimated by $\hat{\lambda} = k/T$, where T is a given (fixed) *cumulative operating time* and k the *total number of failures during* T (Section 7.2.2.1). This result is a consequence of the *memoryless property* of the exponential distribution function (Eqs. (2.12), (A6.87), (A7.41)). The mean of the failure-free operating time is given in general by

$$MTTF = E[\tau] = \int_0^\infty R(t)\,dt, \tag{1.8}$$

where *MTTF* stands for *mean time to failure* (Eq. (A6.38)). In the case of a constant failure rate $\lambda(t) = \lambda$, $E[\tau]$ assumes the value $E[\tau] = \int_0^\infty e^{-\lambda t}\,dt = 1/\lambda$. It is common usage to define

$$MTBF = \frac{1}{\lambda}, \tag{1.9}$$

where *MTBF* stands for *mean operating time between failures* (formerly *mean time between failures*). Also because of the frequently used point estimation $\hat{MTBF} = T/k$, with T as the given (fixed) *cumulative operating time* and k as the *total number of failures during T, MTBF* should be reserved for items with a constant (time independent) failure rate λ.

The failure rate of a *large population* of *statistically identical*, independent items exhibits often the typical bathtub curve depicted in Fig. 1.2, with the following three phases:

1. *Early failures*: $\lambda(t)$ decreases rapidly with time; failures in this phase are generally attributable to randomly distributed weaknesses in materials, components, or production processes.
2. *Failures with constant (or nearly so) failure rate*: $\lambda(t)$ is approximately constant and equal to λ; failures in this period are *Poisson distributed* and often cataleptic.
3. *Wearout failures*: $\lambda(t)$ increases with time; failures in this period are generally attributable to aging, wearout, fatigue, etc.

Early failures are often due to momentary (randomly distributed) *weaknesses* in materials, components, or in the item's production process. They are *not deterministic* (in contrast to *systematic failures*) and thus appear in general randomly distributed in time and over the items. During the early failure period, $\lambda(t)$ must not *necessarily decrease* as in Fig. 1.2, in some cases it can also oscillate. To eliminate early failures, *burn-in* or *environmental stress screening* (ESS) is used.

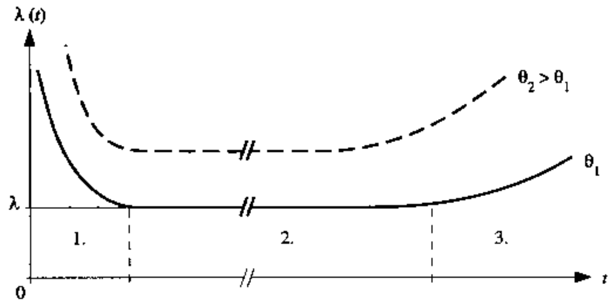


Figure 1.2 Typical shape of the failure rate of a large population of statistically identical (independent) items (the dashed line shows the basic shift of the curve for a higher stress, e.g. ambient temperature θ for electronic components)

Early failures should be distinguished from *systematic failures*, which are deterministic and caused by *errors* or *mistakes*, and whose elimination requires a *change* in the design, production process, operational procedure, documentation, or other. The length of the early failure period varies greatly in practice. However, in most applications it will be shorter than a few thousand hours. The presence of a period with *constant* (or nearly so) *failure rate* $\lambda(t) = \lambda$ is useful for calculations. The *memoryless property*, which characterizes this period, leads to a *Poisson process* for the flow of failures and to a *Markov process* for the time behavior of a repairable item if also *constant repair rates* can be assumed (Sections 2.3.5, 6.4.1, 6.5.1, 6.8). An *increasing failure rate* after a given operating time appears because of degradation phenomena due to aging or wearout, and is typical for most items.

A *plausible explanation* for the shape of $\lambda(t)$, as given in Fig. 1.2, is that the population of n statistically identical, independent items contains $n\,p_f$ weak elements and $n(1-p_f)$ good ones. The distribution of the failure-free operating time can then be expressed by a *weighted sum* of the form $F(t) = p_f F_1(t) + (1-p_f)F_2(t)$. For calculation purposes, $F_1(t)$ can be a gamma distribution with $\beta < 1$ and $F_2(t)$ a shifted Weibull distribution with $\beta > 1$, see Eqs. (A6.34), (A6.96), and (A6.97).

The failure rate introduced above depends strongly upon the item's operating conditions. For semiconductor devices, experience shows for example that the value of λ doubles for an operating temperature increase of 10 to 20°C, and becomes more than an order of magnitude higher if the device is exposed to elevated mechanical stresses (Table 2.3). Typical *figures* for λ are 10^{-10} to $10^{-7}\,h^{-1}$ for components, and 10^{-7} to $10^{-5}\,h^{-1}$ for assemblies.

The concept of failure rate also applies to humans and a shape similar to that depicted in Fig. 1.2 can be obtained from a mortality table.

1.2.4 Maintenance, Maintainability

Maintenance defines the set of activities performed on an item to *retain* it in or to *restore* it to a specified state. Maintenance is thus subdivided into *preventive maintenance* (carried out at predetermined intervals and according to prescribed procedures, in particular to reduce wearout failures) and *corrective maintenance* (carried out after fault recognition, and intended to put the item into a state in which it can again perform the required function). Corrective maintenance is also known as *repair*, it can include any or all of the following steps: localization, isolation, disassembly, exchange, reassembly, alignment, checkout. The aim of preventive maintenance is also to detect and repair *hidden failures*, i.e. failures in redundant elements. To simplify computations it is generally assumed that the item (element in the reliability block diagram for which a maintenance action has been performed) is *as-good-as-new* after each maintenance. This assumption is valid for the wohle equipment or system in the case of *constant failure rates*.

Maintainability is a *characteristic* of an item, expressed by the *probability* that *preventive maintenance* or *repair* of the item will be performed within a stated *time interval* for given *procedures and resources* (number and skill level of personnel, spare parts, test facilities, etc.). From a qualitative point of view, maintainability can be defined as the *ability of an item to be retained in or restored to a specified state*. The expected value (*mean*) of the repair time is denoted by *MTTR* (mean time to repair), and that of a preventive maintenance by *MTTPM* (mean time to preventive maintenance). Often used is for instance also *MTBUR* (mean time between unscheduled removals). Maintainability has to be *built into* complex equipment or systems during the design and development phase, by planning and realizing a *maintenance concept*. Due to the increasing maintenance costs, maintainability aspects have been growing in importance. However, maintainability achieved in field largely depends on the correct installation of the equipment or system, the resources (personnel and material) available for maintenance, and the maintenance organization, i.e. on the *logistical support*.

1.2.5 Logistical Support

Logistical support designates all activities undertaken to provide effective and economical use of an item during its operating phase. To be effective, logistical support should be integrated into the *maintenance concept* of the item under consideration and include after-sales service.

An emerging aspect, related to maintenance and logistical support, is that of *obsolescence management*, i.e. how to assure operation over for instance 20 years, *when technology is rapidly evolving* and components need for maintenance are no longer manufactured. Consideration shall be given in particular to the *design aspects*, to assure interchangeability during the equipment's useful life without important redesigns. Standardization in this direction is in progress [1.7].

1.2.6 Availability

Point availability is a *characteristic* of an item, often designated by PA(t), expressed by the *probability* that the item will perform its *required function* under *given conditions* at a stated *instant of time t*. From a qualitative point of view, *point availability* can be defined as the *ability of an item to perform its required function under given conditions at a stated instant of time*, the term *dependability* is then often used.

Availability calculations are generally difficult, as *logistical support* and *human factors* should be considered in addition to reliability and maintainability. *Ideal* human and logistical support conditions are thus often assumed, leading to the *intrinsic availability*. For simplicity in the notation, in Chapter 6 availability will be used instead of intrinsic availability. Further assumptions for calculations are continuous operation and complete renewal (the repaired element in the rel. block diagram is as-good-as-new after each repair). Taking this into account, the point availability PA(t) converges rapidly to the *steady-state value* (Eqs. (6.48), (6.57))

$$PA = \frac{MTTF}{MTTF + MTTR}. \tag{1.10}$$

PA is also the steady-state value of the *average availability* (*AA*, or simply *A* for *availability*) and expresses the expected value (*mean*) of the *percentage of the time* during which the item performs its required function. Depending on the application considered, other types of availability can be defined (Section 6.2). In contrast to reliability analyses for which no failure at the item level is allowed (only redundant parts may fail and be repaired), availability analysis allows failures at the item level.

1.2.7 Safety, Risk, and Risk Acceptance

Safety is the ability of an item not to cause injury to persons, nor significant material damage or other unacceptable consequences during its use. Safety evaluation must consider the following two aspects: safety when the item functions and is operated correctly and safety when the item or a part of it has failed. The first aspect deals with *accident prevention*, for which a large number of national and international regulations exist. The second aspect is that of *technical safety* which is investigated using the same tools as for reliability. However, a distinction between technical safety and reliability is necessary. While safety assurance examines measures which allow an item to be brought into a *safe state* in the case of a failure (*fail-safe behavior*), reliability assurance deals more generally with measures for minimizing the total number of failures. Moreover, technical safety must also take into account of the effects of *external influences* like human errors, catastrophes, sabotage, etc. The safety level of an item has great influence on the number of *product liability* claims. However, measures designed to increase safety can reduce reliability. To some extent, the concept of safety can be applied to arbitrary products or services.

Closely related to the concept of (technical) safety are those of *risk, risk management,* and *risk acceptance,* including risk analysis and risk assessment [1.20, 1.24, 2.35]. Risk problems are generally *interdisciplinary* and have to be solved in close *cooperation* between engineers, psychologists, and politicians to find common solutions to controversial questions. An appropriate weighting between *probability of occurrence* and *effect* (consequence) of a given accident is important. The multiplicative rule is one of several possibilities. Also it is necessary to consider the different *causes* (machine, machine & human, human) and *effects* (location, time, involved people, effect duration) of an accident. Statistical tools can support *risk assessment.* However, although the behavior of a homogenous human population is often known, experience shows that the reaction of a *single person* can become unpredictable. Similar difficulties also arise in the evaluation of *rare events* in complex systems. Considerations on risk and risk acceptance should take into account that the probability p_1 for a given accident which can be caused by one of n statistically identical and *independent* items, each of them with occurrence probability p, is for np small nearly equal to np as per

$$p_1 = np(1-p)^{n-1} \approx npe^{-np} \approx np(1-np) \approx np. \tag{1.11}$$

Equation (1.11) follows from the binomial distribution and the Poisson approximation (Eqs. (A6.120) and (A6.129)). It also applies with $np = \lambda_{tot} T$ to the case in which one assumes that the accident occurs randomly in the interval $(0, T]$, caused by one of n independent items (systems) with failure rates $\lambda_1, \ldots, \lambda_n$, where $\lambda_{tot} = \lambda_1 + \ldots + \lambda_n$ or $\lambda_{tot} = n\lambda$ for $\lambda_1 = \ldots = \lambda_n = \lambda$. This is because the *sum of n independent Poisson processes is again a Poisson process* (Example 7.7) and the probability $\lambda_{tot} T e^{-\lambda_{tot} T}$ for one failure in the interval $(0, T]$ is nearly equal to $\lambda_{tot} T$. Thus, for $np << 1$ or $\lambda_{tot} T << 1$ it holds that

$$p_1 \approx np \approx (\lambda_1 + \ldots + \lambda_n) T. \tag{1.12}$$

Also by assuming a reduction of the individual occurrence probability p (or failure rate λ_i), one recognizes that in the future it will be necessary either to *accept greater risks* p_1 or to keep the spread of *high-risk technologies under tighter control.* Similar considerations could also be made for the problem of *environmental stresses* caused by mankind. Aspects of *ecologically acceptable* production, use, disposal, and *recycling* or *reuse* of products will become a subject for international regulations in the general context of a *sustainable development.*

In the context of a *product development process,* risks related to feasibility and time to market within the given cost constraints must also be considered.

Mandatory for risk management are in particular psychological aspects related to *risk awareness* and *safety communication.* As long as a *danger for risk* is not perceived, people often do not react (the coconut effect is a good example for such a situation [1.24 (1997)]). Knowing that a *safety behaviour* presupposes a risk awareness, *communication* is an important tool to avoid that a risk related to the system considered will be underestimated. Psycholinguistic principles can help to facilitate communication between all people involved [1.24 (1993)].

1.2.8 Quality

Quality is understood today as the totality of features and characteristics of an item (product or service) that bear on its ability to satisfy stated or implied needs. This definition is general and has the advantage of accounting for all objective and subjective attributes and characteristics of an item. A disadvantage for technical products can be a reduced strength compared to earlier definitions.

1.2.9 Cost and System Effectiveness

All previously introduced concepts are interrelated. Their relationship is best shown through the concept of cost effectiveness, as given in Fig. 1.3. *Cost effectiveness* is a measure of the ability of an item to meet a service demand of stated quantitative characteristics, with the best possible usefulness to life-cycle cost ratio. It is often referred to as *system effectiveness.* From Fig. 1.3, one recognizes the central role of *quality assurance,* bringing together *all* assurance activities (Section 1.5) and of *dependability* (collective term for availability performance and its influencing factors). *Quality assurance* is used here in the sense also of *quality management* as per *TQM.*

As shown in Fig. 1.3, the *life-cycle cost* (LCC) is the sum of the costs for acquisition, operation, maintenance, and disposal of an item. For complex systems, higher reliability in general leads to a higher acquisition cost and lower operating cost, so that the optimum of life-cycle cost seldom lies at extremely low or extremely high reliability figures. For such a system, with say 10 years useful life, the acquisition cost often accounts for 40 to 60% of the life-cycle cost and experience shows that up to 80% of the life-cycle cost is frequently generated by decisions early in the design phase. In the future, life-cycle cost will take more into account current and deferred damage to the *environment* caused by the production, use, and disposal of an item. Life-cycle cost *optimization* is generally project specific and falls within the framework of *cost effectiveness* or of *systems engineering.* It can be positively influenced by *concurrent engineering* [1.10, 1.14, 1.21]. Figure 1.4 shows as an example the influence of the attainment level of quality and reliability targets on the sum of costs for quality assurance and for the assurance of reliability, maintainability, and logistical support for two complex systems [2.2 (1986)]. To introduce this model let us first consider Example 1.1.

Example 1.1

An assembly contains n independent components each with a *defective probability* p. Let c_k be the cost to replace k defective components. Determine (i) the expected value (mean) $C_{(i)}$ of the total replacement cost (no defective components are allowed in the assembly) and (ii) the mean of the total cost (test and replacement) if the components are submitted to an incoming inspection which reduces defective percentage from p to p_0 (test cost c_t per component).

Solution

(i) The solution makes use of the *binomial distribution* (Appendix A6.10.7) and question (i) is also solved in Example A6.18. The probability of having exactly k defective components in a lot of size n is given by

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}. \qquad (1.13)$$

The mean $C_{(i)}$ of the total cost (deferred cost) caused by the defective components follows then from

$$C_{(i)} = \sum_{k=1}^{n} c_k p_k = \sum_{k=1}^{n} c_k \binom{n}{k} p^k (1-p)^{n-k}. \qquad (1.14)$$

(ii) To the cost caused by the defective components, computed from Eq. (1.14) with p_0 instead of p, one must add the incoming inspection cost $n\,c_t$

$$C_{(ii)} = n c_t + \sum_{k=1}^{n} c_k \binom{n}{k} p_0^k (1-p_0)^{n-k}. \qquad (1.15)$$

The difference between $C_{(i)}$ and $C_{(ii)}$ gives the gain obtained by introducing the incoming inspection, allowing thus a *cost optimization*.

With similar considerations to those in Example 1.1 one obtains for the expected value (*mean*) of the total repair cost C_{cm} during the cumulative operating time T of an item with failure rate λ and cost c_{cm} per repair

$$C_{cm} = \lambda\, T\, c_{cm} = \frac{T}{MTBF}\, c_{cm}. \qquad (1.16)$$

In Eq. (1.16), the term $\lambda\, T$ gives the mean value of the number of failures during T (Eq. (A7.40)).

From the above considerations, the following equation expressing the *mean* C of the sum of the costs for quality assurance and for the assurance of reliability, maintainability, and logistical support of a system can be obtained

$$C = C_q + C_r + C_{cm} + C_{pm} + C_l + \frac{T}{MTBF_S}\, c_{cm} + (1 - OA_S)\, T\, c_{off} + n_d\, c_d, \qquad (1.17)$$

where q denotes quality, r reliability, cm corrective maintenance, pm preventive maintenance, l logistical support, off downtime, and d defects. $MTBF_S$ and OA_S

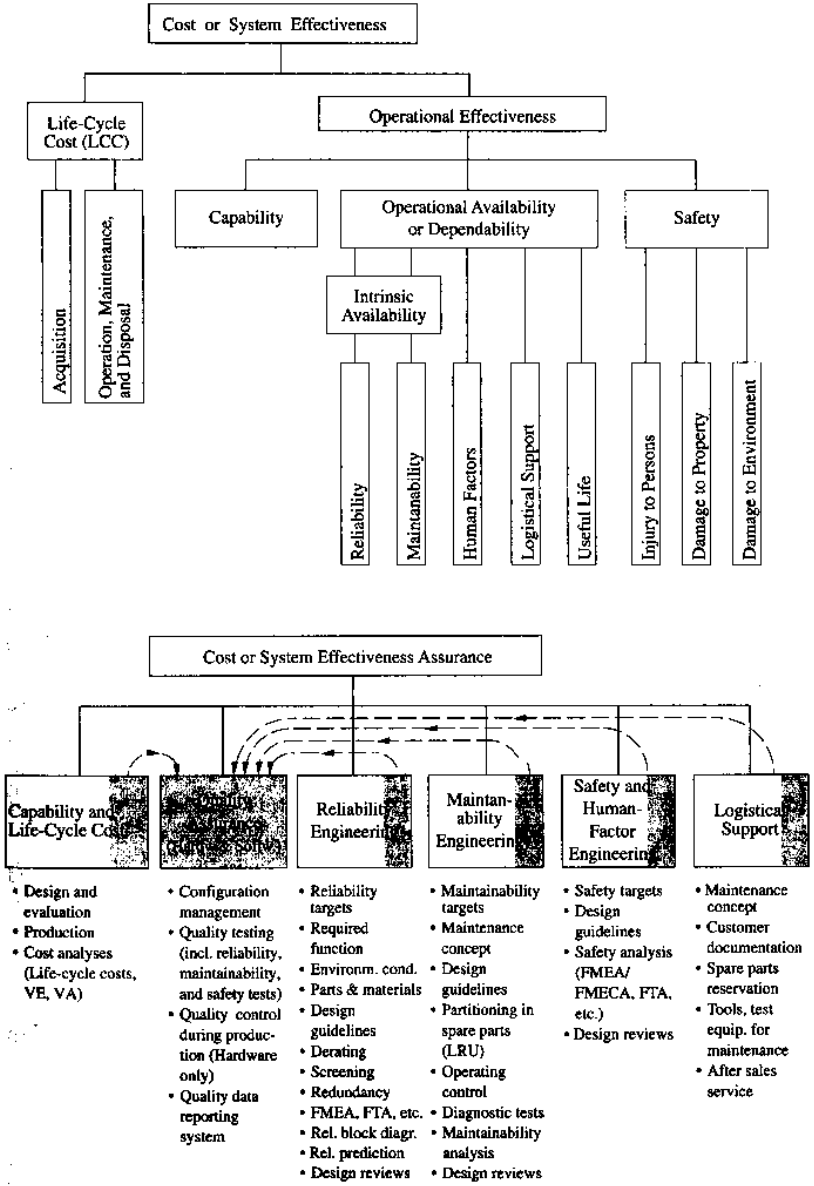


Figure 1.3 Cost or system effectiveness for *complex equipment and systems with high quality and reliability requirements* (dependability can be used instead of operational availability for a qualitative meaning, Quality Assurance in the sense also of Quality Management as per TQM)

are the system mean time between failures and the system steady-state *overall availability* (Eq. (6.201) with T_{pm} instead of T_{PM}), T is the total system operating time (useful life), n_d is the number of *hidden defects* discovered (and eliminated) in the field. C_q, C_r, C_{cm}, C_{pm}, and C_l are the costs for quality assurance and for the assurance of reliability, repairability, serviceability, and logistical support, respectively. c_{cm}, c_{off}, and c_d are the cost per repair, downtime hours (caused by failures), and hidden defect discovered in the field (preventive maintenance costs are included in c_{off}, as preventive maintenance is taken into account in OA_S). The first five terms in Eq. (1.17) represent a part of the *acquisition cost*, the last three terms are *deferred costs* occurring during field operation. A model for investigating the cost C according to Eq. (1.17) was developed in [2.2 (1986)] by assuming C_q, C_r, C_{cm}, C_{pm}, C_l $MTBF_S$, OA_S, T, c_{cm}, c_{off}, and c_d as parameters and investigating the variation of the total cost expressed by Eq. (1.17) as a function of the level of attainment of the specified targets, i.e. by introducing the variables $g_q = QA / QA_g$, $g_r = MTBF_S / MTBF_{Sg}$, $g_{cm} = MTTR_{Sg} / MTTR_S$, $g_{pm} = MTTPM_{Sg} / MTTPM_S$, and $g_l = MLD_{Sg} / MLD_S$, where the subscript g denotes the specified target for the corresponding quantity. A power relationship

$$C_i = C_{ig}\, g_i^{m_i} \qquad (1.18)$$

was assumed between the actual cost C_i the cost C_{ig} to reach the specified target (goal) of the considered quantity, and the *level of attainment* g_i of the specified target ($0 < m_l < 1$ and all other $m_i > 1$). The following relationship between the number of hidden defects discovered in the field and the ratio C_q / C_{qg} was also included in the model

$$n_d = \frac{1}{(C_q / C_{qg})^{m_d}} - 1 = \frac{1}{g_q^{m_q m_d}} - 1. \qquad (1.19)$$

The final equation for the cost C as function of the variables g_q, g_r, g_{cm}, g_{pm}, and g_l is then given by

$$C = C_{qg} g_q^{m_q} + C_{rg} g_r^{m_r} + C_{cmg} g_{cm}^{m_{cm}} + C_{pmg} g_{pm}^{m_{pm}} + C_{lg} g_l^{m_l} + \frac{T c_{cm}}{g_r MTBF_{Sg}}$$
$$+ \Big(1 - \frac{1}{1 + \frac{1}{g_r g_{cm}} \cdot \frac{MTTR_{Sg}}{MTBF_{Sg}} + \frac{1}{g_r g_l} \cdot \frac{MLD_{Sg}}{MTBF_{Sg}} + \frac{MTTPM_{Sg}}{g_{pm} T_{pm}}}\Big) T c_{off} + \Big(\frac{1}{g_q^{m_q m_d}} - 1\Big) c_d . \qquad (1.20)$$

The relative cost C / C_g given in Fig. 1.4 is obtained by dividing C by the value C_g obtained form Eq. (1.20) with all $g_i = 1$. Extensive analyses with different values for m_i, C_i, $MTBF_S$, OA_S, T, c_{cm}, c_{off}, and c_d have shown that the value C / C_g is only moderately sensitive to the parameters m_i.

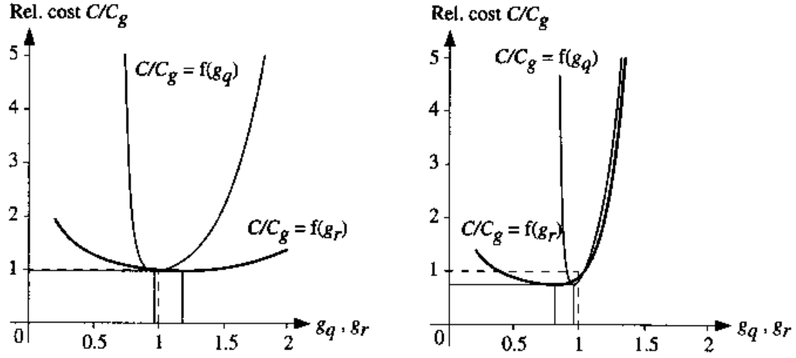


Figure 1.4 Sum of the relative cost C / Cg for quality assurance and for the assurance of reliability, maintainability, and logistical support of two complex systems with different mission profiles, as a function of the level of attainment of the specified quality and reliability targets g_q and g_r, respectively (the specified targets are dashed)

1.2.10 Product Liability

Product liability is the onus on a manufacturer (producer) or others to compensate for losses related to injury to persons, material damage, or other unacceptable consequences caused by a product (item). The manufacturer has to specify a safe operational mode for the product. The term *product* is used here to indicate in general *hardware* only. Furthermore, in legal documents related to product liability, the term *defective product* is used instead of *defective or failed product*. Responsible in a product liability claim are all those people involved in the design, production, and sale of the product, inclusive suppliers. Basically, *strict liability* is applied (the manufacturer has to demonstrate that the product was free from defects). This holds in particular in the USA but also partially in Europe [1.11, 1.13, 1.19]. However, in Europe the causality between damage and defect has still to be demonstrated by the user.

The rapid increase of product liability claims (alone in the USA, 50,000 in 1960 and over one million in 1980) cannot be ignored by manufacturers. Although the reason for such a situation also lies in the peculiarity of US legal procedures, *configuration management* (Section A3.3.5) and *safety analyses* (Section 2.6) should be performed to avoid product liability claims. This is true even if the growth of liability claims in Europe will remain under better control than in the USA.

1.2.11 Historical Development

Methods and procedures of quality assurance and reliability engineering have been developed extensively over the last 50 years. Table 1.1 summarizes the major steps. Figure 1.5 depicts the *basic* distribution of the relative effort between quality assurance and reliability engineering during this period of time. Because of the rapid progress of microelectronics, considerations on *redundancy, fault-tolerance, testability, test strategies*, and *software quality* have increased in importance.

Table 1.1 Historical development of quality assurance and reliability engineering

before 1940	Quality attributes and characteristics are defined. In-process and final tests are carried out, usually in a department within the production area. The concept of *quality of manufacture* is introduced.
1940 - 50	Defects and failures are systematically collected and analyzed. *Corrective actions* are carried out. *Statistical quality control* is developed. It is recognized that quality must be *built into* an item. The concept *quality of design* becomes important.
1950 - 60	*Quality assurance* is recognized as a means for developing and manufacturing an item with a specified quality level. *Preventive measures* are added to tests and corrective actions. It is recognized that correct short-term functioning does not also signify *reliability*. *Design reviews* and systematic analysis of failures and failure data, generally in the research and development area, lead to important reliability improvements.
1960 - 70	Difficulties with respect to reproducibility and change control, as well as interfacing problems during the integration phase, require a refinement of the concept of *configuration management*. Reliability engineering is recognized as a means of developing and manufacturing an item with specified reliability. *Reliability estimation* and *demonstration* tests are developed. It is recognized that reliability cannot easily be demonstrated by an *acceptance test*. Instead of a reliability figure (*MTBF*), the contractual requirement is for a *reliability assurance program*. *Maintainability, availability*, and *logistical support* become important.
1970 - 80	Due to the increasing complexity and costs for maintenance of electronic equipment and systems, the aspects of *man-machine interface* and *life-cycle cost* become important. The terms *product assurance, cost effectiveness* and *systems engineering* are introduced. *Product liability* becomes important. Quality and reliability assurance activities are made *project specific* and carried out in close *cooperation* with all engineers involved in a project. Customers require demonstration of reliability and maintainability during the warranty period.
1980 - 90	The aspect of *testability* gains in significance. *Test and screening strategies* are developed to reduce testing costs and warranty services. Because of the rapid progress in microelectronics, greater possibilities are available for *redundant structures*. The concept of *software quality* is introduced.
after 1990	The necessity to further shorten the development time leads to the concept of *concurrent engineering*. *Total Quality Management* (TQM) appears as a refinement to the concept of quality assurance as given at the end of the seventies.

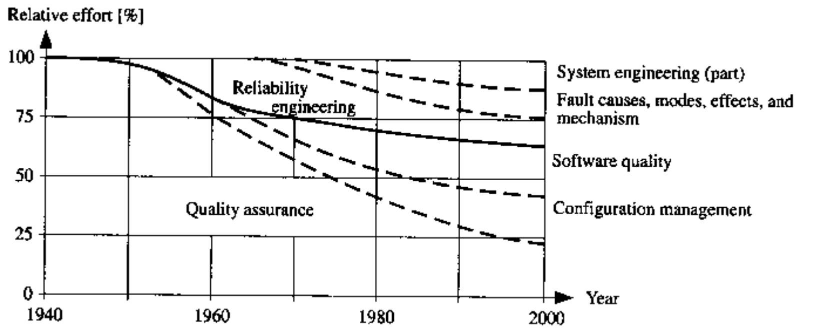


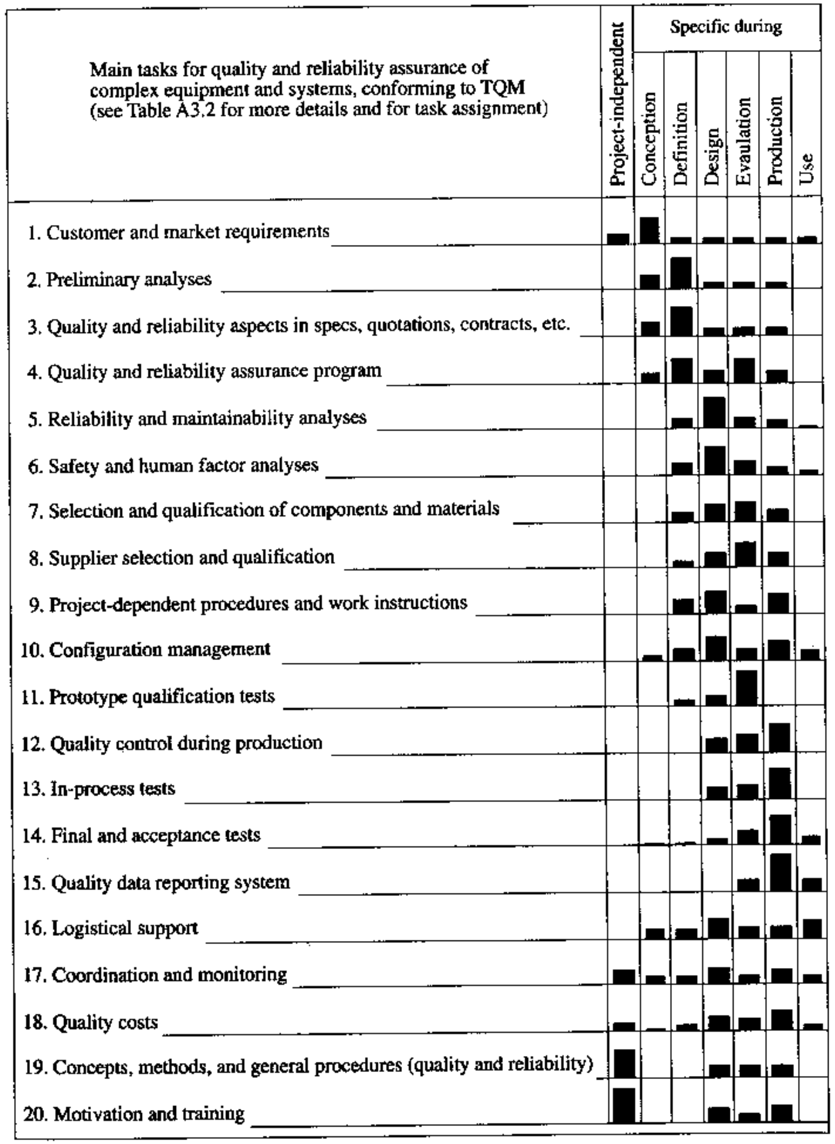
Figure 1.5 Approximate distribution of the relative effort between quality assurance and reliability engineering for complex equipment and systems

1.3 Quality and Reliability Assurance Tasks for Complex Equipment and Systems

Experience shows that the development and production of complex equipment and systems with specified reliability, maintainability, availability, and safety targets, as well as a prescribed quality level, is only possible when *specific activities* are performed during *all life-cycle phases* of the item considered. Figure 1.6 defines the *life-cycle phases* for complex equipment and systems, also indicating the *inputs* and *outputs* of each phase (the output of a given phase corresponds with the input of the next phase, *disposal and recycling* mean ecologically acceptable or environmentally compatible disposal and recycling).

The *main tasks* for quality and reliability assurance of complex equipment and systems are listed in Table 1.2, see also Table A3.2 for further details and a *task assignment matrix*. Depicted in Table 1.2 is the period of time during the life-cycle phases over which the tasks have to be performed. Many tasks extend over *several* project phases and must therefore be *coordinated*. A reinforcement of this coordination in the sense of simultaneous activities leads to the concept of *concurrent engineering*. Within a project, the tasks of Table 1.2 must be refined in a project-specific quality and reliability *assurance program*. The content of such a program is discussed in Appendix A3, starting with Table A3.2 which presents a *detailed description* on the tasks outlined in Table 1.2.

Table 1.2 Main tasks for quality and reliability assurance of complex equipment and systems, conforming to TQM (the bar height is a measure of the relative effort)



Main tasks for quality and reliability assurance of complex equipment and systems, conforming to TQM (see Table A3.2 for more details and for task assignment)	Project-independent	Specific during					
		Conception	Definition	Design	Evaulation	Production	Use
1. Customer and market requirements	■	■	■	■	■	■	■
2. Preliminary analyses		■	■	■	■	■	
3. Quality and reliability aspects in specs, quotations, contracts, etc.		■	■	■	■	■	
4. Quality and reliability assurance program		■	■	■	■	■	
5. Reliability and maintainability analyses			■	■	■	■	■
6. Safety and human factor analyses			■	■	■	■	■
7. Selection and qualification of components and materials			■	■	■	■	
8. Supplier selection and qualification			■	■	■	■	
9. Project-dependent procedures and work instructions			■	■	■	■	
10. Configuration management		■	■	■	■	■	■
11. Prototype qualification tests			■	■	■		
12. Quality control during production				■	■	■	
13. In-process tests				■	■	■	
14. Final and acceptance tests		■	■	■	■	■	■
15. Quality data reporting system					■	■	■
16. Logistical support		■	■	■	■	■	■
17. Coordination and monitoring	■	■	■	■	■	■	■
18. Quality costs	■	■	■	■	■	■	■
19. Concepts, methods, and general procedures (quality and reliability)	■			■	■	■	
20. Motivation and training	■			■	■	■	

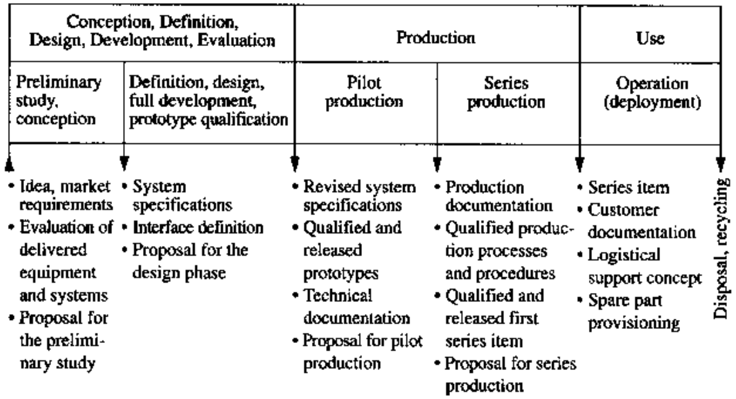


Figure 1.6 Life-cycle phases of complex equipment and systems

1.4 Basic Quality and Reliability Assurance Rules

Based on the considerations in Section 1.3 (refined in Appendix A3), the *basic rules* for a quality and reliability assurance optimized with respect to cost and time schedule, and conforming to Total Quality Management (TQM), can be summarized as follows:

1. Quality and reliability targets should be just as high as necessary to satisfy real customer needs

   → *apply the rule "as-good-as-necessary".*

2. Activities for quality and reliability assurance should be performed continuously throughout *all project phases*, from definition to operating phase

   → *do not change the project manager before ending the pilot production.*

3. Activities must be performed in close cooperation between all engineers involved in the project (Table A3.2)

   → *use TQM and concurrent engineering approaches.*

4. Quality and reliability assurance activities should be monitored by a central quality and reliability assurance department (Q&RA), which cooperates actively in all project phases (Fig. 1.7 and Table A3.2)

   → *establish a quality and reliability assurance department active in the projects.*

5. To ensure the necessary independence, the quality and reliability assurance department should report directly to the upper management level (Fig. 1.7)

   → *assign to the central Q&RA department the competence it needs.*

Figure 1.7 shows an organization which embodies the above rules. As set out in Table A3.2, the assignment of quality and reliability assurance tasks should be such that every engineer in a project *bears his/her own responsibilities.* A design engineer should for instance be responsible for all aspects of his own product (e.g. an assembly) including reliability, and the production department should be able to manufacture and test such an item within its own competence. Following the suggestions given in Table A3.2 for a manufacturer of complex equipment and systems with high quality and reliability requirements, the central quality and reliability assurance department (Q&RA in Fig. 1.7) should be responsible for

- setting targets for reliability and quality levels,
- coordination of all activities belonging to quality and reliability assurance,
- preparation of guidelines and working documents,
- qualification, testing and screening of components and material,
- release of manufacturing processes (quality and reliability aspects),
- development and operation of the quality data reporting system,
- solution of quality and reliability problems at the equipment and system level,
- final and acceptance testing.

This central quality and reliability department should not be too small (credibility) nor should it be too large (sluggishness). A reasonable size is between 2 and 4% of the total manpower of a company, depending upon company size, product area, and assignment of tasks.

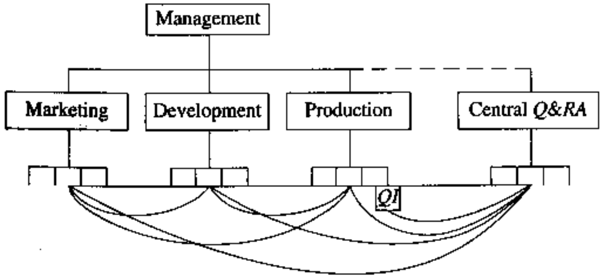


Figure 1.7 Basic organizational structure for quality and reliability assurance in a company producing complex equipment and systems with high quality (Q) and reliability (R) requirements (connecting lines indicate close cooperation; *A* denotes assurance/management, *I* inspection)

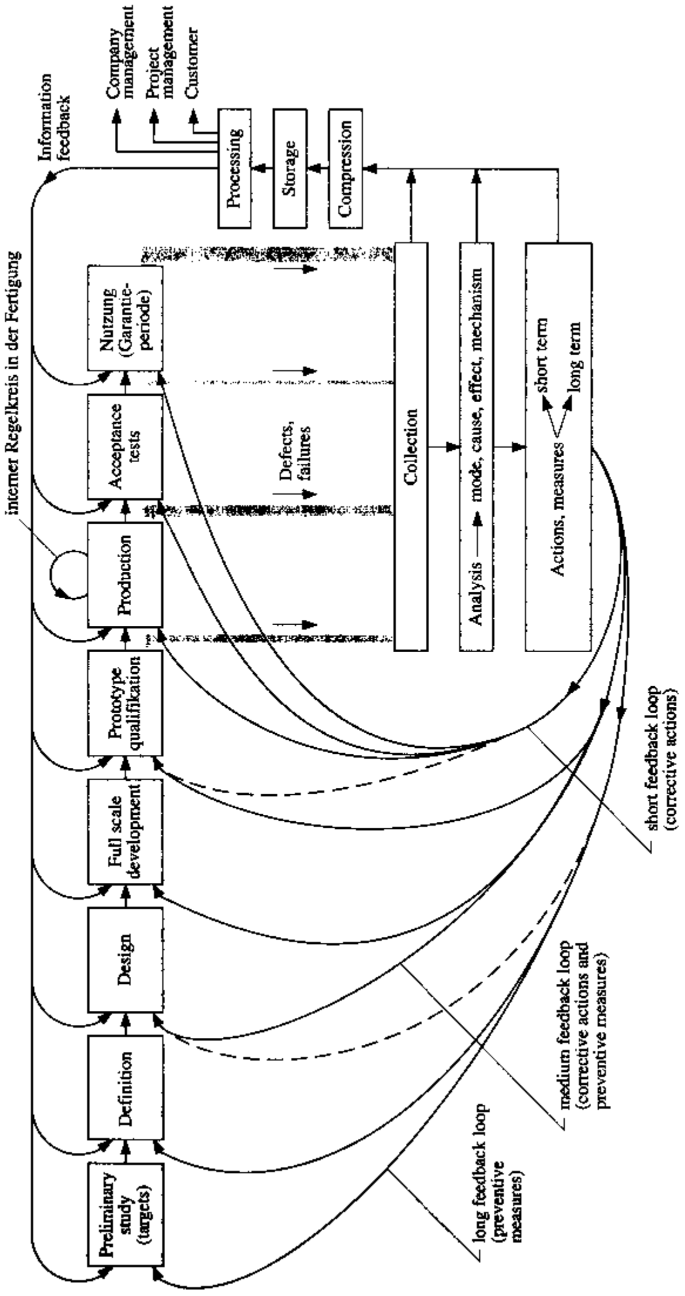


Figure 1.8 Principle of a quality data reporting system

Table 1.3 Information status for PCBs from a quality data reporting system

a) Defects and failures at PCB level

Period: . . . .

PCB	No. of PCBs			Rough classification				No. of faults		Measures		Cost		
	tested	with faults	%	assembling	soldering	board	component	total	per PCB	short term	long term	production	Q A	other areas

b) Defects and failures at component level

Period: . . . .   PCB: . . . .   No. of PCBs: . . . .

Component	Manufacturer	No. of components		Number of faults	%	No. of faults per place of occurrence			
		Same type	Same application			incoming inspection	in-process test	final test	warranty period

c) Cause analysis for defects and failures due to components

Period: . . . .

Component	PCB	Cause			Percent defective (%)		Failure rate ($10^{-9}h^{-1}$)		Measures	
		systematic	inherent failure	not identified	observed	predicted	observed	predicted	short term	long term

d) Correlation between components and PCBs

Period: . . . .

PCB / Component					

1.5 Elements of a Total Quality Management System

As stated in Sections 1.3 and 1.4, many of the tasks associated with quality assurance (in the sense also of quality management as per TQM) are *interdisciplinary* in nature. In order to have a minimum impact on cost and time schedules, their solution requires the *concurrent efforts* of all engineers involved in a project, in particular those from development and production. The related activities, distributed along the project phases and among different company departments (marketing, development, production), must be integrated. This leads to the concept of *quality assurance*, as depicted in Fig. 1.3.

In order to improve the coordination of the quality assurance activities it is useful to group these activities into the following areas:

1. *Configuration Management*: Procedure used to specify, describe, audit, and release the configuration of an item, as well as to control it during modifications or changes (configuration includes all of an item's functional and physical characteristics as given in the documentation and present in the hardware and/or software). Configuration management is one important tool for quality assurance during the development phase. It is subdivided into configuration *identification*, *auditing*, *control*, and *accounting* (Appendix A3.2.5). Auditing of a configuration is carried out through *design reviews*.
2. *Quality Tests*: Tests to verify whether an item conforms to specified requirements. Quality tests include incoming inspections, as well as qualification tests, production tests, and acceptance tests. They also cover reliability, maintainability, and safety aspects. To be cost effective, quality tests must be coordinated and integrated into a *test and screening strategy*.
3. *Quality Control During Production*: Control/monitoring of the production processes and procedures in order to reach a stated quality of manufacturing.
4. *Quality Data Reporting System*: A system to collect, analyze, and correct all defects and failures (faults) occurring during the production and test of an item, as well as to evaluate and feedback the corresponding quality and reliability data. Such a system is generally computer assisted. Analysis of failures and defects must be traced to the *cause* in order to determine the best corrective actions necessary to *avoid repetition* of the same problem. Ideally, a quality data reporting system should remain active during the operating phase (Fig. 1.8 and Tables 1.2 and 1.3).

The configuration management of an item extends from the definition phase until the installation in the field, with emphasis on *technical documentation*, *design reviews*, *production documentation*, and *customer documentation* (Appendices A3 and A4). Quality tests are emphasized in the *prototype qualification* and *production tests* (Chapters 3, 7, and 8). The concept of a quality data reporting system is shown in Fig. 1.8 (see Appendix A5 for basic requirements). To be effective, the information supplied by a quality data reporting system must be *correct/true*, presented *on time*, and in a *concise form*. Table 1.3 shows (as an example) data reporting sheets for populated printed circuit board's (PCBs) evaluation.

1.6 Quality and Reliability Assurance Handbook

The quality and reliability assurance/management system must be described in an appropriate *handbook* supported by the company management. The handbook should describe the *present* system and indicate *how* tasks are performed. Such a handbook is required in all *quality assurance standards* (Appendix A2). To fulfill its aims it should

- clearly define *responsibilities* and *competence,*
- serve as a *reference work,*
- be dynamic,
- be issued as a result of *teamwork* between all affected areas in the company,
- increase customer confidence.

It is useful to subdivide the handbook into three parts: *General Guidelines and Procedures* (can be given to the customer), *Detailed Guidelines and Procedures* (should be accessible to the customer for informative purposes only), and *Motivation and Training*.

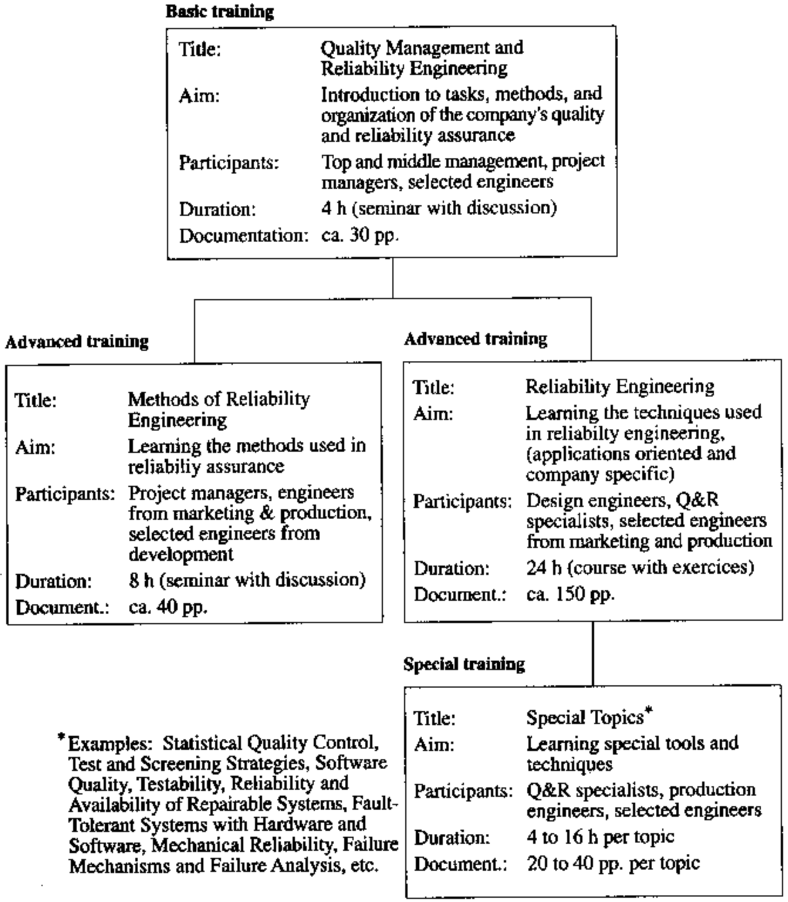
The quality and reliability assurance/management system with its associated handbook must be *tailored* to the complexity and requirements of the item being developed or produced. The following is an example for the table of contents of the first part of a quality and reliability assurance handbook for a company producing *complex equipment and systems with high reliability, maintainability, availability, and safety requirements*:

1. General
    - Quality policy
    - Company organization
    - Organizational responsibility diagram
2. Project Organization
3. Quality and Reliability Assurance Program
    - Preparation and release
    - Development phase
    - Production phase

4. Reliability Engineering
    - Design guidelines
    - Selection and qualification of components and materials
    - Failure rate handbook
    - Reliability models, reliability calculations
    - Fault modes, effects, and criticality analysis
    - Reliability testing
    - Screening strategies
    - Failure analysis, corrective actions
5. Maintainability Engineering
    - Design guidelines
    - Maintenance concept
    - Maintainability tests
    - Logistical support
6. Safety Engineering
    - Accident prevention
    - Design guidelines
    - Safety analysis
    - Safety tests
7. Quality Assurance/Management
    - Configuration management
    - Quality tests
    - Quality control during production
    - Quality data reporting system
    - Software quality assurance

1.7 Motivation and Training

Cost effective quality and reliability assurance can only be achieved if every engineer involved in a project is responsible for his/her assigned activities, conforming to TQM and depicted in Table A3.2. To do this, an appropriate, practice oriented, *motivation and training* program can be necessary. Figure 1.9 shows a motivation and training concept for quality and reliability assurance in a company dealing with complex equipment and systems with high quality and reliability requirements. The courses are aimed at company management, project management, and engineering level.



Basic training

Title:	Quality Management and Reliability Engineering
Aim:	Introduction to tasks, methods, and organization of the company's quality and reliability assurance
Participants:	Top and middle management, project managers, selected engineers
Duration:	4 h (seminar with discussion)
Documentation:	ca. 30 pp.

Advanced training

Title:	Methods of Reliability Engineering
Aim:	Learning the methods used in reliabiliy assurance
Participants:	Project managers, engineers from marketing & production, selected engineers from development
Duration:	8 h (seminar with discussion)
Document.:	ca. 40 pp.

Advanced training

Title:	Reliability Engineering
Aim:	Learning the techniques used in reliabilty engineering, (applications oriented and company specific)
Participants:	Design engineers, Q&R specialists, selected engineers from marketing and production
Duration:	24 h (course with exercices)
Document.:	ca. 150 pp.

Special training

Title:	Special Topics*
Aim:	Learning special tools and techniques
Participants:	Q&R specialists, production engineers, selected engineers
Duration:	4 to 16 h per topic
Document.:	20 to 40 pp. per topic

*Examples: Statistical Quality Control, Test and Screening Strategies, Software Quality, Testability, Reliability and Availability of Repairable Systems, Fault-Tolerant Systems with Hardware and Software, Mechanical Reliability, Failure Mechanisms and Failure Analysis, etc.

Figure 1.9 Example for a training and motivation concept in a company producing complex equipment and systems with high quality (Q) and reliability (R) requirements

2 Reliability Analysis During the Design Phase

Reliability analysis during the design and development phase is important to detect and eliminate *reliability weaknesses* as early as possible and to *perform comparative* studies with respect to reliability. Such an analysis includes *failure rate* and *failure mode* investigations, verification that *design guidelines for reliability* have been considered, and cooperation in *design reviews*. This chapter presents methods and tools for the *failure rate* and *failure mode* analysis of complex electronic and electromechanical equipment and systems. Design guidelines for reliability, maintainability, and software quality are given in Chapter 5. Design reviews are discussed in Appendices A3 and A4. Reliability tests are considered in Chapters 3, 7, and 8. After a short introduction (Section 2.1), Section 2.2 deals with series/parallel structures. Complex structures, elements with more than one failure mode, and parallel models with load sharing (e. g. for standby redundancy) are considered in Section 2.3. Reliability allocation is introduced in Section 2.4. Stress/strength and drift analyses are discussed in Section 2.5. Section 2.6 deals with failure mode analyses and Section 2.7 gives a list of questions for reliability aspects in design reviews. Computer aided analysis is considered in Section 6.8.2. Theoretical foundations for this chapter are given in Appendix A6.

2.1 Introduction

An important part of the reliability analysis during the design and development of complex equipment and systems deals with failure rate and failure mode investigations as well as with the verification of the adherence to appropriate design guidelines for reliability. *Failure mode analyses* are discussed in Section 2.6, and *design guidelines* are given in Sections 5.1 to 5.3 for reliability, maintainability, and software quality. Sections 2.2 to 2.5 are devoted to *failure rate analyses*.

Investigating the failure rate of a complex item (equipment or system) leads to the calculation of the *predicted reliability*, i.e. that reliability which can be computed from the structure of the item under consideration and the reliability of its elements. Such a prediction is necessary for an *early detection of reliability weaknesses*, for *comparative studies*, for the investigation of the *reliability* and *availability* (taking care of *maintainability* and *logistical support*), as well as for the definition of *quantitative reliability targets* for subcontractors. However, because of different kind of uncertainties, the predicted reliability can often be only given with a limited precision. To these uncertainties belong

- simplifications in the mathematical modeling (independent elements, complete and sudden failures, no mistakes during design and manufacturing, no damages),
- insufficient consideration of faults caused by internal or external interference (switching, transients, EMC, etc.),
- inaccuracies in the data used for the computation of the component failure rates.

On the other hand, the *true reliability* of an item can only be determined by *reliability tests*, often performed at the qualification of the prototypes. Experience also shows that with an experienced reliability engineer, the predicted failure rate at equipment or system level often agree *reasonably well* (within a factor 2) with field data. Moreover, relative values obtained by comparative studies generally have a much greater accuracy than absolute values. All these reasons support the efforts for a *reliability prediction* during the design of equipment and systems with specified reliability targets.

Besides theoretical considerations, discussed in the following sections, *practical aspects* have to be considered when designing reliable equipment or systems, for instance with respect to operating conditions and to the mutual influence between elements (input/output, load sharing, effects of failures, transients, etc.). Concrete possibilities for reliability improvement are

- reduction of thermal, electrical, and mechanical stresses,
- correct interfacing of components and materials,
- simplification of design and construction,
- use of qualitatively better components and materials,
- protection against ESD and EMC,
- screening of critical components and assemblies,
- use of redundancy,

in that order. *Design guidelines* (Chapter 5) and *design reviews* (Appendices A3.3 and A4, Tables 2.8 and 4.3) are mandatory to support such improvements. This chapter deals with nonrepairable (up to system failure) equipment and systems. Reliability and availability of repairable equipment and systems is considered in depth in Chaper 6.

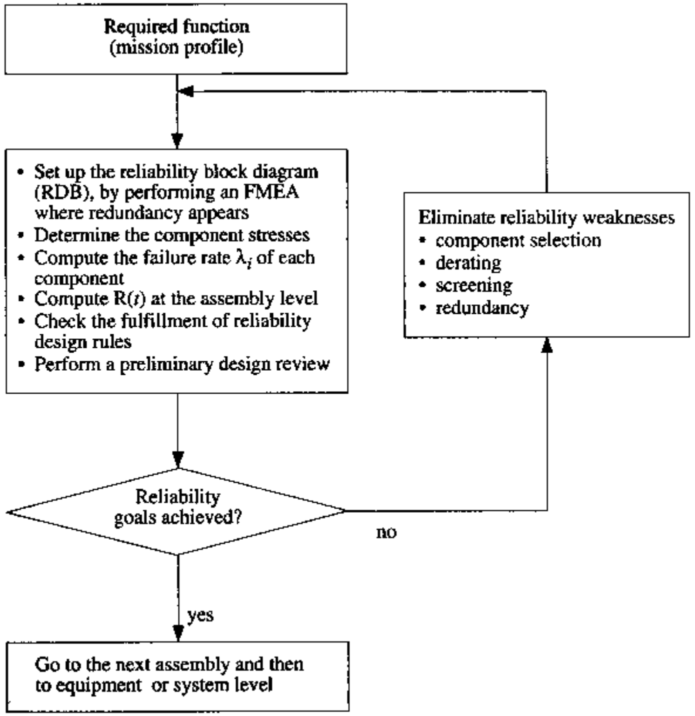


Figure 2.1 Reliability analysis procedure at assembly level

Taking into account the above remarks, Fig. 2.1 shows the reliability analysis procedure for an assembly as a flow chart. Also included are a failure mode analysis (FMEA/FMECA as described in Section 2.6) to check the validity of the assumed *failure modes*, and a verification of the fulfillment of *design guidelines for reliability* (Section 5.1) in a *preliminary design review* (Appendix A3.2.5). Verification of the assumed failure mode is essential where *redundancy* appears, in particular because of the *series element* in the reliability block diagram, see Figs. 2.8 and 2.9, as well as Figs. 6.17 and 6.18, for a comparison. To simplify the notation, *reliability* will be used instead of *predicted reliability* and *system* will be used for *technical system*, i.e. for a system with ideal human factors and logistical support.

2.2 Predicted Reliability of Equipment and Systems with Simple Structures

Simple structures are those for which a reliability block diagram *exists* and can be reduced to a *series/parallel form* with *independent* elements. For such an item, the *predicted reliability* is calculated according to the following procedure, see Fig. 2.1:

1. Definition of the required function and of its associated mission profile.
2. Derivation of the corresponding reliability block diagram (RBD).
3. Determination of the operating conditions of each element of the RBD.
4. Determination of the failure rate for each element of the RBD.
5. Computation of the reliability for each element of the RBD.
6. Computation of the item (system) reliability function $R_S(t)$.
7. Elimination of reliability weaknesses and return to step 1 or 2, as necessary.

This section discusses at some length steps 1 to 6, see Example 2.6 for the application to a simple, practice oriented, situation.

2.2.1 Required Function

The *required function* specifies the item's task. Its definition is the starting point for any analysis, as it defines failures. For practical purposes, parameters should be defined with tolerances and not merely as fixed values.

In addition to the required function, global *environmental conditions* at system level must also be defined. Among these, ambient temperature (e.g. +40°C), storage temperature (e.g. –20 to +60°C), humidity (e.g. 40 to 60%), dust, corrosive atmosphere, vibration (e.g. $0.5 g_n$, at 2 to 60 Hz), shock, noise (e.g. 40 to 70 dB), and power supply voltage variations (e.g. ±20%). From these global environmental conditions, the constructive characteristics of the system, and the internal loads, *operating conditions* (actual stresses) of every element of the system can be determined (step 3 of the above procedure).

Required function and environmental conditions are often *time dependent*, leading to the *mission profile*. A representative mission profile and the corresponding reliability targets should be defined in the system specifications (initially as a rough description and then refined step by step).

2.2.2 Reliability Block Diagram

The *reliability block diagram* (RBD) is an *event* diagram. It answers the following question: *Which elements of the item under consideration are necessary for the fulfillment of the required function and which can fail without affecting it?* Setting up an RBD involves first *partitioning* the item into elements with clearly defined tasks. The elements which are necessary for the required function are then connected *in series*, while elements which can fail with no effect on the required function (redundancy) are connected *in parallel*. Obviously, the ordering of the series elements is arbitrary. Elements which are not relevant (used) for the required function under consideration are removed (and put into a reference list), after having verified (FMEA) that their failure does not affect elements involved in the required function. These considerations make it clear that for a given system, each required function has its reliability block diagram.

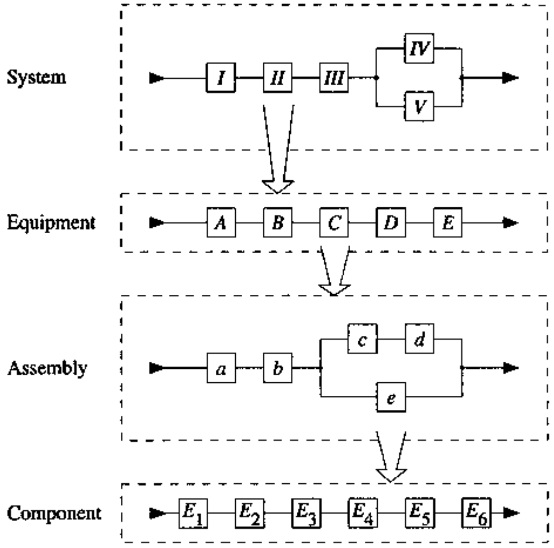


Figure 2.2 Procedure for setting up the reliability block diagram (RBD) of a system with four levels

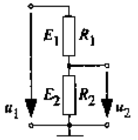
In setting up the reliability block diagram, care must be taken regarding the fact that only *two states* (good or failed) and *one failure mode* (e.g., opens or shorts) can be considered for each element. Particular attention must also be paid to the correct identification of the parts which appear *in series with a redundancy* (Section 6.9). For large equipment or systems the reliability block diagram is derived top down as indicated in Fig. 2.2 (for 4 levels as an example). At each level, the corresponding required function is derived from the one at the next higher level.

The technique of setting up reliability block diagrams is shown in the Examples 2.1 to 2.3 (see also Examples 2.5, 2.6, 2.14, and 2.15). One recognizes easily that a reliability block diagram differs basically from a *functional block diagram*.

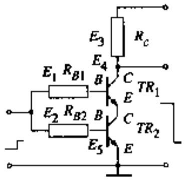
Examples 2.2 and 2.3 also show that one or more elements can appear *more than once* in a reliability block diagram, while the corresponding element is physically present only once in the item considered. To point out the strong dependence created by this fact, it is helpful to use *a box form other* than a square for these elements (in Example 2.2, when E_2 fails, the required function is fulfilled *only* if E_1, E_3, and E_5 work), see also Tab. 2.1. To avoid confusion, each physically different element of the item should bear its own number.

Example 2.1

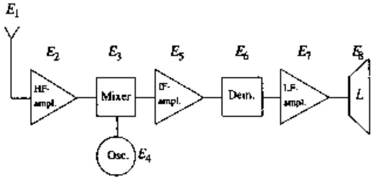
Set up the reliability block diagrams for the following circuits:



(i) Res. voltage divider



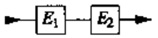
(ii) Electronic switch



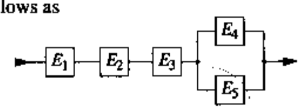
(iii) Simplified radio receiver

Solution

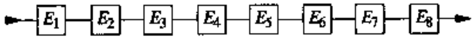
Cases (i) and (iii) exhibit no redundancy, i.e. for the required function (tacitly assumed here) all elements must work. In case (ii), transistors TR_1 and TR_2 are redundant if their failure mode is a *short* between emitter and collector (the failure mode for resistors is generally an open). From these considerations, the reliability block diagrams follows as



(i) Resistive voltage divider



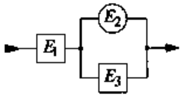
(ii) Electronic switch



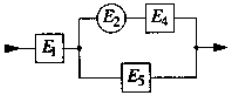
(iii) Simplified radio receiver

Example 2.2

An item is used for two different missions with the corresponding reliability block diagrams given in the figures below. Give the reliability block diagram for the case in which both functions are simultaneously required in a common mission.



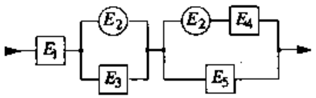
Mission 1



Mission 2

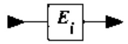
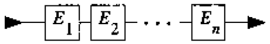
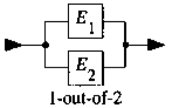
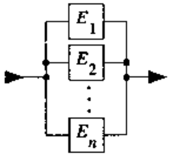
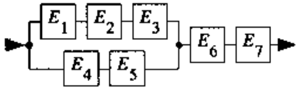
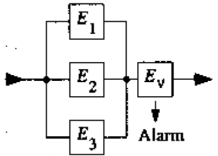
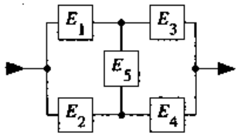
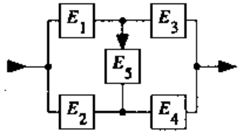
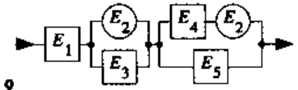
Solution

The simultaneous fulfillment of both required functions leads to the series connection of both reliability block diagrams (event diagram). Simplification is possible for element E_1 ($e_1 \cap e_1 = e_1$, see the notation used for Eq. (2.14)) but not for element E_2.



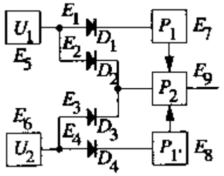
Mission 1 and 2

Table 2.1 Basic reliability block diagrams and associated reliability functions (nonrepairable until system failure, *independent elements* (except E_2 in 9), active redundancy; examples 7 to 9 are *complex structures* and can not be reduced to a series/parallel structure with independent elements)

Reliability Block Diagram	Reliability Function ($R_S = R_S(t)$, $R_i = R_i(t)$)	Remarks
 1	$R_S = R_i$	One-item structure, for $\lambda(t) = \lambda \Rightarrow R_i(t) = e^{-\lambda_i t}$
 2	$R_S = \prod_{i=1}^{n} R_i$	Series structure $\lambda_S(t) = \lambda_1(t) + \ldots + \lambda_n(t)$
 1-out-of-2 3	$R_S = R_1 + R_2 - R_1 R_2$	1-out-of-2 redundancy, $R_1(t) = R_2(t) = e^{-\lambda t} \Rightarrow$ $R_S(t) = 2e^{-\lambda t} - e^{-2\lambda t}$
 *k*-out-of-*n* 4	$R_1 = \ldots = R_n = R$ $R_S = \sum_{i=k}^{n} \binom{n}{i} R^i (1-R)^{n-i}$	*k*-out-of-*n* redundancy, for $k = 1 \Rightarrow$ $R_S = 1 - (1-R)^n$
 5	$R_S = (R_1 R_2 R_3 + R_4 R_5 - R_1 R_2 R_3 R_4 R_5) R_6 R_7$	Series / parallel structure
 2-out-of-3 6	$R_1 = R_2 = R_3 = R$ $R_S = (3R^2 - 2R^3) R_V$	Majority redundancy (general case $(n+1)$-out-of-$(2n+1)$)
 7	$R_S = R_5 (R_1 + R_2 - R_1 R_2)(R_3 + R_4 - R_3 R_4) + (1 - R_5)(R_1 R_3 + R_2 R_4 - R_1 R_2 R_3 R_4)$	Bridge structure (bidirectional on E_5)
 8	$R_S = R_4 [R_2 + R_1 (R_3 + R_5 - R_3 R_5) - R_1 R_2 (R_3 + R_5 - R_3 R_5)] + (1 - R_4) R_1 R_3$	Bridge structure (unidirectional on E_5)
 9	$R_S = R_2 R_1 (R_4 + R_5 - R_4 R_5) + (1 - R_2) R_1 R_3 R_5$	The element E_2 appears twice in the reliability block diagram

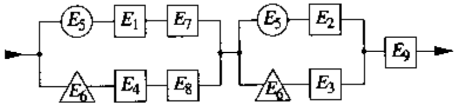
Example 2.3

Set up the reliability block diagram for the electronic circuit shown on the right. The required function asks for operation of P_2 (main assembly) and of P_1 or $P_{1'}$ (control cards).



Solution

This example is not as trivial as Examples 2.1 and 2.2. A good way to derive the reliability block diagram is to consider the mission " P_1 or $P_{1'}$ must work" and " P_2 must work" *separately*, and then to put both missions together as in Example 2.2.



The typical structures of reliability block diagrams are summarized in Table 2.1. Also given in Table 2.1 are the associated reliability functions for the case of *nonrepairable* elements (until system failure) with *active* (parallel) redundancy and statistically *independent elements* (Sections 2.2.5, 2.2.6, 2.3.1).

Table 2.2 Parameters influencing the failure rate of electronic components

Component	Ambient temp. (θ_A)	Junction temp. (θ_J)	Power stress (S)	Voltage stress (S)	Current stress (S)	Breakdown voltage	Technology	Complexity	Package	Application	Contact construction	Range	Production maturity	Environment (π_E)	Quality (π_Q)
Digital and linear ICs		D			x		x	x	x				x	x	x
Hybrid circuits	D	D	D	D	D	x	x	x	x	x	x	x	x	x	x
Bipolar transistors		D	D			x	x	x	x	x	x	x	x	x	x
FETs		D	D			x	x	x	x	x	x		x	x	x
Diodes		D				x	x		x	x	x	x	x	x	x
Thyristors		D				x	x		x		x	x	x	x	x
Optoelectronic components		D		x	x		x	x	x				x	x	x
Resistors	D		D				x					x	x	x	x
Capacitors	D			D			x					x	x	x	x
Coils, transformers	D		x	x			x						x	x	x
Relays, switches	D			x	x		x	x		x	x		x	x	x
Connectors	D				x		x		x	x	x	x	x	x	x

D denotes dominant, x denotes important

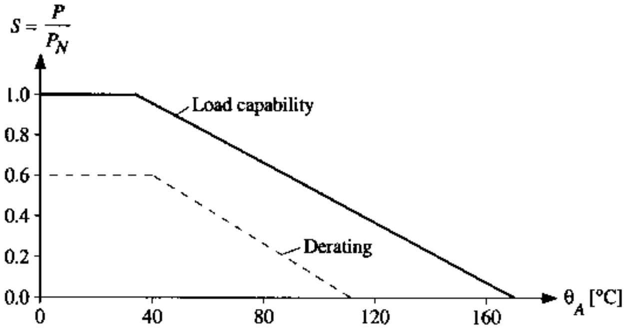


Figure 2.3 Load capability and typical derating curve (dashed) for a bipolar Si-transistor as function of the ambient temperature θ_A (P = dissipated power, P_N = rated power)

2.2.3 Operating Conditions at Component Level, Stress Factors

The *operating conditions* of each element in the reliability block diagram influence the item's reliability and have to be considered. These operating conditions are function of the *environmental conditions* (Section 3.1.1) and *internal loads,* in operating and dormant state. Table 2.2 gives the parameters relevant to the determination of electronic component failure rates.

A basic assumption is that components are in no way *overstressed.* In this context it is important to consider that the *load capability* of many electronic components decreases with increasing *ambient temperature.* This in particular for power, but also for voltage and current, as shown in Fig. 2.3 for the power dissipation P as a function of the ambient temperature θ_A for a bipolar Si transistor (under the assumption of a constant thermal resistance R_{JA}). The continuous line represents the *load capability.* To the right of the break point the junction temperature is nearly equal to 175°C (max. specified operating temperature). The dashed line gives a typical *derating curve* for such a device. *Derating* is the intentional nonutilization of the full load capability of a component with the purpose to reduce its failure rate. The *stress factor* (*stress ratio, stress*) S is defined as

$$S = \frac{\text{applied load}}{\text{rated load at } 40^\circ\text{C}}. \tag{2.1}$$

To give an idea, Figs. 2.4 to 2.6 show the influence of the temperature (ambient θ_A, case θ_C, or junction θ_J) and of the stress factor S on the failure rate of some electronic components (from IEC 61709 [2.24] as an example). Experience shows that for a good design and $\theta_A \le 40°C$ one should have $0.1 < S < 0.6$ for power, voltage, and current, $S \le 0.8$ for fan-out, and $S \le 0.7$ for U_{in} of lin. ICs (Table 5.1). $S < 0.1$ should also be avoided.

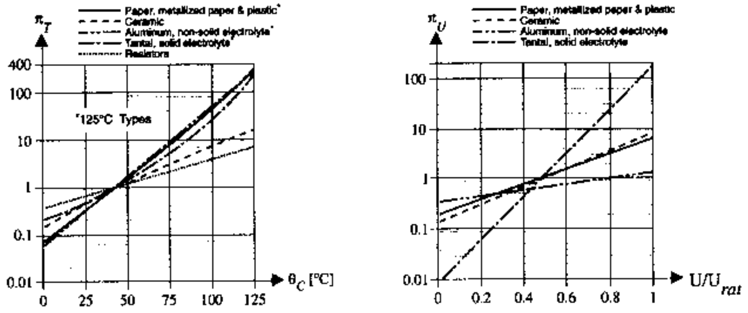


Figure 2.4 Factor π_T as function of the case temperature θ_C for capacitors and resistors, and factor π_U as function of the voltage stress for capacitors (examples from *IEC 61709* [2.24])

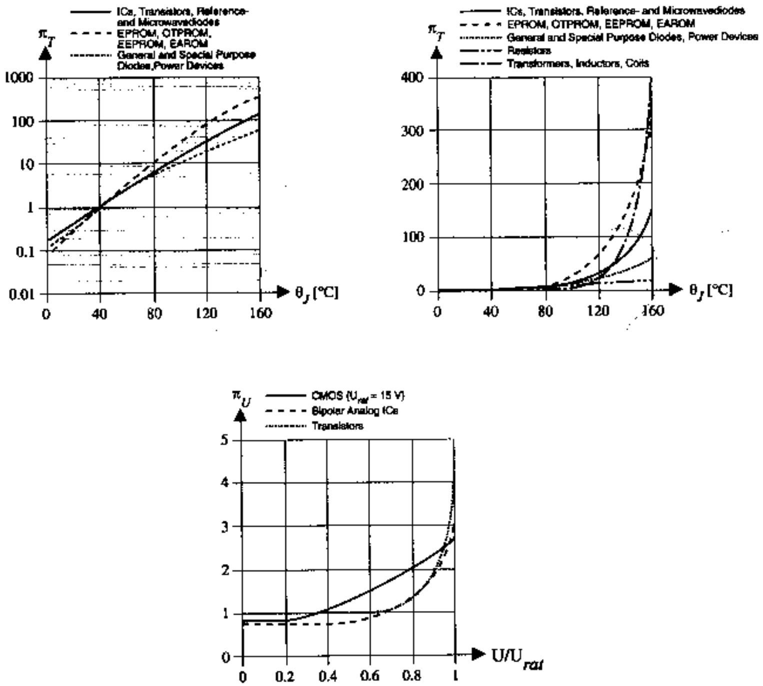


Figure 2.5 Factor π_T as function of the junction temperature θ_J (left, half log for semiconductors and right, linear for semiconductors, resistors, and coils) and factor π_U as function of the power supply voltage for semiconductors (examples from *IEC 61709* [2.24])

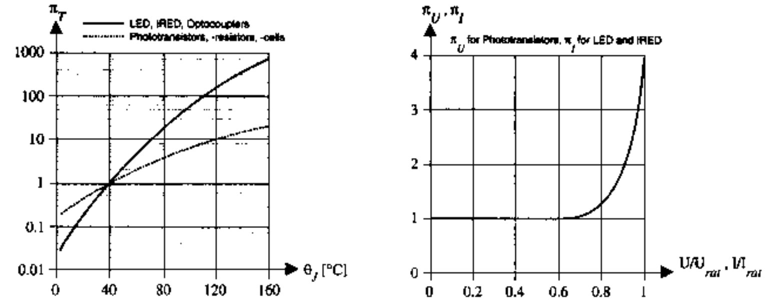


Figure 2.6 Factor π_T as function of the junction temperature θ_J and factors π_U and π_I as function of voltage and current stress for optoelectronic devices (examples from *IEC 61709* [2.24])

2.2.4 Failure Rate of Electronic Components

The failure rate $\lambda(t)$ of an item is the probability (referred to δt) of a failure in the interval $(t, t+\delta t]$ given that the item was new at $t=0$ and did not fail in the interval $(0, t]$. For a *large* population of *statistically identical,* independent items, $\lambda(t)$ exhibits three successive phases: one of *early failures*, one with *constant* (or nearly so) *failure rate* and one involving *failures* due to *wearout.* Early failures should be eliminated through *screening* (Chapter 8). *Wearout* failures can be expected for some few electronic components (electrolytic capacitors, power and optoelectronic devices, ULSI-ICs) as well as for mechanical and electromechanical components. They must be considered on a case-by-case basis by setting up a preventive maintenance strategy.

To simplify computations, reliability prediction is often performed by assuming a *constant* (time independent) *failure rate* during the *useful life*

$$\lambda(t) = \lambda.$$

This approximation greatly simplify computations, since a constant failure rate leads to a flow of failures described by a *Poisson process*, i.e. to a process with a *memoryless property* (Eqs. (A6.87) and (A7.41)). The failure rate of components can be assessed *experimentally* by accelerated reliability tests or from field data (if operating conditions are known) with appropriate statistical data analysis (Sections 7.2.2 and 7.4). For established electronic/electromechanical components, appropriate figures/models are given in *failure rate/models handbooks* [2.21 to 2.29]. Among these are *Bellcore TR-332* (1997) [2.22], *CNET RDF 93* [2.23], *IEC 61709* (1996) [2.24], and *MIL-HDBK-217 F* (Not.2, 1995) [2.26]. New or announced

Table 2.3 Indicative figures for environmental conditions and for the corresponding environmental factor (π_E) according to *MIL-HDBK-217 F* and *CNET RDF 93*

Environment	Stress (indicative values)					π_E factor			
	Vibrations	Fog	Dust	RH (%)	Mech. shocks	ICs	DS	R	C
G_B (Ground benign)	2 – 200 Hz $\le 0.1\ g_n$	l	l	40 – 70	$\le 5\ g_n$ / 22 ms	1*	1	1	1
G_F (Ground fixed)	2 – 200 Hz $1\ g_n$	m	m	5 – 100	$\le 20\ g_n$ / 6 ms	2.5	2.5**	2.5 – 3	2.5 – 3
G_M (Ground mobile)	2 – 500 Hz $3\ g_n$	m	m	5 – 100	$30\ g_n$ / 11 ms to $100\ g_n$ / 6 ms	5	5**	5.5 – 9	5.5 – 9
N_S (Nav. sheltered)	2 – 200 Hz $2\ g_n$	l	l	5 – 100	$10\ g_n$ / 11 ms to $30\ g_n$ / 6 ms	4	4**	4 – 7	4 – 7
N_U (Nav. unsheltered)	2 – 200 Hz $5\ g_n$	h	m	10 – 100	$10\ g_n$ / 11 ms to $50\ g_n$ / 2.3 ms	6	6**	7 – 12	7 – 12

C = capacitors, DS = discrete semiconductors, RH = relative humidity, R = resistors, h = high, m = medium, l = low, $g_n = 10\,\text{m/s}^2$, * 0.5 in *MIL HDBK-217 F*

are *CNET RDF 99* [2.23] and *RAC (FRM)* [2.28]. *IEC 61709* gives the *laws of dependency* of the failure rate on the different stresses (temperature, voltage, etc.) and must be supported by a set of *reference failure rates* λ_{ref}, valid for a *standard environment* (40°C ambient temperature θ_A, G_B as per Table 2.3, and steady-state conditions in the field). The time should be ripe for *international agreements* on failure rate figures/models for *reliability predictions at equipment and system level*, see [1.2 (1996)] and the discussion on p. 40. *IEEE Std 1413* [2.25] has been issued recently to identify required elements for a credible reliability prediction.

Failure rates are often taken from one of the above handbooks (or from one's own field experience) for the calculation of the predicted reliability. The *models* in these handbooks have a simple structure, of the form

$$\lambda = \lambda_0\,\pi_T\,\pi_E\,\pi_Q\,\pi_A \tag{2.2}$$

for discrete components and

$$\lambda = \pi_Q\,\pi_L\,(C_1\,\pi_T + C_2\,\pi_E) \tag{2.3}$$

for ICs. The models given in *IEC 61709* are further simplified, basically to

$$\lambda = \lambda_{ref}\,\pi_T\,\pi_U\,\pi_I, \tag{2.4}$$

by taking $\pi_E = \pi_Q = 1$, because of the assumed standard (industrial) environment (G_B in Table 2.3) and standard quality level (CECC qualification in Table 2.4). Indicative figures for the factors π_E and π_Q are given in Tables 2.3 and 2.4.

The value of λ lies between about $10^{-10}\,\text{h}^{-1}$ for passive components and $10^{-7}\,\text{h}^{-1}$ for VLSI ICs (see Table 3.2 and Example 2.4). The unit $10^{-9}\,\text{h}^{-1}$ is often designated by *FIT* (failures in time).

Table 2.4 Indicative figures for the quality factor (π_Q) according to *CNET RDF 93*

	Qualification (π_{Q1})			Evaluation (π_{Q2})	
	Reinforced	CECC*	no special	with	without
Monolithe ICs	0.7	1.0	1.3	1.0	1.3
Hybrid ICs	0.2	1.0	1.5	1.0	1.5
Discrete Semiconductors	0.2	1.0	2.0	1.0	2.0
Resistors	0.1	1.0	2.0	1.0	2.0
Capacitors	0.1	1.0	2.0	1.0	2.0

* correspond approximately to *MIL-HDBK-217 F* classes B–1, JANTX, M

In general, λ_0 and λ_{ref} increase exponentially with temperature. Figures 2.4 to 2.6 show this for some components (examples from *IEC 61709* [2.24]). In addition, the influence of the *stress factor S* is illustrated by the factors π_U and π_I. For the factor π_T as a function of the junction temperature θ_J, an *Arrhenius Model* is often used as an approximation. In the case of only one *dominant failure mechanism*, Eq. (7.56) gives the ratio of the π_T factors at two temperatures T_2 and T_1

$$\frac{\pi_{T_2}}{\pi_{T_1}} = A \approx e^{\frac{E_a}{k}\left(\frac{1}{T_1}-\frac{1}{T_2}\right)},$$

where A is the *acceleration factor*, k the Boltzmann's constant ($8.6\cdot 10^{-5}$ eV/K), T the *junction temperature* (in Kelvin degrees), and E_a the *activation energy* in eV. As in Figs. 2.4 to 2.6, experience shows that activation energies often lie between 0.3eV and 0.7eV for Si devices. The design guideline $\theta_J \le 100$°C, if possible $\theta_J \le 80$°C, given in Section 5.1 for semiconductor devices is based on this consideration. The models in *IEC 61709* assumes for π_T two dominant failure mechanisms with activation energies E_{a_1} and E_{a_2} (about 0.3eV for E_{a_1} and 0.7eV for E_{a_2}). The corresponding equation for π_T in this case takes the form

$$\pi_T = \frac{A\,e^{z\,E_{a_1}} + (1-A)\,e^{z\,E_{a_2}}}{A\,e^{z_{ref}\,E_{a_1}} + (1-A)\,e^{z_{ref}\,E_{a_2}}}, \tag{2.5}$$

where $0 \le A \le 1$ is a constant, $z = (1/T_{ref} - 1/T_2)/k$, and $z_{ref} = (1/T_{ref} - 1/T_1)/k$ with $T_{ref} = 313\,\text{K}$ (40°C).

Assuming $\pi_E = \pi_Q = 1$, failure rate computations lead to figures which for industrial applications agree often *reasonably well* with field data. This holds at the *equipment and system level*, although deviations up to a factor 10 can occur at the component level, depending on the failure rate catalog used (Example 2.4).

Discussions over comparison with obsolete data should be dropped, and it would seem to be opportune at this time to *unify models and data*. Thereby, *international standards* should be limited to the laws for *dominant failure mechanisms*, failure rate models should cover basically only intrinsic failures, *simple models* should be available for predictions in practice, *models based on failure mechanisms* should be developed as a basis for the simplified models (taking into account important production parameters), and the assumption of $\lambda \approx 0$ (or $\lambda < 10^{-9}\,\mathrm{h}^{-1}$) should be confined to components with established manufacturing processes and possessing a reserve with respect to the technological limits, see also the remark at the end of Section 7.4.

Example 2.4

For indicative purpose, the following table gives the failure rate of some electronic components computed according to different data bases for $\theta_A = 40°\mathrm{C}$, $\theta_J = 50°\mathrm{C}$, $S = 0.5$, G_B, and π_Q for CECC certified or B–1, JANTX, and P devices (λ in $10^{-9}\,\mathrm{h}^{-1}$). Experimental data for memories [3.2 (1993), 3.7] agree well with the lowest values given below.

	MIL-HDBK-217 F	*RDF 93*	*SN 29500*	λ_{ref}*
4 M DRAM	37	61	34	30
1 M SRAM	103	88	56	50
1 M EPROM	32	54	101	30
80486 µP	509**	150	48	100
LM741 op amp	24	23	9	10
Dig. CMOS, 30,000gates, 40 pins (ASIC)	144**	34	59	20
100 mA GP diode	2	2	2	2
LED	1.5	2	2	2
1 W bip. transistor	0.5	3	3.5	2
1 W MOSFET	27	4	27	2
1 nF ceramic capacitor (125°C, class 2)	1.5	2	2	1
1 µF foil capacitor	3	2	1	1
100 µF Ta solid capacitor (125°C, ≤ 0.4 Ω / V)	2	13	2	5
100 µF Al wet capacitor (125°C)	18	10	4	10
100 kΩ metal film resistor	0.5	0.3	0.1	0.2
50 kΩ cermet potentiometer	41	16	40	10

* λ_{ref} is the failure rate assumed here as a *possible* reference for computations according to *IEC 61709* [2.24]

** obviously too high

2.2.5 Reliability of One-Item Structures

A *one-item* (nonrepairable) structure is characterized by the distribution function $\mathrm{F}(t) = \Pr\{\tau \le t\}$ of its failure-free operating time τ. Its *reliability function* $\mathrm{R}(t)$, i.e. the probability of no failure in the interval $(0, t]$, is given by (Eq. (A6.26))

$$\mathrm{R}(t) = \Pr\{\text{no failure in } (0, t]\} = \Pr\{\tau > t\} = 1 - \mathrm{F}(t). \tag{2.6}$$

$\mathrm{R}(0) = 1$ is often tacitly assumed. The expected value (mean) of the failure-free operating time τ, designated as *MTTF* (*mean time to failure*), can be computed from Eq. (A6.38) as

$$MTTF = \mathrm{E}[\tau] = \int_0^\infty \mathrm{R}(t)\,dt. \tag{2.7}$$

Equation (2.7) is an important relationship. It is valid not only for a one-item structure, but it also holds for an item of arbitrary complexity. $\mathrm{R}_S(t)$ and $MTTF_S$ will be used to emphasize this point (the index S stands for system and designates the highest integration level considered)

$$MTTF_S = \int_0^\infty \mathrm{R}_S(t)\,dt. \tag{2.8}$$

Furthermore, Eq. (2.7) can also be used for *repairable* items. Assuming that a failed item is replaced by a statistically equivalent one, a new failure-free operating time τ with the *same* distribution function as the former one is started after replacement, thus yielding *the same expected value*. With this, *MTTF* or $MTTF_S$ defines the mean time to failure of a given item, independently of whether it is repairable or not (Chapter 6). The only implicit assumption is that after replacement or repair, the item is *as-good-as-new*. If this is not the case, i.e. for instance only the failed element of the item has been replaced, a new *MTTF* has to be considered, for example $MTTF_{Si}$ in Table 6.2. Note also that only the failure-free operating time and not the time between consecutive failures (sum of a failure-free operating time and a repair or replacement time) is considered.

Should an item exhibit a *useful life* limited to T_L, i.e. $\mathrm{R}(t) = 0$ for $t > T_L$, Eq. (2.7) converts to

$$MTTF_L = \int_0^{T_L} \mathrm{R}_S(t)\,dt.$$

In the following, $T_L = \infty$ and $\mathrm{R}(t)$ continuous will be tacitly assumed.

For a one-item structure, the failure rate $\lambda(t)$, as per Eq. (A6.27), is given by

$$\lambda(t) = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{t < \tau \le t + \delta t \mid \tau > t\} = -\frac{d\,\mathrm{R}(t)/dt}{\mathrm{R}(t)}.$$

For $R(0) = 1$ it follows that

$$R(t) = e^{-\int_0^t \lambda(x)dx} , \qquad (2.9)$$

from which

$$R(t) = e^{-\lambda t} \qquad (2.10)$$

for the special case of $\lambda(t) = \lambda$. The mean time to failure is in this case equal to $1/\lambda$ and it is common usage to define

$$\frac{1}{\lambda} = MTBF, \qquad (2.11)$$

where $MTBF$ stands for *mean operating time between failures* (formerly *mean time between failures*).

As shown in Eq. (2.9), the reliability function of a one-item structure is *completely determined* by the failure rate $\lambda(t)$. In the case of electronic components, $\lambda(t) = \lambda$ can be often assumed. The failure-free operating time then exhibits an *exponential distribution* (Eq. (A6.81)). For a nonconstant failure rate, the distribution function of the failure-free operating time can often be approximated by the weighted sum of a Gamma distribution (Eq. (A6.97)) with $\beta < 1$ and a shifted Weibull distribution (Eq. (A6.96)) with $\beta > 1$ (Example 7.16).

Equations (2.9) and (2.10) assume that the one-item structure is *new* at time $t = 0$. Also of interest in some applications is the probability of failure free operation during an interval $(0, t]$ under the condition that the item has already operated without failure for x_0 time units before $t = 0$. This quantity is a conditional probability designated by $R(t, x_0)$ and given as per Eq. (A6.29) by

$$R(t, x_0) = \Pr\{\tau > t + x_0 \mid \tau > x_0\} = \frac{R(t + x_0)}{R(x_0)} = e^{-\int_{x_0}^{t+x_0} \lambda(x)dx} . \qquad (2.12)$$

For $\lambda(x) = \lambda$, Eq. (2.12) reduces to Eq. (2.10). This *memoryless property* occurs *only* with *constant failure rate*. Its use will simplify calculations in the next sections and in Chapter 6.

2.2.6 Reliability of Series-Parallel Structures

Calculation of the reliability of equipment or systems can often be performed using the basic models of Table 2.1. The one-item structure was introduced in Section 2.2.5. Series, parallel, and series/parallel structures are considered in this Section (the last three models of Table 2.1 follow then in Section 2.3).

2.2.6.1 Items without Redundancy

From a reliability point of view, an item has *no redundancy* (and constitutes a series model) if all elements must work in order to fulfill the required function. The reliability block diagram consists of the series connection of all elements (E_1 to E_n) of the item (Table 2.1). For calculation purposes it is often assumed that each element operates or fails *independently* from every other element. This assumption is often satisfied, *at least until the first failure* (system failure because of the series structure). For the calculation, let e_i be the event

$e_i \equiv$ element E_i works without failure in the interval $(0, t]$.

The probability of this event is the reliability function $R_i(t)$ of the element E_i, i.e.

$$\Pr\{e_i\} = \Pr\{\tau_i > t\} = R_i(t). \qquad (2.13)$$

The item as a whole does not fail in the interval $(0, t]$ *only* when *all* elements, $E_1, \ldots, E_n$ do not fail in that interval, thus

$$R_S(t) = \Pr\{e_1 \cap \ldots \cap e_n\}. \qquad (2.14)$$

Here and in the following, the index S stands for system (technical system) and refers to the highest integration level of the considered item. Due to the assumed independence among the elements $E_1, \ldots, E_n$ and thus among the events $e_1, \ldots, e_n$, it follows (Eq. (A6.9)) that for the *reliability function* $R_S(t)$

$$R_S(t) = \prod_{i=1}^{n} R_i(t). \qquad (2.15)$$

The *failure rate* of the item is then given, considering Eq. (2.9), by

$$\lambda_S(t) = \sum_{i=1}^{n} \lambda_i(t). \qquad (2.16)$$

Equation (2.16) leads to the following important conclusion:

> ***The failure rate of an item (equipment or system) without redundancy that consists of independent elements is equal to the sum of the failure rates of its elements.***

The item's *mean time to failure* follows from Eq. (2.8). The special case in which all elements have a *constant failure rate* $\lambda_i(t) = \lambda_i$ leads to

$$R_S(t) = e^{-\lambda_S t}, \qquad \lambda_S(t) = \lambda_S = \sum_{i=1}^{n} \lambda_i, \qquad \frac{1}{\lambda_S} = MTBF_S. \qquad (2.17)$$

2.2.6.2 Concept of Redundancy

High reliability, availability, and/or safety at equipment or system level can often only be reached with the help of redundancy. *Redundancy* is the existence of more than one means for performing a required function in an item. Redundancy does not just imply a *duplication of hardware* because it can also often be implemented by coding, at the software level, or in another form. However, to avoid *common mode failures*, redundant elements should be realized (designed and produced) *independently* from each other. Irrespective of the *failure mode* (e.g. shorts or opens), redundancy still appears in *parallel* on the reliability block diagram. In setting up the reliability block diagram, particular attention must be paid to the element which must be put *in series* to a redundancy, an FMEA is often *necessary* for such a decision. Should the redundant elements fulfill only a part of the required function a *pseudo redundancy* exist. From the operating point of view, one distinguishes between active, warm, and standby redundancy:

1. *Active Redundancy* (parallel, hot): Redundant (reserve) elements are subjected from the beginning to the *same load* as operating elements, *load sharing* is possible (Section 2.3.5); failure rate in the reserve state is the same as in the operating state.
2. *Warm Redundancy* (lightly loaded): Redundant elements are subjected to a *lower load* until one of the operating element fails, *load sharing* is also possible (Section 2.3.5); failure rate in the reserve state is lower than in the operating state.
3. *Standby Redundancy* (cold, nonloaded): Redundant elements are subjected to *no load* until one of the operating element fails, no *load sharing* is possible; failure rate in the reserve state is *assumed* to be zero.

Important redundant structures with independent elements in active redundancy are investigated in Sections 2.2.6.3 to 2.3.4. Analysis of warm and standby redundancy, as well as cases in which the failure of an element has an *influence* on the failure rate of other elements (*load sharing* for example) are given in Section 2.3.5 and Chapter 6 (results for a repair rate $\mu = 0$).

2.2.6.3 Parallel Models

A parallel model consists of n (often statistically identical) elements in *active redundancy*, of which k are necessary to perform the required function and the remaining $n-k$ are in reserve. Such a structure is designated as a *k-out-of-n redundancy*, also known as *k-out-of-n: G.*

Let us consider first the case of an active *1-out-of-2 redundancy* as given in Table 2.1 (third line). The required function is fulfilled if at least one of the elements E_1 or E_2 works without failure in the interval $(0, t]$. With the same notation as for Eq. (2.14) it follows that

$$R_S(t) = \Pr\{e_1 \cup e_2\} = \Pr\{e_1\} + \Pr\{e_2\} - \Pr\{e_1 \cap e_2\}. \qquad (2.18)$$

Assuming that the elements E_1 and E_2 work or fail *independently* of each other, Eq. (2.18) yields for the *reliability function* $R_S(t)$, see Eqs. (A6.13) and (A6.8),

$$R_S(t) = \Pr\{e_1\} + \Pr\{e_2\} - \Pr\{e_1\}\Pr\{e_2\} = R_1(t) + R_2(t) - R_1(t)R_2(t). \qquad (2.19)$$

The *mean time to failure* $MTTF_S$ can be computed from Eq. (2.8). The special case of two identical elements with constant failure rate ($R_1(t) = R_2(t) = e^{-\lambda t}$) leads to

$$R_S(t) = 2e^{-\lambda t} - e^{-2\lambda t} \qquad (2.20)$$

and

$$MTTF_S = \frac{2}{\lambda} - \frac{1}{2\lambda} = \frac{3}{2\lambda}. \qquad (2.21)$$

Generalization to an active *k-out-of-n redundancy* with identical elements ($R_1(t) = \ldots = R_n(t) = R(t)$) follows from the equation for the *binomial distribution* (Eq. (A6.120)) by setting $p = R(t)$

$$R_S(t) = \sum_{i=k}^{n} \binom{n}{i} R^i(t)(1-R(t))^{n-i}. \qquad (2.22)$$

$R_S(t)$ can be interpreted as the probability of observing at least k successes in n Bernoulli trials with $p = R(t)$. The mean time to failure $MTTF_S$ can be computed from Eq. (2.8). For $k = 1$ and $R(t) = e^{-\lambda t}$ it follows that

$$R_S(t) = 1 - (1 - e^{-\lambda t})^n \qquad \text{and} \qquad MTTF_S = \frac{1}{\lambda}(1 + \frac{1}{2} + \ldots + \frac{1}{n}). \qquad (2.23)$$

The improvement in $MTTF_S$ shown by Eqs. (2.21) and (2.23) becomes much greater when *repair* without interruption of operation at system level is allowed, factor $\mu / 2\lambda$ instead of $3/2$ for an active 1-out-of-2 redundancy, where $\mu = 1/MTTR$ is the constant repair rate (Tables 6.6 and 6.8). However, as shown in Fig. 2.7, the increase of the reliability function $R_S(t)$ caused by redundancy is very *important* for *short missions* ($t << 1/\lambda$), even in the nonrepairable case. Other comparisons between series/parallel structures are given in Figs. 2.8 and 2.9 (Figs. 6.17 and 6.18 for the repairable case).

In addition to the k-out-of-n redundancy described by Eq. (2.22), attention has been paid in the literature to cases in which the fulfillment of the required function asks that *not more* than $n-k$ *consecutive* elements fail. Such a structure, known as *consecutive k-out-of-n system*, is theoretically more reliable than the corresponding k-out-of-n redundancy. For a k-out-of-n consecutive system with n identical and independent elements in active redundancy (each with reliability R) it holds that

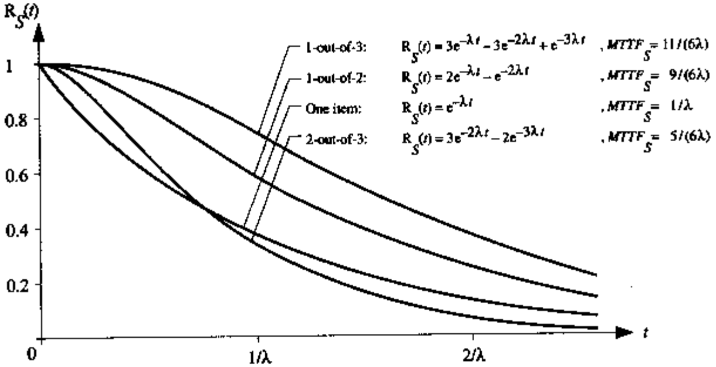


Figure 2.7 Reliability function for the one-item structure (as reference) and for some active redundancies which are nonrepairable until system failure (constant failure rates, identical elements, no load sharing, see Section 2.3.5 for load sharing)

$$R_S = \Pr\{\text{no block with more than } n-k \text{ consecutive failed elements}\}$$
$$= \sum_{i=0}^{n} g(n,i)\, R^{n-i}(1-R)^i, \tag{2.24}$$

with $g(n,i) = \binom{n}{i}$ for $i \le n-k$, $g(a,a) = 0$ for $a \ge n-k+1$ and $g(a,b) = g(a-1,b) + g(a-2,b-1) + \ldots + g(a-n+k-1, b-n+k)$ otherwise [2.34, 2.38]. For example, $n = 5$ and $k = 3$ yields from Eq. (2.22)

$$R_S = R^5 + 5R^4(1-R) + 10R^3(1-R)^2$$

and from Eq. (2.24)

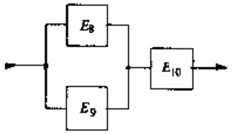
$$R_S = R^5 + 5R^4(1-R) + 10R^3(1-R)^2 + 7R^2(1-R)^3 + R(1-R)^4. \tag{2.25}$$

Examples of consecutive k-out-of-n systems are conveying systems and relay stations. However, for these kinds of application importance should be given to the verification of the assumption that all elements are *independent* (common stresses, external influences, load sharing).

2.2.6.4 Series/Parallel Structures

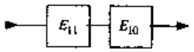
Series/parallel structures can be investigated through successive use of the results for series and parallel models. This is in particular true for *nonrepairable* items with *active redundancy* and *independent* elements. To demonstrate the procedure, let us consider Example 5 in Table 2.1:

1st step: The series elements E_1, E_2, E_3 are replaced by E_8, E_4 and E_5 by E_9, and E_6 and E_7 by E_{10}, yielding



with
$R_8(t) = R_1(t)R_2(t)R_3(t)$
$R_9(t) = R_4(t)R_5(t)$
$R_{10}(t) = R_6(t)R_7(t)$.

2nd step: The 1-out-of-2 redundancy E_8 and E_9 is replaced by E_{11}, giving



with $R_{11}(t) = R_8(t) + R_9(t) - R_8(t)R_9(t)$

3rd step: From steps 1 and 2, the *reliability function* of the item follows as

$$R_S = R_{11} R_{10} = (R_1 R_2 R_3 + R_4 R_5 - R_1 R_2 R_3 R_4 R_5) R_6 R_7 \tag{2.26}$$

with $R_S = R_S(t)$ and $R_i = R_i(t)$, $i = 1, \ldots, 7$.

The mean time to failure can be computed from Eq. (2.8) with $R_S(t)$ as per Eq. (2.26). Should all elements have a constant failure rate (λ_1 to λ_7), then

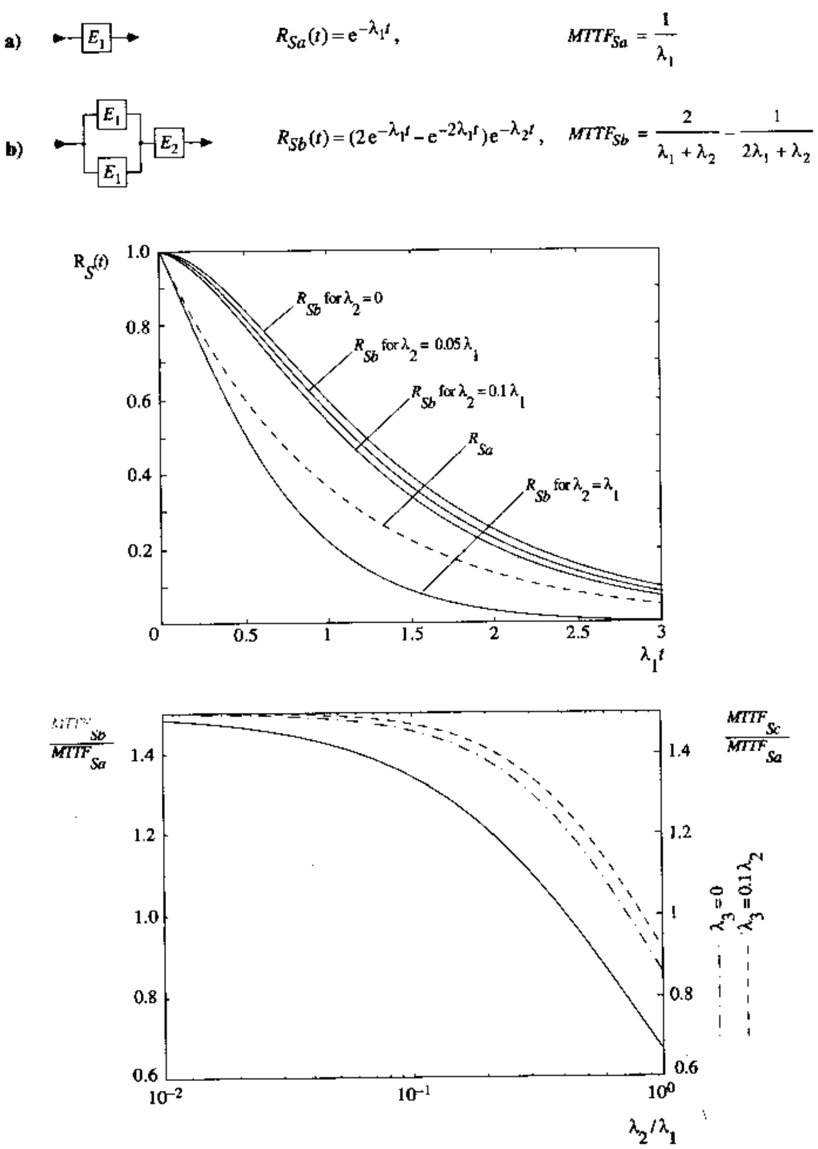
$$R_S(t) = e^{-(\lambda_1+\lambda_2+\lambda_3+\lambda_6+\lambda_7)t} + e^{-(\lambda_4+\lambda_5+\lambda_6+\lambda_7)t} - e^{-(\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7)t}$$

and

$$MTTF_S = \frac{1}{\lambda_1+\lambda_2+\lambda_3+\lambda_6+\lambda_7} + \frac{1}{\lambda_4+\lambda_5+\lambda_6+\lambda_7} - \frac{1}{\lambda_1+\lambda_2+\lambda_3+\lambda_4+\lambda_5+\lambda_6+\lambda_7}. \tag{2.27}$$

Under the assumptions of active redundancy, independent elements, no repair until system failure, and constant failure rates, the *reliability function* $R_S(t)$ of a system with series/parallel structure is given by an algebraic *sum of exponential functions*. The *mean time to failure* $MTTF_S$ follows then directly from the exponent terms of $R_S(t)$, see Eqs. (2.26) and (2.27) for an example.

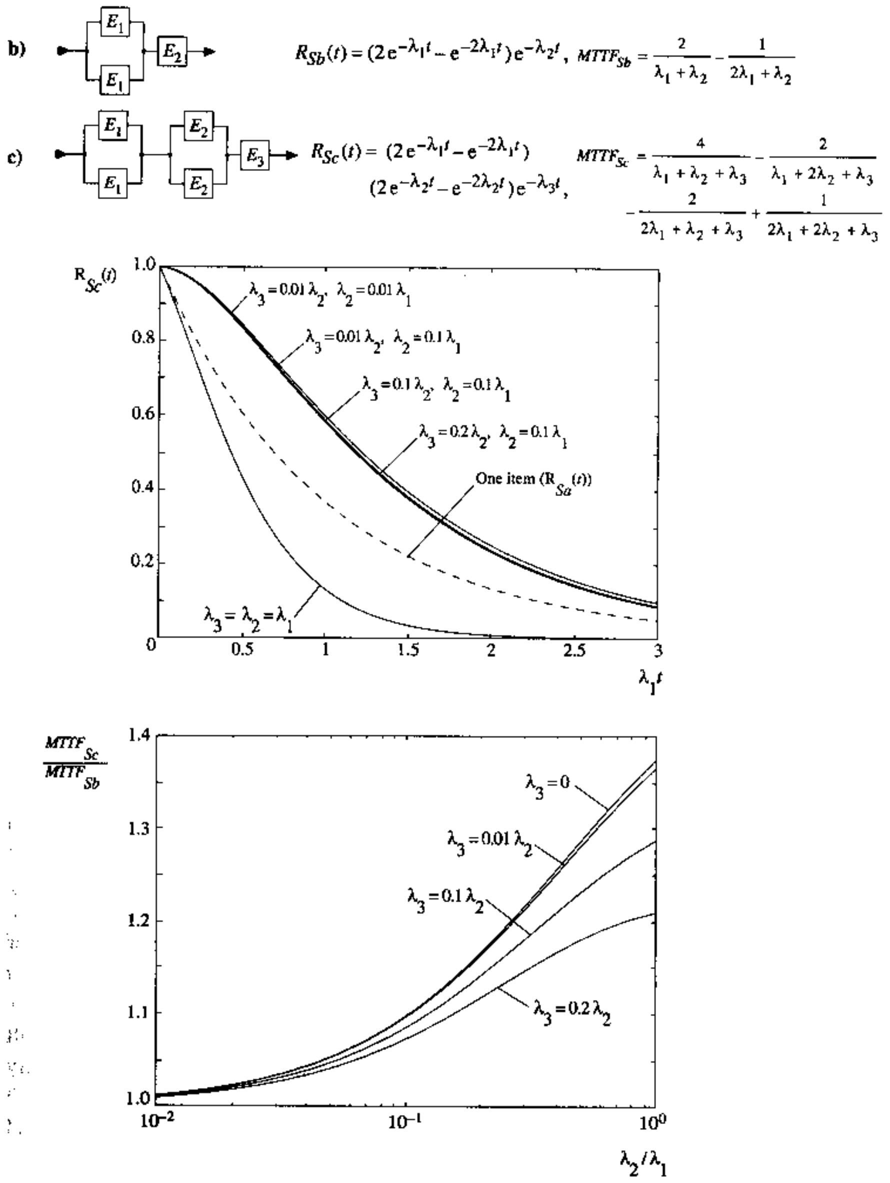
The use of *redundancy* implies the introduction of a *series element* in the reliability block diagram which takes into account the parts which are common to the redundant elements, creates the redundancy (as in the case of majority redundancy, see Example 2.5), or assumes a control and/or *switching function*. For a design engineer it is important to evaluate the influence of the series element on the reliability of the redundant structure. Figures 2.8 and 2.9 allow such an evaluation to be made for the case in which *constant failure rates*, *independent elements*, and *active redundancy* can be assumed. In Fig. 2.8, a one-item structure (element E_1 with failure rate λ_1) is compared with a 1-out-of-2 redundancy with a series element



a) $R_{Sa}(t)=e^{-\lambda_1 t}$, $MTTF_{Sa}=\dfrac{1}{\lambda_1}$

b) $R_{Sb}(t)=(2e^{-\lambda_1 t}-e^{-2\lambda_1 t})e^{-\lambda_2 t}$, $MTTF_{Sb}=\dfrac{2}{\lambda_1+\lambda_2}-\dfrac{1}{2\lambda_1+\lambda_2}$

Figure 2.8 Comparison between the one-item structure and a 1-out-of-2 active redundancy with series element (nonrepairable until system failure, independent elements, constant failure rates λ_1 and λ_2, λ_1 remains the same in both structures, equations according to Table 2.1; given on the right-hand side is $MTTF_{Sc}$ / $MTTF_{Sa}$ with $MTTF_{Sc}$ from Fig. 2.9; see Fig. 6.17 for the repairable case)



b) $R_{Sb}(t)=(2e^{-\lambda_1 t}-e^{-2\lambda_1 t})e^{-\lambda_2 t}$, $MTTF_{Sb}=\dfrac{2}{\lambda_1+\lambda_2}-\dfrac{1}{2\lambda_1+\lambda_2}$

c) $R_{Sc}(t)=(2e^{-\lambda_1 t}-e^{-2\lambda_1 t})(2e^{-\lambda_2 t}-e^{-2\lambda_2 t})e^{-\lambda_3 t}$,

$$MTTF_{Sc}=\frac{4}{\lambda_1+\lambda_2+\lambda_3}-\frac{2}{\lambda_1+2\lambda_2+\lambda_3}-\frac{2}{2\lambda_1+\lambda_2+\lambda_3}+\frac{1}{2\lambda_1+2\lambda_2+\lambda_3}$$

Figure 2.9 Comparison between basic series/parallel structures (nonrepairable until system failure, active redundancy, independent elements, constant failure rates λ_1 to λ_3, λ_1 and λ_2 remain the same in both structures, equations according to Table 2.1; see Fig. 6.18 for the repairable case) (element E_2 with failure rate λ_2). In Fig. 2.9, the 1-out-of-2 redundancy with a series element E_2 is compared with the structure which would be obtained if a 1-out-of-2 redundancy for element E_2 with a series element E_3 would become necessary, obviously for $\lambda_3 < \lambda_2 < \lambda_1$ (the limiting cases $\lambda_1 = \lambda_2$ for Fig. 2.8 and $\lambda_1 = \lambda_2 = \lambda_3$ for Fig. 2.9 have an indicative purpose only). The three cases are labeled a), b), and c), respectively. The upper part of Figs. 2.8 and 2.9 depict the reliability functions and the lower part the ratios $MTTF_{Sb} / MTTF_{Sa}$ and $MTTF_{Sc} / MTTF_{Sb}$, respectively. The comparison between case a) of Fig. 2.8 and case c) of Fig. 2.9 given as $MTTF_{Sc} / MTTF_{Sa}$ on Fig. 2.8, shows a much lower dependency on λ_2 / λ_1. From Figs. 2.8 and 2.9 following *design guideline* can be derived:

> *The failure rate λ_2 of the series element in a nonrepairable (until system failure) 1-out-of-2 active redundancy should not be larger than 10% of the failure rate of the redundant elements λ_1 (the 10% rule applies also for the case of λ_3 in Fig. 2.9), i.e.*
>
> $$10\lambda_3 \le \lambda_2 < 0.1\lambda_1. \tag{2.28}$$

The investigation of the structures given in Figs. 2.8 and 2.9 in the repairable case (with $\mu = 1/MTTR$ as constant repair rate) leads in Section 6.6 to more severe conditions ($\lambda_2 \le 0.01\lambda_1$ in general, and $\lambda_2 \le 0.002\lambda_1$ for $\mu/\lambda_1 > 500$), see Figs. 6.17 and 6.18.

2.2.6.5 Majority Redundancy

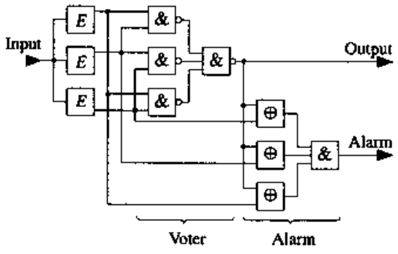
Majority redundancy is a special case of a k-out-of-n redundancy, used primarily in digital circuits. $2n+1$ outputs are fed to a *voter* whose output represents the majority of its $2n+1$ input signals. The investigation is based on the previously described procedure for series/parallel structures, see for example the case of $n = 1$ (active redundancy 2-out-of-3 in series with the voter E_v) given in the sixth line of Table 2.1. The majority redundancy realizes in a very simple way a *fault-tolerant structure without the need for control or switching elements*, the required function is performed without interruption until a second failure occurs, the first failure will automatically be masked by the majority redundancy. The *voter* for a majority redundancy with $n = 1$ consists of three two-input NAND and one three-input NAND gate (per bit). An *alarm circuit* can be implemented with three two-input EXOR and one three-input NAND gate, see Example 2.5.

Example 2.5

Realize a majority redundancy for $n = 1$ inclusive voter and alarm signal at the first failure of a redundant element..

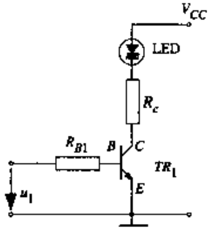
Solution

Using the same notation as for Eq. (2.14), the 2-out-of-3 active redundancy can be implemented by $(e_1 \cap e_2) \cup (e_1 \cap e_3) \cup (e_2 \cap e_3)$. With this, the functional block diagram per bit of the voter for a majority redundancy with $n = 1$ is simply obtained as the realization of the logic equation related to the above expression. The alarm circuit giving a logic 1 at the occurrence of the first failure is also easy to implement.



Example 2.6

Compute the predicted reliability of the following circuit, for which the required function asks that the LED must light when the control voltage u_1 is high and the environmental conditions correspond to G_B in Table 2.3, with ambient temperature $\theta_A = 50°C$ inside the equipment and 30°C at the location of the LED, quality factor $\pi_Q = 1$ as per Table 2.4.



u_1 : 0.1 V and 4 V
V_{CC} : 5 V
LED : 1 V at 20 mA, $I_{max} = 100$ mA
R_C : 150 Ω, 1/2 W, MF
TR_1 : Si, 0.3 W, 30 V, β > 100, plastic
R_{B1} : 10 kΩ, 1/2 W, MF

Solution

The solution is based on the procedure given in Fig 2.1.

1. The transistor works as an electronic switch with $I_C \approx 20\,\text{mA}$ and $I_B \approx 0.33\,\text{mA}$ in the on state (saturated), the off state is assured by $u_1 = 0.1\,\text{V}$, the required functiont can be fulfilled.
2. Since all elements are involved in the required function, the reliability block diagram consists of the series the connection of the five items E_1 to E_5, where E_5 represents the printed circuit with soldering joints.

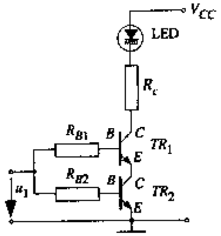
$E_1 \triangleq$ LED, $E_2 \triangleq R_C$, $E_3 \triangleq R_{B1}$, $E_4 \triangleq TR_1$
$E_5 \triangleq$ PCB and solder joints

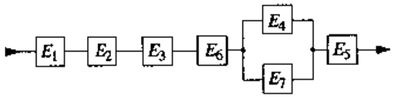
3. The stress factor of each element can be easily determined from the circuit and the given rated values. A stress factor 0.1 is assumed for all elements when the transistor is off. When the transistor is on, the stress factor is 0.2 for the diode and about 0.1 or less for all other elements. The ambient temperature is 30°C for the LED and 50°C for the remaining elements.
4. The failure rates of the individual elements is determined (approximately) with data from Section 2.2.4 (Example 2.4, Figs. 2.4 to 2.6, Tables 2.3 to 2.4 with $\pi_E = \pi_Q = 1$). Thus,

LED : $\lambda_1 \approx 1.3 \cdot 10^{-9}\,\text{h}^{-1}$
Transistor : $\lambda_4 \approx 3 \cdot 10^{-9}\,\text{h}^{-1}$
Resistor : $\lambda_2 = \lambda_3 \approx 0.3 \cdot 10^{-9}\,\text{h}^{-1}$,

when the transistor is on. For the printed circuit board and soldering joints, $\lambda_5 = 1 \cdot 10^{-9}\,\mathrm{h}^{-1}$ is assumed. The above values for λ remain practically unchanged when the transistor is off due to the low stress factors (the stress factor in the off state was set at 0.1).

5. Based on the results of Step 4, the reliability function of each element can be determined as $\mathrm{R}_i(t) = e^{-\lambda_i t}$
6. The reliability function $\mathrm{R}_S(t)$ for the whole circuit can now be computed. Equation (2.17) yields $\mathrm{R}_S(t) = e^{-5.9 \cdot 10^{-9} t}$. For 10 years of continuous operation, for example, the predicted reliability of the circuit is > 0.999.
7. *Supplementary result:* Assume now, to discuss this example further, that the failure rate of the transistor is too high (e.g. for safety reasons) and that no transistor of better quality can be obtained. Then, redundancy should be implemented for this element. Assuming the *failure mode* for this transistor to be a short between emitter and collector, the resulting circuit and the corresponding reliability block diagram are





E_1 to E_5 as in point 2
$E_6 \triangleq R_{B2} \triangleq R_{B1}$, $E_7 \triangleq TR_2 \triangleq TR_1$

Due to the very small stress factor, computation of the individual element failure rates yields the same values as without redundancy. Thus, for the reliability function of the circuit one obtains (assuming independent elements)

$$\mathrm{R}_S(t) = e^{-3.2 \cdot 10^{-9} t}\,(2\,e^{-3 \cdot 10^{-9} t} - e^{-6 \cdot 10^{-9} t}),$$

from which it follows that

$$\mathrm{R}_S(t) \approx e^{-3.2 \cdot 10^{-9} t} \qquad \text{for} \quad t \le 10^6\,\mathrm{h}.$$

Circuit reliability is then practically no longer influenced by the transistor. This confirms the discussion made with Fig. 2.7 for $\lambda t << 1$. If the *failure mode* of the transistors were an open between collector and emitter, both elements E_4 and E_7 would appear in series in the reliability block diagram, redundancy would be a *disadvantage* in this case.

2.2.7 Part Count Method

In an early phase of a project, for logistical needs, or in some well defined simple applications, a rough estimate of the *predicted reliability* is often required. For such an analysis, it is assumed that the item under consideration is *without redundancy* (series structure as in Section 2.2.6.1) and the computation of the failure rate at component level is made either using *field data* or by considering the technology, environmental, and quality factor only. The procedure is known as *part count method* and differs basically from the *part stress method* introduced in Section 2.2.4. Its advantage is the great simplicity, but its usefulness is often limited to specific applications.

2.3 Reliability of Systems with Complex Structures

Complex structures arise in many applications, for example in telecommunications, defense, and aerospace equipment and systems. In the context of this book, a structure is *complex* when the reliability block diagram either cannot be reduced to a *series/parallel structure* with *independent* elements or does *not exist*. A reliability block diagram does not exist, for instance, if more than two states (good/failed) or one failure mode (e.g. short or open) must be considered for an element. The reduction of a reliability block diagram to a series/parallel structure with independent elements is generally not possible with distributed (meshed) structures, and when elements appear in the diagram more than once.

Analysis of complex structures can become difficult and time consuming. However, methods are well developed should the reliability block diagram *exist* and the item satisfy the following requirements:

1. Only active (parallel) redundancy is considered.
2. Some elements can appear more than once in the reliability block diagram, but all different elements are independent.
3. On/off operations are either 100% reliable, or their effect has been considered in the reliability block diagram according to the above restrictions.

Under these assumptions, analysis can be done using Boolean models. Important special cases can also be investigated with simple heuristically-oriented methods. *Heuristic methods* will be introduced in Sections 2.3.1 to 2.3.3, *Boolean models* in Section 2.3.4, items with *load sharing* or *warm redundancies* in Section 2.3.5, and elements with *two failure modes* in Section 2.3.6. Techniques based on stress/strength analysis, for mechanical parts and drift failures, are discussed in Section 2.5. For a computer aided investigation of complex systems one can refer to Section 6.8.2. Computer aided reliability analyses are discussed in Section 6.8.2.

2.3.1 Key Item Method

The *key item method* is based on the theorem of *total probability* (Eq. (A6.17)). The event

the item operates failure free in the interval $(0, t]$

or in a short form

system *up* in $(0, t]$

can be split into the following two complementary events

Element E_i *up* in $(0, t] \cap$ system *up* in $(0, t]$

and

Element E_i fails in $(0, t] \cap$ system *up* in $(0, t]$.

From this it follows that, for the *reliability function* $\mathrm{R}_S(t)$

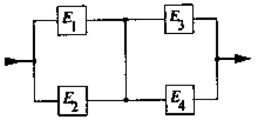
$$\begin{aligned}\mathrm{R}_S(t) &= \mathrm{R}_i(t)\,\Pr\{\text{system } up \text{ in } (0,t] \mid E_i\ up \text{ in } (0,t]\}\\ &\quad + (1-\mathrm{R}_i(t))\,\Pr\{\text{system } up \text{ in } (0,t] \mid E_i \text{ failed in } (0,t]\}.\end{aligned} \tag{2.29}$$

Where $\mathrm{R}_i(t) = \Pr\{E_i\ up \text{ in } (0,t]\} = \Pr\{e_i\}$ as in Eq. (2.13). The element E_i must be chosen in such a way that a series/parallel structure is obtained for the reliability block diagrams conditioned by the events $\{E_i\ up$ in $(0, t]\}$ and $\{E_i$ failed in $(0, t]\}$. Successive application of Eq. (2.29) is also possible (Examples 2.9 and 2.15). Sections 2.3.1.1 and 2.3.1.2 present two typical situations.

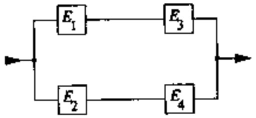
2.3.1.1 Bridge Structures

The reliability block diagram of a *bridge structure* with a bidirectional connection is shown in Fig. 2.10. Element E_5 can work with respect to the required function in *both directions*, from E_1 via E_5 to E_4 and from E_2 via E_5 to E_3. It is therefore in a *key position* (key element).

This property is used to compute the reliability function by means of Eq. (2.29) with $E_i = E_5$. For the conditional probabilities in Eq. (2.29), the corresponding reliability block diagrams are derived from Fig. 2.10.



E_5 did not fail in $(0, t]$



E_5 failed in $(0, t]$

From Eq. (2.29), it follows that

$$R_S = R_5(R_1+R_2-R_1R_2)(R_3+R_4-R_3R_4)+(1-R_5)(R_1R_3+R_2R_4-R_1R_2R_3R_4). \tag{2.30}$$

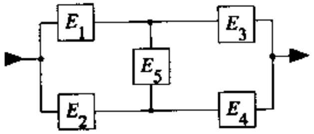
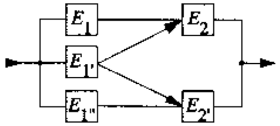


Figure 2.10 Reliability block diagram of a bridge circuit

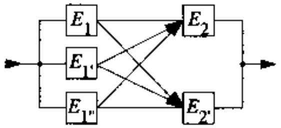
The same considerations apply to the bridge structure with a directed connection (here, i must be different from 5 and $E_i = E_4$ has been used in Table 2.1).

Example 2.7

Calculate the reliability of the item according to case a) below. How much would the reliability be improved if the structure were be modified according to case b)? (Assumptions: nonrepairable until system failure, active redundancy, independent elements, $\mathrm{R}_{E1}(t) = \mathrm{R}_{E1'}(t) = \mathrm{R}_{E1''}(t) = \mathrm{R}_1(t)$ and $\mathrm{R}_{E2}(t) = \mathrm{R}_{E2'}(t) = \mathrm{R}_2(t)$).



Case a)



Case b)

Solution

Element $E_{1'}$ is in a key position in case a). Thus, similarly to Eq. (2.30), one obtains $R_a = R_1(2R_2 - R_2^2) + (1-R_1)(2R_1R_2 - R_1^2R_2^2)$ with $R_a = \mathrm{R}_a(t)$, $R_1 = \mathrm{R}_1(t)$ and $R_2 = \mathrm{R}_2(t)$. Case b) represents a series connection of a 1-out-of-3 redundancy with a 1-out-of-2 redundancy. From Sections 2.2.6.3 and 2.2.6.4 it follows that $R_b = R_1R_2(3 - 3R_1 + R_1^2)(2 - R_2)$, with $R_b = \mathrm{R}_b(t)$, $R_1 = \mathrm{R}_1(t)$ and $R_2 = \mathrm{R}_2(t)$. From this,

$$R_b - R_a = 2R_1R_2(1-R_2)(1-R_1)^2. \tag{2.31}$$

The difference $R_b - R_a$ reaches as maximum the value $2/27$ for $R_1 = 1/3$ and $R_2 = 1/2$, i.e. $R_b = 57/108$ and $R_a = 49/108$; the advantage of case b) is thus small, at least as far as the reliability figure is concerned.

2.3.1.2 Reliability Block Diagram in Which at Least One Element Appears More than Once

In practice, situations often occur in which an element appears more than once in the reliability block diagram, although there is only one such element in the item considered. These situations can be investigated with the *key item method* introduced in Section 2.3.1.1. Examples 2.8 and 2.9 present two concrete cases.

Example 2.8

Give the reliability for the equipment introduced in Example 2.2.

Solution

In the reliability block diagram of Example 2.2, element E_2 is in a key position. Similarly to Eq. (2.30) it follows that

$$R_S = R_2R_1(R_4 + R_5 - R_4R_5) + (1-R_2)R_1R_3R_5, \tag{2.32}$$

with $R_S = \mathrm{R}_S(t)$ and $R_i = \mathrm{R}_i(t)$, $i = 1, \ldots, 5$.

Example 2.9

Calculate the reliability for the redundant circuit of Example 2.3.

Solution

In the reliability block diagram of Example 2.3, U_1 and U_2 are in a key position. Using the method introduced in Section 2.3.1 successively on U_1 and U_2, i.e. on E_5 and E_6, yields.

$$R_S = R_9\{R_5[R_6(R_1R_7 + R_4R_8 - R_1R_4R_7R_8)(R_2 + R_3 - R_2R_3) + (1 - R_6)R_1R_2R_7] + (1 - R_5)R_3R_4R_6R_8\}.$$

With

$$R_1 = R_2 = R_3 = R_4 = R_D, \quad R_5 = R_6 = R_U, \quad R_7 = R_8 = R_I, \quad \text{and } R_9 = R_{II}$$

it follows that

$$R_S = R_U\, R_{\text{II}}\, [R_U(2\,R_D\,R_{\text{I}} - R_D^2\,R_{\text{I}}^2)(2\,R_D - R_D^2) + 2(1 - R_U)\,R_D^2\,R_{\text{I}}], \tag{2.33}$$

with $R_S = \text{R}_S(t)$, $R_U = \text{R}_U(t)$, $R_D = \text{R}_D(t)$, $R_I = \text{R}_I(t)$ and $R_{II} = \text{R}_{II}(t)$.

2.3.2 Successful Path Method

In this and in the next section, two general, closely related methods are introduced. For simplicity, the presentation will be based on the reliability block diagram given in Fig. 2.11. As in Section 2.2.6.1, e_i stands for the *event*

element E_i *up* in the interval $(0, t]$,

hence $\Pr\{e_i\} = \text{R}_i(t)$ and $\Pr\{\bar{e}_i\} = 1 - \text{R}_i(t)$. The *successful path method* is based on the following concept:

> *An item fulfills its required function if there is at least one path between the input and the output upon which all elements perform their required function.*

Paths must lead from left to right and may not contain any loops. Only the given direction is possible along a directed connection. The following successful paths exist in the reliability block diagram of Fig. 2.11

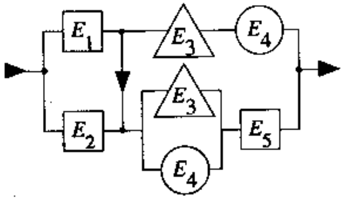


Figure 2.11 Reliability block diagram of a complex structure (elements E_3 and E_4 appear each twice in the RBD, the directed connection has reliability 1)

$$e_1 \cap e_3 \cap e_4, \quad e_1 \cap e_3 \cap e_5, \quad e_1 \cap e_4 \cap e_5, \quad e_2 \cap e_3 \cap e_5, \quad e_2 \cap e_4 \cap e_5.$$

Consequently it follows that

$$\begin{aligned} R_S(t) = \Pr\{&(e_1 \cap e_3 \cap e_4) \cup (e_1 \cap e_3 \cap e_5) \cup (e_1 \cap e_4 \cap e_5) \\ &\cup (e_2 \cap e_3 \cap e_5) \cup (e_2 \cap e_4 \cap e_5)\}. \end{aligned} \tag{2.34}$$

Using the addition theorem of probability theory (Eq. (A6.17)), Eq. (2.34) leads to

$$\begin{aligned} R_S = {}& R_1\,R_3\,R_4 + R_1\,R_3\,R_5 + R_1\,R_4\,R_5 + R_2\,R_3\,R_5 + R_2\,R_4\,R_5 - 2\,R_1\,R_3\,R_4\,R_5 \\ & - R_1\,R_2\,R_3\,R_5 - R_1\,R_2\,R_4\,R_5 - R_2\,R_3\,R_4\,R_5 + R_1\,R_2\,R_3\,R_4\,R_5, \end{aligned} \tag{2.35}$$

with $R_S = \text{R}_S(t)$ and $R_i = \text{R}_i(t)$, $i = 1, \ldots, 5$.

2.3.3 State Space Method

This method is based on the following concept:

> *Every element E_i is assigned an indicator $\zeta_i(t)$ with the following property: $\zeta_i(t) = 1$ as long as E_i does not fail, and $\zeta_i(t) = 0$ if E_i has failed. The vector with components $\zeta_i(t)$ determines the item state at time t. Since each element in the interval $(0, t]$ functions or fails independently of the others, 2^n states are possible for an item with n elements. After listing the 2^n possible states at time t, all those states are determined in which the item performs the required function. The probability that the item is in one of these states is the reliability function $\text{R}_S(t)$ of the item considered.*

The 2^n possible conditions at time t for the reliability block diagram of Fig. 2.11 are

E_1	1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0
E_2	1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0
E_3	1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0
E_4	1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
E_5	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
S	1 1 1 0 1 1 1 0 1 1 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0

A "1" in this table means that the element or item considered has not failed before t. For the Example of Fig. 2.11, the event

item *up* in the interval $(0, t]$

is equivalent to the event

$$
\begin{aligned}
&(e_1 \cap e_2 \cap e_3 \cap e_4 \cap e_5) \cup (\bar{e}_1 \cap e_2 \cap e_3 \cap e_4 \cap e_5) \cup (e_1 \cap \bar{e}_2 \cap e_3 \cap e_4 \cap e_5) \\
&\cup (e_1 \cap e_2 \cap \bar{e}_3 \cap e_4 \cap e_5) \cup (\bar{e}_1 \cap e_2 \cap \bar{e}_3 \cap e_4 \cap e_5) \cup (e_1 \cap \bar{e}_2 \cap \bar{e}_3 \cap e_4 \cap e_5) \\
&\cup (e_1 \cap e_2 \cap e_3 \cap \bar{e}_4 \cap e_5) \cup (\bar{e}_1 \cap e_2 \cap e_3 \cap \bar{e}_4 \cap e_5) \cup (e_1 \cap \bar{e}_2 \cap e_3 \cap \bar{e}_4 \cap e_5) \\
&\cup (e_1 \cap e_2 \cap e_3 \cap e_4 \cap \bar{e}_5) \cup (e_1 \cap \bar{e}_2 \cap e_3 \cap e_4 \cap \bar{e}_5).
\end{aligned}
$$

After appropriate simplification, this reduces to

$$
\begin{aligned}
&(e_2 \cap e_3 \cap e_5) \cup (e_1 \cap e_3 \cap e_4 \cap \bar{e}_5) \cup (e_1 \cap \bar{e}_2 \cap e_3 \cap \bar{e}_4 \cap e_5) \\
&\cup (e_1 \cap \bar{e}_2 \cap e_4 \cap e_5) \cup (e_2 \cap \bar{e}_3 \cap e_4 \cap e_5),
\end{aligned}
$$

from which

$$
\begin{aligned}
\mathrm{R}_S(t) = \Pr\{&(e_2 \cap e_3 \cap e_5) \cup (e_1 \cap e_3 \cap e_4 \cap \bar{e}_5) \cup (e_1 \cap \bar{e}_2 \cap e_3 \cap \bar{e}_4 \cap e_5) \\
&\cup (e_1 \cap \bar{e}_2 \cap e_4 \cap e_5) \cup (e_2 \cap \bar{e}_3 \cap e_4 \cap e_5)\}.
\end{aligned} \tag{2.36}
$$

Evaluation of Eq. (2.36) leads to Eq. (2.35). In contrast to the successful path method, all events in the state space method (columns in the state space table and terms in Eq. (2.36)) are *mutually exclusive*.

2.3.4 Boolean Function Method

The *Boolean function method* generalizes and formalizes the methods introduced in Sections 2.3.2 and 2.3.3. For this analysis, it is assumed that the item considered is *coherent*, that is, the following holds:

1. The state of the item depends on the states of all of its elements (all elements are relevant and the item is up if all its elements are up).
2. If the item is down, no additional failure of any element can bring it in an up state (monotony).

In the case of *repairable* items, the second property must be extended to: If the item is up, it remains up if any element is repaired. Not every item is necessarily coherent. In the following, *up* denotes *item in an operating state* and *down* denotes *item in a failed state* (being repaired or waiting for repair).

The states of a coherent item can be described by a system function. A *system function* ϕ is a *Boolean function*

$$
\phi = \phi(\zeta_1, \ldots, \zeta_n) = \begin{cases} 1 & \text{for item up} \\ 0 & \text{for item down} \end{cases} \tag{2.37}
$$

of the indicators $\zeta_i = \zeta_i(t)$, defined in Section 2.3.3 ($\zeta_i = 1$ if element E_i is *up* and $\zeta_i = 0$ if element E_i is *down*), for which the following applies (coherent item):

1. ϕ depends on all the variables ζ_i $(i = 1, \ldots, n)$.
2. ϕ is nondecreasing in all variables, $\phi = 0$ for all $\zeta_i = 0$ and $\phi = 1$ for all $\zeta_i = 1$.

Since the indicators ζ_i and the system function ϕ take on only the values 0 and 1,

$$
\mathrm{R}_i(t) = \Pr\{\zeta_i(t) = 1\} = \mathrm{E}[\zeta_i(t)] \tag{2.38}
$$

applies for the *reliability function* of the element E_i, and

$$
\mathrm{R}_S(t) = \mathrm{E}[\phi(\zeta_i, \ldots, \zeta_n)] \tag{2.39}
$$

applies for the *reliability function* of the item, with $\zeta_i = \zeta_i(t)$.

The Boolean function method thus transfers the problem of calculating $\mathrm{R}_S(t)$ to that of the determination of the system function $\phi(\zeta_1, \ldots, \zeta_n)$. Two methods are available for this purpose:

1. *Minimal Path Set* approach: A set $\mathcal{P}_i$ of elements is a *minimal path* if the item is up *only* when $\zeta_j = 1$ for all $E_j \in \mathcal{P}_i$ and $\zeta_k = 0$ for all $E_k \notin \mathcal{P}_i$, but this does not apply for any subset of $\mathcal{P}_i$. The elements E_j within $\mathcal{P}_i$ form a *series model* with system function

   $$
   \phi_{\mathcal{P}_i} = \prod_{E_j \in \mathcal{P}_i} \zeta_j . \tag{2.40}
   $$

   If for a given item there exist *r* minimal paths, then these form an *active 1-out-of-r redundancy*, i.e.

   $$
   \phi = \phi(\zeta_1, \ldots, \zeta_n) = 1 - \prod_{i=1}^{r} (1 - \phi_{\mathcal{P}_i}) = 1 - \prod_{i=1}^{r} (1 - \prod_{E_j \in \mathcal{P}_i} \zeta_j). \tag{2.41}
   $$

2. *Minimal Cut Set* approach: A set C_i is a *minimal cut* if the item is down *only* when $\zeta_j = 0$ for all $E_j \in C_i$ and $\zeta_k = 1$ for all $E_k \notin C_i$, but this does not apply for any subset of C_i. The elements E_j within C_i form a *parallel model* (*active redundancy with* $k = 1$) with system function

   $$
   \phi_{C_i} = 1 - \prod_{E_j \in C_i} (1 - \zeta_j) . \tag{2.42}
   $$

   If for a given item there exist *m* minimal cuts, then these form a *series model*, i.e.

   $$
   \phi = \phi(\zeta_1, \ldots, \zeta_n) = \prod_{i=1}^{m} \phi_{C_i} = \prod_{i=1}^{m} (1 - \prod_{E_j \in C_i} (1 - \zeta_j)). \tag{2.43}
   $$

A series model with elements $E_1, \ldots, E_n$ has one path and *n* cuts, a parallel model (1-out-of-*n*) with elements $E_1, \ldots, E_n$ has one cut and *n* paths. Because of Eqs. (2.41) and (2.43) it holds that

$$
\mathrm{R}_S(t) = \Pr\{\phi_{\mathcal{P}_1} = 1 \cup \ldots \cup \phi_{\mathcal{P}_r} = 1\} = 1 - \Pr\{\phi_{C_1} = 0 \cup \ldots \cup \phi_{C_m} = 0\}. \tag{2.44}
$$

Considering now that for arbitrary events $A_1, \ldots, A_n$ one has (Appendix A6.4.4)

$$\sum_{i=1}^{n} \Pr\{A_i\} - \frac{1}{2} \sum_{i \neq j} \Pr\{A_i \cap A_j\} \leq \Pr\{A_1 \cup \ldots \cup A_n\} \leq \sum_{i=1}^{n} \Pr\{A_i\}, \tag{2.45}$$

it can be shown (Example 2.10) that

$$1 - \sum_{i=1}^{m} \Pr\{\phi_{C_i} = 0)\} \leq \mathrm{R}_S(t) \leq 1 - (1 - \frac{(m-1)q}{2}) \sum_{i=1}^{m} \Pr\{\phi_{C_i} = 0)\}, \tag{2.46}$$

with $q = \max\{1 - \mathrm{R}_1(t), \ldots, 1 - \mathrm{R}_n(t)\}$. Equation (2.46) is a useful *approximation* for many practical applications, as an alternative to that discussed with Eq. (A6.16).

Example 2.10

Prove Eqs. (2.45) and (2.46).

Solution

For arbitrary events $A_1, \ldots, A_n$ it holds from the two sided inequality $S_1 - S_2 \leq S \leq S_1$ as per Eq. (A6.16) that

$$\sum_{i=1}^{n} \Pr\{A_i\} \; - \sum_{1 \leq i < j \leq n}^{n} \Pr\{A_i \cap A_j\} \leq \Pr\{A_1 \cup \ldots \cup A_n\} \leq \sum_{i=1}^{n} \Pr\{A_i\}.$$

Equation (2.45) follows then by considering

$$\sum_{1 \leq i < j \leq n} \Pr\{A_i \cap A_j\} = \frac{1}{2} \sum_{i \neq j} \Pr\{A_i \cap A_j\}.$$

Define now as A_i the event {all elements $E_k \in C_i$ are down}, $i = 1, \ldots, m$, assume that all elements are independent, set as before $R_i = \Pr\{\text{Element } E_i \text{ } \textit{up} \text{ at time } t\}$, and let $q = \max\{1 - R_1, \ldots, 1 - R_n\}$, with $R_i = \mathrm{R}_i(t)$. For fixed i and j ($i \neq j$) assume further that there is at least one k $(1 \leq k \leq n)$ for which $E_k \in C_j$ and $E_k \notin C_i$, then $\Pr\{A_j \mid A_i\} = \Pr\{A_j \cap A_i\} / \Pr\{A_i\} \leq \Pr\{E_k \textit{ down} \text{ at } t\} / \Pr\{A_i\} \leq \Pr\{E_k \textit{ down} \text{ at } t\} \leq q$ and thus

$$\Pr\{A_i \cap A_j\} = \Pr\{A_i\} \Pr\{A_j \mid A_i\} \leq \Pr\{A_i\} q.$$

Summation over $i < j$ leads to

$$\sum_{1 \leq i < j \leq m} \Pr\{A_i \cap A_j\} = \frac{1}{2} \sum_{i \neq j} \Pr\{A_i \cap A_j\} = \frac{1}{2} \sum_{i=1}^{m} \Pr\{A_i\} \sum_{j \neq i} \Pr\{A_j \mid A_i\} \leq \frac{(m-1)}{2} q \sum_{i=1}^{m} \Pr\{A_i\}.$$

From Eq. (2.45) it follows then

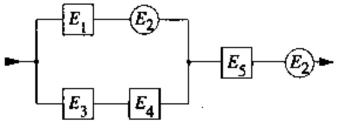
$$\sum_{i=1}^{m} \Pr\{A_i\} \geq \Pr\{A_1 \cup \ldots \cup A_m\} \geq \sum_{i=1}^{m} \Pr\{A_i\} - \frac{(m-1)}{2} q \sum_{i=1}^{m} \Pr\{A_i\},$$

and thus Eq. (2.46) by considering $\mathrm{R}_S(t) = 1 - \Pr\{A_1 \cup \ldots \cup A_m\}$ as per Eq. (2.44).

The paths according to Eq. (2.34) constitute a minimal path set for the reliability block diagram of Fig. 2.11. Using Eq. (2.41) this would lead to the *system function* $\phi(\zeta_1, \ldots, \zeta_n) = 1 - (1 - \zeta_1 \zeta_3 \zeta_4)(1 - \zeta_1 \zeta_3 \zeta_5)(1 - \zeta_1 \zeta_4 \zeta_5)(1 - \zeta_2 \zeta_3 \zeta_5)(1 - \zeta_2 \zeta_4 \zeta_5)$. The investigation of the block diagram of Fig. 2.11 by the method based on minimal cut sets is much more laborious. Obviously, minimal path sets and minimal cut sets deliver the same system function $\phi(\zeta_1, \ldots, \zeta_n)$ with different effort depending on the structure of the reliability block diagram considered (structures with many series elements can be treated easily by the minimal path set method, see Example 2.11).

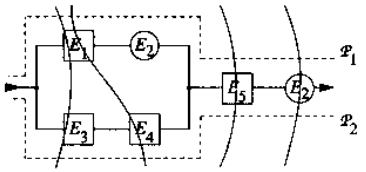
Example 2.11

Give the system function according to the minimal path set and the minimal cut set approach for the following reliability block diagram, and calculate the reliability function assuming independent elements.



Solution

For the above reliability block diagram, there exist 2 minimal path sets $\mathcal{P}_1$, $\mathcal{P}_2$ and 4 minimal cut sets $C_1, \ldots, C_4$.



The system function is then given by (minimal path sets, Eq. (2.41) with $\zeta_i \zeta_i = \zeta_i$, $\zeta_i \zeta_j = \zeta_j \zeta_i$)

$$\phi(\zeta_1, \ldots, \zeta_5) = 1 - (1 - \zeta_1 \zeta_2 \zeta_5)(1 - \zeta_2 \zeta_3 \zeta_4 \zeta_5) = \zeta_1 \zeta_2 \zeta_5 + \zeta_2 \zeta_3 \zeta_4 \zeta_5 - \zeta_1 \zeta_2 \zeta_3 \zeta_4 \zeta_5$$

or by (minimal cut sets, Eq. (2.43) with $\zeta_i \zeta_i = \zeta_i$, $\zeta_i \zeta_j = \zeta_j \zeta_i$)

$$\begin{aligned} \phi(\zeta_1, \ldots, \zeta_5) &= [1 - (1 - \zeta_1)(1 - \zeta_3)][1 - (1 - \zeta_1)(1 - \zeta_4)][1 - (1 - \zeta_5)][1 - (1 - \zeta_2)] \\ &= (\zeta_1 + \zeta_3 - \zeta_1 \zeta_3)(\zeta_1 + \zeta_4 - \zeta_1 \zeta_4) \zeta_2 \zeta_5 \\ &= \zeta_1 \zeta_2 \zeta_5 + \zeta_2 \zeta_3 \zeta_4 \zeta_5 - \zeta_1 \zeta_2 \zeta_3 \zeta_4 \zeta_5. \end{aligned}$$

Assuming independence for the (different) elements, it follows for reliability function (in both cases)

$$R_S = R_1 R_2 R_5 + R_2 R_3 R_4 R_5 - R_1 R_2 R_3 R_4 R_5$$

with $R_S = \mathrm{R}_S(t)$ and $R_i = \mathrm{R}_i(t)$, $i = 1, \ldots, 5$. Computation with the key item method leads directly to $R_S = R_2 (R_1 + R_3 R_4 - R_1 R_3 R_4) R_5 + (1 - R_2) \cdot 0 = R_2 (R_1 + R_3 R_4 - R_1 R_3 R_4) R_5$.

In the case of elements which are *independent* and *nonrepairable* until system failure, the item's reliability function $\mathrm{R}_S(t) = \mathrm{E}[\phi(\zeta_1, \ldots, \zeta_n)]$ can often be obtained directly from the system function by considering in Eqs. (2.41) and (2.43) the *idempotency* property ($\zeta_i \zeta_i = \zeta_i$) and substituting $\mathrm{R}_i(t)$ for ζ_i (Eq. (A6.69)). A further possibility is to use the *disjunctive normal form* $\phi_D(\zeta_1, \ldots, \zeta_n)$ or its equivalent *linear form* $\phi_L(\zeta_1, \ldots, \zeta_n)$ of the system function $\phi(\zeta_1, \ldots, \zeta_n)$, yielding, for *coherent items* (systems) with *independent elements* [6.10, 2.43, 2.44],

$$\mathrm{R}_S(t) = \phi_D(R_1, \ldots, R_n) = \phi_L(R_1, \ldots, R_n), \tag{2.47}$$

with $R_i = \mathrm{R}_i(t)$, $i = 1, \ldots, n$.

For *coherent repairable items* with *independent elements*, i.e. if every element works and is repaired independently from every other element (one repair crew per element), Eq. (2.39) or Eq. (2.47) can be used to compute the *point availability* $\mathrm{PA}_S(t)$ of the item

$$\mathrm{PA}_S(t) = \phi_D(PA_1, \ldots, PA_n) = \phi_L(PA_1, \ldots, PA_n), \tag{2.48}$$

with $PA_i = \mathrm{PA}_i(t)$ according to Eq. (6.44) for element E_i. However, in practical applications it is not usual to have a repair crew for each element in the reliability block diagram of a system. Nevertheless, Eq. (2.48) can be used as an *approximation* (upper bound) for $\mathrm{PA}_S(t)$. For *repairable* elements the *indicators* $\zeta_i(t)$ as in Eq. (2.37) are defined as $\zeta_i(t) = 1$ if element E_i is up (operating) and $\zeta_i(t) = 0$ if element E_i is in repair. Numerical computations are generally easier when performed for the *unavailability* $1 - \mathrm{PA}_S(t)$.

2.3.5 Parallel Models with Constant Failure Rates and Load Sharing

In the redundancy structures investigated in the previous sections, all elements are operating under the same conditions. For this type of redundancy, denoted as *active* (parallel) redundancy, the assumed statistical independence of the elements implies in particular that there is no *load sharing*. This assumption does not arise in many practical applications, for example, at component level or in the presence of power elements. The investigation of the reliability function in the case of load sharing or of other kinds of dependency involves the use of *stochastic processes*. The situation is simple if one can assume that the failure rate of each element changes *only* when a failure occurs. In this case, the general model for a *k-out-of-n redundancy* is given in Fig. 2.12 by a *death process*. $Z_0, \ldots, Z_{n-k+1}$ are the states of the process. In state Z_i, exactly i elements are down. At state Z_{n-k+1} the system is down. Assuming

$$\lambda = \text{failure rate of an element in the } \textit{operating state} \tag{2.49}$$

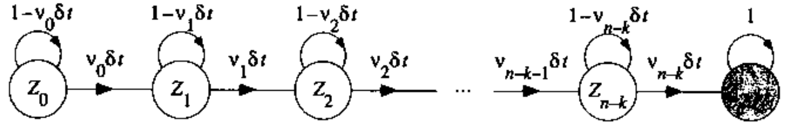


Figure 2.12 Diagram of the transition probabilities in $(t, t+\delta t]$ of a k-out-of-n redundancy (nonrepairable until system failure, constant failure rates *during the sojourn time in each state* (not necessarily at a state change, e.g. because of load sharing), t arbitrary, $\delta t \to 0$, Markov process)

and

$$\lambda_r = \text{failure rate of an element in the } \textit{reserve state} \quad (\lambda_r \le \lambda), \tag{2.50}$$

the model of Fig. 2.12 considers in particular the following cases:

1. Active redundancy without load sharing

$$\nu_i = (n-i)\lambda, \qquad i = 0, \ldots, n-k, \tag{2.51}$$

   λ is the same for all states.

2. Active redundancy with *load sharing* ($\lambda = \lambda(i)$)

$$\nu_i = (n-i)\lambda(i), \qquad i = 0, \ldots, n-k, \tag{2.52}$$

   $\lambda(i)$ increases at each state change.

3. Warm (lightly loaded) redundancy

$$\nu_i = k\lambda + (n-k-i)\lambda_r, \qquad i = 0, \ldots, n-k, \tag{2.53}$$

   λ and λ_r are the same for all states.

4. Standby (cold) redundancy

$$\nu_i = k\lambda, \qquad i = 0, \ldots, n-k, \tag{2.54}$$

   λ is the same for all states.

For a *standby redundancy* one assumes that the failure rate in the reserve state is $\equiv 0$ (the reserve elements are switched on when needed). *Warm redundancy* is somewhere between active and standby ($0 < \lambda_r < \lambda$). It should be noted that the *k*-out-of-*n* active, warm, or standby redundancies are only the *simplest* representatives of the general concept of redundancy (series/parallel structures, voting techniques, bridges, and more complex structures are frequently used; redundancy can also appear in the coding, at the software level, or in other forms). Furthermore, the benefit of redundancy can be limited by the involved *failure modes*, as well as by the *switching elements*. For the analysis of the model shown in Fig. 2.12, let

$$P_i(t) = \Pr\{\text{the process is in state } Z_i \text{ at time } t\} \tag{2.55}$$

be the *state probabilities* ($i = 0, \dots, n-k+1$). $P_i(t)$ is obtained by considering the process at two adjacent time points t and $t+\delta t$ and by making use of the *memoryless property* resulting from the *constant* failure rate assumed *between consecutive state changes* (Appendix A7.5). The function $P_i(t)$ thus satisfies the following *difference equation*

$$P_i(t+\delta t) = P_i(t)(1-\nu_i\,\delta t) + P_{i-1}(t)\,\nu_{i-1}\,\delta t, \qquad i = 1, \dots, n-k. \tag{2.56}$$

For $\delta t \to 0$, there follows then a system of differential equations describing the *death process*

$$\begin{aligned} \dot P_0(t) &= -\nu_0\, P_0(t) \\ \dot P_i(t) &= -\nu_i\, P_i(t) + \nu_{i-1}\, P_{i-1}(t), \qquad i = 1, \dots, n-k, \\ \dot P_{n-k+1}(t) &= \nu_{n-k}\, P_{n-k}(t). \end{aligned} \tag{2.57}$$

Assuming, for instance, the initial conditions $P_i(0) = 1$ at $t = 0$, the solution (generally obtained using the Laplace transform) leads to $P_i(t)$, $i = 0, \dots, n-k+1$. Knowing $P_i(t)$, one can evaluate the *reliability function* $R_S(t)$

$$R_S(t) = \sum_{i=0}^{n-k} P_i(t) = 1 - P_{n-k+1}(t) \tag{2.58}$$

and the *mean time to failure* from Eq. (2.8). Assuming $P_0(0) = 1$ as initial condition, one obtains for the Laplace transform of $R_S(t)$, defined as

$$\tilde R_S(s) = \int_0^\infty R_S(t)\, e^{-st}\, dt, \tag{2.59}$$

the expression

$$\tilde R_S(s) = \frac{(s+\nu_0)\dots(s+\nu_{n-k}) - \nu_0 \dots \nu_{n-k}}{s(s+\nu_0)\dots(s+\nu_{n-k})}. \tag{2.60}$$

The *mean time to failure* follows then from

$$MTTF_S = \tilde R_S(0) \tag{2.61}$$

and leads to

$$MTTF_S = \sum_{i=0}^{n-k} \frac{1}{\nu_i}. \tag{2.62}$$

For a *k-out-of-n standby redundancy* with ν_i given by Eq. (2.54), one obtains in particular

$$R_S(t) = \sum_{i=0}^{n-k} \frac{(k\lambda t)^i}{i!} e^{-k\lambda t} \tag{2.63}$$

and

$$MTTF_S = \frac{n-k+1}{k\lambda}. \tag{2.64}$$

Equation (2.63) shows the relation existing between the *Poisson distribution* and the occurrence of *exponentially* distributed events.

For the case of a *k*-out-of-*n active redundancy* without load sharing, it follows from Eqs. (2.62) and (2.51) that

$$MTTF_S = \frac{1}{\lambda}\left(\frac{1}{k} + \dots + \frac{1}{n}\right), \tag{2.65}$$

see also Table 6.8 with $MTTF_{S0} = MTTF_S$, $\mu = 0$, and $\lambda_r = \lambda$. Expressions for $R_S(t)$ are given in Fig. 2.7 for some different values of n and k.

2.3.6 Elements With More Than One Failure Mode

In the previous sections, it was assumed that each element exhibits only one *failure mode*, for example a short or an open. In reality, many components can fail in different ways (Table 3.5). This section will consider as an example the case of a diode. Let

$$\begin{aligned} R(t) &= \Pr\{\text{no failure in } (0, t]\} \\ \overline{R}(t) &= 1 - R(t) = \Pr\{\text{failure in } (0, t]\} \\ \overline{R}_U(t) &= \Pr\{\text{open in } (0, t]\} \\ \overline{R}_K(t) &= \Pr\{\text{short in } (0, t]\}. \end{aligned}$$

Obviously (Example 2.12)

$$1 - R(t) = \overline{R}(t) = \overline{R}_U(t) + \overline{R}_K(t). \tag{2.66}$$

Example 2.12

In an accelerated test of 1000 diodes, 100 failures occur, of which 30 are opens and 70 shorts. Give the values $\overline{R}$, $\overline{R}_U$, and $\overline{R}_K$.

Solution

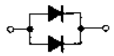
The maximum likelihood estimate of an unknown probability p is, according to Eq. (A8.29) $\hat p = k/n$. Hence, $\overline{R} = 0.1$, $\overline{R}_U = 0.03$ and $\overline{R}_K = 0.07$.

The series connection of two diodes exhibits a circuit failure if either one open or two shorts occur. From this,

$$\overline{R}_S = 1 - (1 - \overline{R}_U)^2 + \overline{R}_K^2 = 2\,\overline{R}_U - \overline{R}_U^2 + \overline{R}_K^2, \qquad (2.67)$$

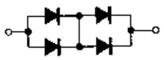
with $R_S = \mathrm{R}_S(t)$, $\overline{R}_K = \overline{\mathrm{R}}_K(t)$ and $\overline{R}_U = \overline{\mathrm{R}}_U(t)$. Similarly, for two diodes in parallel,

$$\overline{R}_S = 2\,\overline{R}_K - \overline{R}_K^2 + \overline{R}_U^2. \qquad (2.68)$$



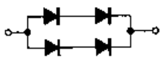
To be *simultaneously* protected against at *least one* failure of *arbitrary mode* (short or open), a *quad redundancy* is necessary. Depending upon whether opens or shorts are more frequent, a quad redundancy with or without a bridge connection is used. For both these cases it follows that (Example 2.13)

$$\overline{R}_S = 2\,\overline{R}_U^2 - \overline{R}_U^4 + (2\,\overline{R}_K - \overline{R}_K^2)^2, \qquad (2.69)$$



and

$$\overline{R}_S = 2\,\overline{R}_K^2 - \overline{R}_K^4 + (2\,\overline{R}_U - \overline{R}_U^2)^2. \qquad (2.70)$$



Equations (2.67) to (2.70) can be obtained using the *state space method* introduced in Section 2.3.3, however with *three states* for every element (good, open (U), and short (K) leading to a *state space* with 3^n elements, see Example 2.13).

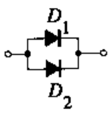
Example 2.13

Using the state space method, give the reliability of two parallel-connected diodes, assuming that opens and shorts are possible.

Solution

Considering the three possible states (good (1), open (U), and short (K)), the state-space for two parallel-connected diodes is

D_1	1	1	1	U	U	U	K	K	K
D_2	1	U	K	1	U	K	1	U	K
S	1	1	0	1	0	0	0	0	0



From the above table,

$$\overline{R}_S = \Pr\{S=0\} = 2R\,\overline{R}_K + \overline{R}_U^2 + 2\,\overline{R}_U\,\overline{R}_K + \overline{R}_K^2$$
$$= 2(1 - \overline{R}_U - \overline{R}_K)\,\overline{R}_K + \overline{R}_U^2 + 2\,\overline{R}_U\,\overline{R}_K + \overline{R}_K^2 = 2\,\overline{R}_K - \overline{R}_K^2 + \overline{R}_U^2.$$

2.3.7 Fault Tolerant Structures

In applications for which high reliability, availability, and/or safety are required, equipment and systems must be *designed* to be *fault tolerant*. This means that autonomously, i.e. without external help, the equipment or system considered should be able to *recognize a fault* (failure or defect) and quickly *reconfigure* itself in such a way as to remain *safe* and possibly also continue to operate with minimal loss in performance.

Methods to investigate *fault tolerant* items have been introduced in Sections 2.2.6.2 through 2.3.6, in particular Sections 2.2.6.5 (*majority redundancy*) and 2.3.6 (*quad redundancy*). The latter is one of the few structures which can support *at least one failure of any mode*, the price paid for this is four devices instead of one. To avoid *common mode* or *single-point failures*, redundant elements should be designed and produced *independently* from each other, in critical cases with different technology, tools, and personnel.

Investigation of all possible *failure modes* during the design of fault tolerant equipment or systems (structures) is mandatory. This is generally done using *fault modes and effects analysis* (FMEA/FMECA), *fault tree analysis* (FTA), *cause-and-effect diagrams* or similar tools, see Section 2.6. Failure mode analysis is important in the presence of *redundancy*, among other to identify the elements which are in series to the ideal redundancy in the reliability block diagram. This also helps to discover *dependencies* and *interactions* between elements of an item and to find appropriate measures to avoid *propagation of faults* (secondary failures).

Protection against *secondary failures* can be often realized with *decoupling elements* such as diodes, resistors, capacitors (diodes D_1 to D_4 in Example 2.3). Other possibilities are the introduction of *standby elements* which are activated only upon failure of active elements, the use of basically different technologies for redundant elements, etc. These and similar methods are used successfully in aerospace and in other equipment and systems with high reliability and/or safety requirements. Quite generally, all parts which are essential for basic functions (clock circuits, bus interfaces, built-in test, monitoring circuits, etc.) have to be selected and designed with care. Adherence to appropriate *design guidelines* is important (Chapter 5). Detection and localization of *hidden failures* (Section 6.10) as well as avoidance of *false alarms* or messages is mandatory. False alarms can also be caused by *synchronization problems* or *glitches*.

Many of the above considerations also apply to *defects*, both in hardware and software, see Sections 5.3.4 and 8.5 for further considerations.

Repairable fault tolerant systems are considered in Chapter 6. Stochastic processes and tools introduced in Appendix A7 and Chapter 6 can also be used to investigate the reliability and availability of *fault tolerant systems* for cases in which a reliability block diagram does not exist (see Section 6.8.1 for some general considerations).

2.4 Reliability Allocation

With complex equipment or systems, it is important to allocate reliability goals at subsystem and assembly levels early in the design phase. Such an allocation motivates the design engineer to consider reliability aspects at all system levels.

Allocation is simple if the item has no redundancy and its components have constant failure rates. The failure rate λ_S of the item is then also constant and equal to the sum of the failure rates of its elements (Eq. (2.17)). In such a case, the allocation of λ_S can be done as follows:

1. Break down the item into elements $E_1, \ldots, E_n$.
2. Define a complexity factor k_i for each element ($0 \le k_i \le 1$, $k_1 + \ldots + k_n = 1$).
3. Determine the *duty cycle* d_i for each element (d_i = operating time of element E_i / operating time of the item).
4. Allocate the item's failure rate λ_S among elements $E_1, \ldots, E_n$ according to

$$\lambda_i = \frac{k_i}{d_i}\lambda_S. \tag{2.71}$$

Should all elements have the same complexity, $k_1 = \ldots = k_n = 1/n$, and the same duty cycle, $d_1 = \ldots = d_n = 1$, then

$$\lambda_i = \frac{1}{n}\lambda_S. \tag{2.72}$$

In addition to the above, costs, technology risks, and failure effects should also be considered. Case-by-case *optimizations* are often possible.

Should the individual element failure rates not be constant and/or the item contain redundancy, allocation of reliability goals is more difficult. The results of Sections 2.2 and 2.3 can be used. If repairable series/parallel structures appear, one can often assume that the failure rate at equipment or system level is fixed by the series elements (Section 6.6), for which Eqs. (2.71) and (2.72) can be used.

2.5 Mechanical Reliability, Drift Failures

As long as the reliability is considered to be the probability R for a mission success (without relation to the distribution of the failure-free operating time), the reliability analysis procedure for mechanical systems is similar to that used for electronic equipment or systems and is based on the following steps:

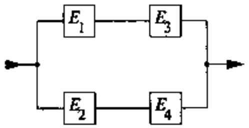
1. Definition of the item and of its associated mission profile.
2. Derivation of the corresponding reliability block diagram.
3. Determination of the reliability for each element of the reliability block diagram.
4. Computation of the item (system) reliability R_S.
5. Elimination of reliability weaknesses and return to step 1 to 2, as necessary.

Such a procedure is established for instance in aerospace applications and is illustrated by Examples 2.14 and 2.15.

Example 2.14

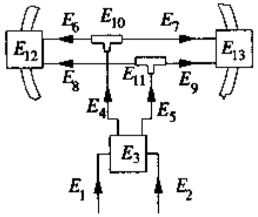
The fastening of two mechanical parts should be easy and reliable. It is done by means of two flanges which are pressed together with 4 clamps E_1 to E_4 placed 90° to each other. Experience has shown that the fastening holds when at least 2 opposing clamps work. Set up the reliability block diagram for this fixation and compute its reliability (each clamp has reliability $R_1 = R_2 = R_3 = R_4 = R$).

Solution



Since at least two opposing clamps (E_1 and E_3 or E_2 and E_4) have to function without failure, the reliability block diagram is obtained as the series connection of E_1 and E_3 in parallel with the series connection of E_2 and E_4, see graph on the right. Under the assumption that clamp is independent from every other one, the item reliability follows from $R_S = 2R^2 - R^4$. *Supplementary result:* If two arbitrary clamps were sufficient for the required function, a 2-out-of-4 active redundancy would apply with $R_S = 6R^2 - 8R^3 + 3R^4$.

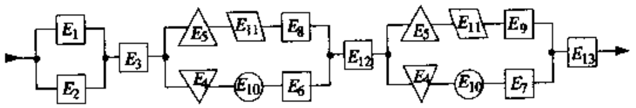
Example 2.15



To separate a satellite's protective shielding, a special electrical-pyrotechnic system described in the block diagram on the right is used. An electrical signal comes through the cables E_1 and E_2 (redundancy) to the electrical-pyrotechnic converter E_3 which lights the fuses. These carry the pyrotechnic signal to explosive charges for guillotining bolts E_{12} and E_{13} of the tensioning belt. The charges can be ignited from two sides, although one ignition will suffice (redundancy). For fulfillment of the required function, both bolts must be exploded simultaneously. Calculate the reliability of this separation system as a function of the reliabilities $R_1, \ldots, R_{13}$ of its elements.

Solution

The reliability block diagram is most easily obtained by considering first the ignition of bolts E_{12} and E_{13} separately and then connecting these two parts of the reliability block diagram in series



Elements E_4, E_5, E_{10}, and E_{11} each appear twice in the reliability block diagram. Repeated application of the *key item method* (successively on E_5, E_{11}, E_4, and E_{10}, see Section 2.3.1 and Example 2.9), by assuming that the elements $E_1, \ldots, E_{13}$ are independent, leads to

$$\begin{aligned} R_S &= R_3 R_{12} R_{13} (R_1 + R_2 - R_1 R_2)\{R_5 \langle R_{11}[R_4 \{R_{10} (R_6 + R_8 - R_6 R_8)(R_7 + R_9 - R_7 R_9) \\ &\quad + (1 - R_{10}) R_8 R_9\} + (1 - R4) R_8 R_9] + (1 - R_{11}) R_4 R_6 R_7 R_{10} \rangle + (1 - R_5) R_4 R_6 R_7 R_{10}\} \\ &= R_3 R_{12} R_{13} (R_1 + R_2 - R_1 R_2)\{R_5 R_{11}[R_4 R_{10} (R_6 + R_8 - R_6 R_8)(R_7 + R_9 - R_7 R_9) \\ &\quad + (1 - R_4 R_{10}) R_8 R_9] + (1 - R_5 R_{11}) R_4 R_6 R_7 R_{10}\}. \end{aligned} \tag{2.73}$$

The situation when the reliability function $\mathrm{R}(t)$ or the distribution function of the failure free time $\mathrm{F}(t) = 1 - \mathrm{R}(t)$ is required, is more complicated. For electronic parts, it is possible to operate with the failure rate; models and data are often available. This is generally not the case for mechanical parts. A possible approach is based on the *stress-strength method*. If $\xi_L(t)$ is the *stress* (load) and $\xi_S(t)$ the *strength*, then a *failure occurs* at the time t for which $|\xi_L(t)| > |\xi_S(t)|$ holds for the first time. Often, $\xi_L(t)$ and $\xi_S(t)$ can be considered as deterministic values and the ratio $\xi_S(t)/\xi_L(t)$ is the *safety factor*. In many practical applications, $\xi_L(t)$ and $\xi_S(t)$ are random variables, often stochastic processes. A practical oriented procedure for the reliability analysis of mechanical systems in these cases is:

1. Definition of the system and of its associated mission profile.
2. Formulation of *failure hypotheses* (buckling, bending, etc.) and validation of them with help of a FMEA/FMECA (Section 2.6); failure hypotheses are (like failure rates) often correlated, this dependence must be identified and considered in the calculations.
3. Evaluation of the stresses applied with respect to the critical failure hypotheses.
4. Evaluation of the strength limits by considering also dynamic stresses, notches, surface condition, etc.
5. Computation of the system reliability (Eqs. (2.74) – (2.80)).
6. Elimination of reliability weaknesses and return to step 1 or 2, as necessary.

Reliability calculation in general leads to one of the following situations:

1. One failure hypothesis, stress and strength are > 0: The *reliability function* is given by

$$\mathrm{R}(t) = \Pr\{\xi_S(x) > \xi_L(x), \quad 0 < x \le t\}. \tag{2.74}$$

2. More than one ($n > 1$) failure hypothesis that can be correlated, stresses and strength are > 0: The *reliability function* is given by

$$\begin{aligned} \mathrm{R}(t) = \Pr\{&(\xi_{S_1}(x) > \xi_{L_1}(x)) \cap (\xi_{S_2}(x) > \xi_{L_2}(x)) \cap \ldots \\ &\cap (\xi_{S_n}(x) > \xi_{L_n}(x)), \quad 0 < x \le t\}. \end{aligned} \tag{2.75}$$

Equation (2.75) can take a complicated form, according to the degree of dependence encountered.

The situation is easier where the stress and strength can be assumed to be *independent positive random variables*. In this case, $\Pr\{\xi_S > \xi_L \mid \xi_L = x\} = \Pr\{\xi_S > x\} = 1 - \mathrm{F}_S(x)$ and the theorem of total probability leads to

$$\mathrm{R}(t) = R = \Pr\{\xi_S > \xi_L\} = \int_0^\infty \mathrm{f}_L(x)(1 - \mathrm{F}_S(x))\,dx. \tag{2.76}$$

Examples 2.16 and 2.17 illustrate the use of Eq. (2.76).

Example 2.16

Let the stress ξ_L of a mechanical joint be normally distributed with mean $m_L = 100\,\mathrm{N/mm^2}$ and standard deviation $\sigma_L = 40\,\mathrm{N/mm^2}$. The strength ξ_S is also normally distributed with mean $m_S = 150\,\mathrm{N/mm^2}$ and standard deviation $\sigma_S = 10\,\mathrm{N/mm^2}$. Compute the reliability of the joint.

Solution

Since ξ_L and ξ_S are normally distributed, their difference is also normally distributed (Example A.615). Their mean and standard deviation are $m_S - m_L = 50\,\mathrm{N/mm^2}$ and $\sqrt{\sigma_S^2 + \sigma_L^2} \approx 41\,\mathrm{N/mm^2}$, respectively. The reliability of the joint is then given by (Table A9.1)

$$R = \Pr\{\xi_S > \xi_L\} = \Pr\{\xi_S - \xi_L > 0\} = \frac{1}{41\sqrt{2\pi}} \int_0^\infty e^{-\frac{(x-50)^2}{2 \cdot 41^2}}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-50/41}^\infty e^{-\frac{y^2}{2}}\,dy \approx 0.89.$$

Example 2.17

Let the strength ξ_S of a rod be normally distributed with mean $m_S = 450\,\mathrm{N/mm^2} - 0.01t\,\mathrm{N/mm^2\,h^{-1}}$ and standard deviation $\sigma_S = 25\,\mathrm{N/mm^2} + 0.001t\,\mathrm{N/mm^2\,h^{-1}}$. The stress ξ_L is constant and equal $350\,\mathrm{N/mm^2}$. Calculate the reliability of the rod at $t = 0$ and $t = 10^4\,\mathrm{h}$.

Solution

At $t = 0$, $m_S = 450\,\mathrm{N/mm^2}$ and $\sigma_S = 25\,\mathrm{N/mm^2}$. Thus,

$$R = \Pr\{\xi_S > \xi_L\} = \frac{1}{\sqrt{2\pi}} \int_{\frac{350-450}{25}}^\infty e^{-\frac{y^2}{2}}\,dy \approx 0.99997$$

After 10,000 operating hours, $m_S = 350\,\mathrm{N/mm^2}$ and $\sigma_S = 35\,\mathrm{N/mm^2}$. The reliability is then

$$R = \Pr\{\xi_S > \xi_L\} = \frac{1}{\sqrt{2\pi}} \int_{\frac{350-350}{35}}^\infty e^{-\frac{y^2}{2}}\,dy = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{y^2}{2}}\,dy = 0.5.$$

Equation (2.76) holds for a one-item structure. For a series model, i.e. in particular for the *series* connection of two independent elements one obtains:

1. Same stress ξ_L

$$R_S = \Pr\{\xi_{S_1} > \xi_L \cap \xi_{S_2} > \xi_L\} = \int_0^\infty \mathrm{f}_L(x)(1-\mathrm{F}_{S_1}(x))(1-\mathrm{F}_{S_2}(x))\,dx. \tag{2.77}$$

2. Independent stresses ξ_{L_1} and ξ_{L_2}

$$\begin{aligned} R_S &= \Pr\{\xi_{S_1} > \xi_{L_1} \cap \xi_{S_2} > \xi_{L_2}\} = \Pr\{\xi_{S_1} > \xi_{L_1}\}\Pr\{\xi_{S_2} > \xi_{L_2}\} \\ &= \Big(\int_0^\infty \mathrm{f}_{L_1}(x)(1-\mathrm{F}_{S_1}(x))\,dx\Big)\Big(\int_0^\infty \mathrm{f}_{L_2}(x)(1-\mathrm{F}_{S_2}(x))\,dx\Big) \triangleq R_1\, R_2\,. \end{aligned} \tag{2.78}$$

For a parallel model, i.e. in particular for the *parallel* connection of two independent elements it follows that:

1. Same stress ξ_L

$$R_S = 1 - \Pr\{\xi_{S_1} \le \xi_L \cap \xi_{S_2} \le \xi_L\} = 1 - \int_0^\infty \mathrm{f}_L(x)\mathrm{F}_{S_1}(x)\mathrm{F}_{S_2}(x)\,dx. \tag{2.79}$$

2. Independent stresses ξ_{L_1} and ξ_{L_2}

$$R_S = 1 - \Pr\{\xi_{S_1} \le \xi_{L_1}\}\Pr\{\xi_{S_2} \le \xi_L\} \triangleq 1-(1-R_1)(1-R_2) = R_1 + R_2 - R_1\, R_2\,. \tag{2.80}$$

As with Eqs. (2.78) and (2.80), the results of Table 2.1 can be applied in the case of *independent* stresses and elements. However, this *ideal situation* is seldom true for mechanical systems, for which Eqs. (2.77) and (2.79) are often more realistic. Moreover, the *uncertainty* about the *exact form* of the distributions for stress and strength far from the mean value, severely reduce the *accuracy* of the results obtained from the above equations in practical applications. For mechanical items, *tests* are thus often the only way to evaluate their reliability. Investigations into new methods are in progress, paying particular attention to the *dependence between stresses* and to a *realistic truncation* of the stress and strength densities.

For electronic items, Eqs. (2.78) and (2.80) can often be used to investigate *drift failures*. Quite generally, all considerations of Section 2.5 could be applied to electronic items. However, the method based on the failure rate, introduced in Section 2.2, is easier to be used and works reasonably well in many practical applications dealing with electronic and electromechanical equipment and systems.

2.6 Failure Mode Analyses

Failure rate analysis, as discussed in Section 2.2, does not account for the *mode* and the *effect* (consequence) of a failure. To understand the mechanism of system failures and in order to identify potential weaknesses of a *fail safe concept* it is necessary to perform a *failure mode analysis* for *critical system parts*. Such an analysis usually includes *failures* and *defects*, i.e. *faults*, and is termed FMEA (Fault Modes and Effects Analysis) or alternatively FMECA (Fault Modes, Effects, and Criticality Analysis) if the *fault severity* is also considered. FMEA/FMECA consists of the systematic analysis of *all* possible fault *modes*, their *causes*, *effects*, and *criticalities* [2.6, 2.81, 2.82, 2.84 to 2.90]. It also investigates ways for avoiding faults and/or for minimizing their consequences, and is performed *bottom-up* by the reliability engineer together with the designer. All possible *fault causes* during the item's design, development, manufacture, and use have to be considered in one run or in several steps (design FMEA/FMECA, process FMEA/FMECA). The procedure for an FMEA/FMECA is easy to understand but time-consuming to apply, see Table 2.5 for the procedure and Table 2.6 for a simple example. Other worksheet forms for FMEA/FMECA than those given in Table 2.6 exist [2.82, 2.84, 2.87, 2.89]. An FMEA/FMECA is *mandatory* for all parts of an item in which *redundancy* appears. This is to verify the effectiveness of the redundancy when failure occurs, and to define the element *in series* with the redundancy on the reliability block diagram. If the item should evolve to a *safe state* after a failure (*fail safe behavior*), an FMEA/FMECA can become necessary for a large part of the item considered.

The procedure for an FMEA/FMECA has been developed for *hardware*, but can also be used for *software* [2.85, 5.54, 5.57]. For mechanical items, the FMEA/FMECA is one of the most important tools for reliability analysis (Section 2.5).

A further possibility to investigate failure-cause-to-effect relationships is the *Fault Tree Analysis* (FTA) [2.6, 2.82 to 2.84, 2.89]. The FTA is a *top-down* procedure in which the undesired event, for example a critical failure at system level, is represented by AND, OR, and NOT combinations of causes at lower levels, see Fig. 2.13 for a simple example.

Compared to the FMEA/FMECA, the FTA can take *external influences* (human and/or environmental) better into account, and handle situations where more than one *primary fault* has to occur in order to cause the undesired event at the system level. However, it does not necessarily go through all possible *fault modes*. The combination of an FTA with an FMEA/FMECA leads to a *causes-and-effects chart*. Such a chart shows the logical relationship between all identified causes and their single or *multiple consequences*.

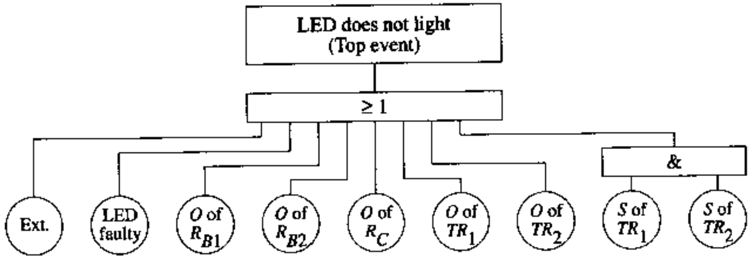


Figure 2.13 Simplified Fault Tree Analysis (FTA) for the circuit of Example 2.6 Point 7 (*O* = open, *S* = short, Ext. are possible external causes, such as power or control voltage out, manufacturing or installation errors, etc.)

Further methods which can support *fault modes and effects analysis* are *sneak analysis* (circuit, path, timing), *worst case analysis, drift analysis, stress-strength analysis, Ishikawa (fishbone) diagrams, Kepner-Tregoe method, Pareto diagrams,* and *Shewhart cycles* (Plan-Analyze-Check-Do), see e.g. [1.6, 1.17, 1.21, 2.7, 2.15].

Table 2.5 Procedure for performing a Fault, Modes, Effects and Criticality Analysis (FMECA)

1. Sequential numbering of the steps.
2. Designation of the element or part under consideration (reference to the reliability block diagram, part list, etc.) and short description of its function.
3. Assumption of a possible failure mode (all possible failure modes have to be considered).
4. Identification of possible causes for the failure mode assumed in point 3 (a cause for a failure can also be a flaw in the design or production phase).
5. Description of the symptoms which will characterize the failure mode assumed in point 3 and of the local effect of this failure mode (output/input relationships, possibilities for secondary failures, etc.).
6. Identification of the consequences of the failure mode assumed in point 3 on the next higher integration levels (up to the system level) and on the mission to be performed.
7. Identification of corrective actions which can mitigate the severity of the failure mode assumed in point 3, reduce the probability of occurrence, or initiate an alternate operational mode which allows continued operation when the failure occurs.
8. Evaluation of the severity of the failure mode assumed in point 3 (1 for minor, 2 for major, 3 for critical, 4 for catastrophic)
9. Estimation of the probability of occurrence (or of the associated failure rate) of the failure mode assumed in point 3 (with consideration of the cause of failure identified in point 4).
10. Formulation of pertinent remarks which complete the information in the previous columns and also of recommendations for corrective actions, which will reduce the consequences of the failure mode assumed in point 3.

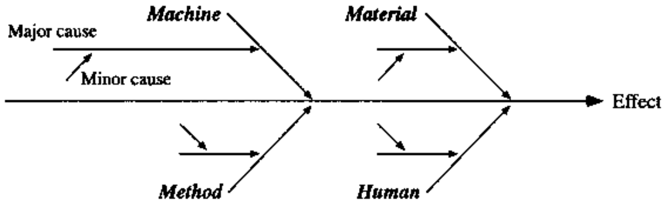


Figure 2.14 Structure of an Ishikawa (fishbone) diagram

Table 2.7 gives a comparison of the most important tools used for fault modes and effects analyses. Figure 2.14 shows the basic structure of an Ishikawa (fishbone) diagram. The Ishikawa diagram is a graphical visualization of the relationships between *causes and effect*, often grouping the causes into four classes: *machine, material, method,* and *human.*

Performing an FMEA/FMECA, FTA, or any other similar investigation presupposes a *detailed* technical knowledge and *thorough understanding* of the item and the technology considered. This is necessary to *identify all relevant* potential flaws (during design, development, manufacture, operation), their causes, and the more appropriate *corrective or preventive actions.*

2.7 Reliability Aspects in Design Reviews

Design reviews are important to find out, discuss, and eliminate *design and development weaknesses.* They are a part of *configuration management* and have been introduced in Appendix A3 (Table A3.3). To be effective, design reviews must be supported by *project-specific check lists.* Table 2.8 gives a catalog of questions which can be used to *generate* project-specific check lists for *reliability aspects* in design reviews (see Table 4.3 for maintainability and Appendix A4 for other aspects). As shown in Table 2.8, checking the reliability aspects during a design review is more than just verifying the value of the predicted reliability. The main purpose of a design review is to discuss the selection and use of components and materials, the adherence to given *design guidelines,* the presence of *potential reliability weaknesses,* and the results of *analyses and tests.* Table 2.9 can be used to support this aim.

Table 2.6 FMECA for the circuit of Example 2.6 (Point 7)

FAULT MODES AND EFFECTS ANALYSIS / FAULT MODES, EFFECTS, AND CRITICALITY ANALYSIS

FMEA/FMECA

Equipment: *control cabinet XYZ*
Item: *LED display circuit*
Prepared by: *E. & J. Thalhammer* Date: *Dec. 6, 1990*

Mission/required function: *fault signaling*
State: *operating phase* Page: *1*

(1) No.	*(2)* Element	*(3)* Assumed fault mode	*(4)* Possible causes	*(5)* Symptoms, local effects	*(6)* Effect on Mission	*(7)* Corrective action to avoid or mitigate the effect in (6)	*(8)* Severity	*(9)* Probability of occurrence	*(10)* Remarks and suggestions
1.	TR_1, NPN Si transistorin plastic package (E_4)	Short BCE, CE	Bad solder joint	Redundancy failed; $U_{CE} = 0$; no consequence to other elements	practically no consequence	—	1	$p = 10^{-5}$	a) λ for $\theta_A = 50°C$ and G_F b) it is possible to notify the failure of TR_1 (Level detector)
			Inherent failure					$\lambda = 0.610^{-9}\ h^{-1}$	
		Short BC	Bad solder joint	LED lights dimly; disappears by bridging CE; no consequence to other elements	Partial failure	Use a transistor of better quality; improve handling, assembly, and soldering procedures	2	$p = 10^{-5}$	
			Inherent failure					$\lambda = 0.3\ 10^{-9}\ h^{-1}$	
		Short BE	Bad solder joint	Circuit faulty; disappears by bridging CE; no consequence to other elements	Complete (possibly partial) failure		3	$p = 10^{-5}$	
			Inherent failure					$\lambda = 0.3\ 10^{-9}\ h^{-1}$	
		Open	Wrong connection, damage, cold solder joint				3	$p = 10^{-3}$	
			Inherent failure					$\lambda = 0.3\ 10^{-9}\ h^{-1}$	
		Intermittent failure	Damage, cold solder joint	Circuit works intermittently; no consequence to other elements	Partial to complete failure	Improve handling, assembly, and soldering procedures	2 to 3	$p = 10^{-4}$	
		Drift	Damage	The circuit works correctly even with large parameter deviations; no consequence to other elements	Practically no consequence	Improve handling	1 to 2	$p = 10^{-4}$	
			Wearout			—		$\lambda = 0.1\ 10^{-9}\ h^{-1}$	
2.	LED (E_1)	Open	Wrong connection, damage, cold solder joint	LED does not light; no consequence to other elements	Complete failure	Improve handling, assembly, and soldering procedures	3	$p = 10^{-3}$	a) λ for $\theta_A = 30°C$ and G_F b) be careful when forming the leads c) Observe the max. soldering time; distance between package and board > 2 mm d) pay attention to the cleaning medium e) hermet. package
			Inherent failure			Reduce θ_A		$\lambda = 0.8\ 10^{-9}\ h^{-1}$	
		Short	Bad solder joint	LED does not light; no consequence to other elements	Complete failure	Improve soldering procedure	3	$p = 10^{-5}$	
			Inherent failure			Reduce θ_A		$\lambda = 0.3\ 10^{-9}\ h^{-1}$	
		Intermittent failure	Damage, cold solder joint	LED lights intermittently; no consequence to other elements	Partial to complete failure	Improve handling, assembly, and soldering procedures	2 to 3	$p = 10^{-4}$	
		Drift	Damage	LED lights dimly; no consequence to other elements	Partial failure	Improve handling	2	$p = 10^{-4}$	
			Wearout			Reduce θ_A		$\lambda = 0.2\ 10^{-9}\ h^{-1}$	
			Corrosion			Protection against humidity		$\lambda \approx 0$	

Table 2.6 (cont.)

(1) No.	*(2)* Element	*(3)* Assumed fault mode	*(4)* Possible causes	*(5)* Symptoms, local effects	*(6)* Effect on Mission	*(7)* Corrective action to avoid or mitigate the effect in (6)	*(8)* Severity	*(9)* Probability of occurrence	*(10)* Remarks and suggestions
3.	R_C, Film resistor to limit the collector current (E_2)	Open	Damage, cold solder joint	Circuit faulty, works again by bridging R_C with an equivalent resistor; no consequence to other elements	Complete failure	Improve handling, assembly, and soldering procedures	3	$p = 10^{-4}$	a) λ for $\theta_A = 50°C$ and G_F b) a short on R_C can produce a short on V_{CC}
			Inherent failure			Use composition resistors (if possible)		$\lambda = 0.5\ 10^{-9}\ h^{-1}$	
		Short	Inherent failure	Circuit faulty; LED lights very brightly; secondary failure of LED and/or TR_1 and/or TR_2	Complete failure	Put 2 resistors in series ($R_C/2$)	3	$\lambda \approx 0$	
		Intermittent failure	Damage, cold solder joint	Circuit works intermittently; no consequence to other elements	Partial to complete failure	Improve handling, assembly, and soldering procedures	2 to 3	$p = 10^{-4}$	
		Drift	Damage	The circuit works correctly even with large parameter deviations; no consequence to other elements	Practically no consequence	Improve handling	1	$p = 10^{-6}$	
			Wearout			—		$\lambda \approx 0$	
4.	R_{B1}, Film resistor to limit the base current (E_3)	Open	Damage, cold solder joint	Circuit faulty, works again by bridging R_{B1} with an equivalent resistor; no consequence to other elements	Complete failure	Improve handling, assembly, and soldering procedures	3	$p = 10^{-4}$	a) λ for $\theta_A = 50°C$ and G_F b) a short on R_{B1} can produce a failure of TR_1
			Inherent failure			Use composition resistors (if possible)		$\lambda = 0.5\ 10^{-9}\ h^{-1}$	
		Short	Inherent failure	Partial failure; TR_1 can fail because of a too high base current	Partial failure	Put 2 resistors in series ($R_{B1}/2$)	2	$\lambda \approx 0$	
		Intermittent failure	Damage, cold solder joint	Circuit works intermittently; no consequence to other elements	Partial to complete failure	Improve handling, assembly, and soldering procedures	2 to 3	$p = 10^{-4}$	
		Drift	Damage	The circuit works correctly even with large parameter deviations; no consequence to other elements	Practically no consequence	Improve handling	1	$p = 10^{-6}$	
			Wearout			—		$\lambda \approx 0$	
5.	R_{B2} (E_6)	cf. Point 4							
6.	TR_2 (E_7)	cf. Point 1							
7.	PCB and solder joints (E_5)	cf. Points 1 to 4							

Table 2.7 Important tools for causes-to-effects-analyses

Tool	Description	Application	Effort
FMEA/FMECA (Failure Mode Effects Analysis/Failure Mode, Effects and Criticality Analysis)*	Systematic *bottom-up* investigation of the effects (consequences) at system (item) level of the failure (fault) *modes* of all parts of the item considered, as well as of manufacturing flaws (and as far as possible) of user's errors/mistakes	Development phase (design FMEA/FMECA) and production phase (process FMEA/FMECA); mandatory for all interfaces, in particular where *redundancy* appears and for *safety* relevant parts	Very large if performed for all elements (0.1 – 0.2 MM for a PCB)
FTA (Fault Tree Analysis)	Quasi-systematic *top-down* investigation of the effects (consequences) of faults (failures and defects) as well as of external influences on the reliability and/or safety of the system (item) considered; the top event (e.g. a specific catastrophic failure) is the result of AND/OR combinations of elementary events	Similar to FMEA/FMECA; however, combination of more than one fault (or elementary event) can be better considered as by an FMEA/FMECA; also is the influence of *external events* (natural catastroph, sabotage etc.) easier to be considered	Large to very large, if many top events are considered
Ishikawa Diagram (Fishbone Diagram)	Graphical representation of the causes-to-effect relationships; the causes are often grouped in four classes: machine, material, method/process, and human dependent	Ideal for team-work discussions, in particular for the investigation of design, development, or production weakenesses	Small to large
Kepner-Tregoe Method	Structured problem detection, analysis, and solution by complex situations; the main steps of the method deal with a careful problem analysis, decision making, and solution weighting	Generally applicable, in particular by complex situations and in inter-disciplinary work-groups	Largely dependent on the specific situation
Pareto Diagram	Graphical presentation of the frequency (hystogram) and (cumulative) distribution of the problem causes, grouped in application specific classes	Supports the objective decision making in selecting the causes of a fault and thus in defining the appropriate corrective action (Pareto rule: 80% of the problems are generated by 20% of the possible causes)	Small
Correlation Diagram	Graphical representation of (two) quantities with possible functional (deterministic or stochastic) relation on an appropriate x/y-cartesian coordinate system	Assessment of a relationship between two quantities	Small

* Faults (failures and defects) is generally used instead of failures; MM stays for man month

Table 2.8 Catalog of questions for the preparation of project-specific checklists for the evaluation of reliability aspects in preliminary design reviews (see Tab. 4.3 and Appendix A4 for other aspects)

1. Is it a new development, a redesign, or only a change/modification?
2. Can existing elements be reused?
3. Will the item or parts thereof be reused in other equipment or systems?
4. Is there test or field data available from similar items? What were the problems?
5. Has a list of preferred components been prepared?
6. Is the selection/qualification of nonstandard components and material specified? How?
7. Have all the important performance parameters of the item been clearly defined?
8. Have the interactions among elements been minimized?
9. Have all the specification requirements of the item been fulfilled? Can individual requirements be reduced? Can interface problems be expected?
10. Has the mission profile been defined? How has it been considered in the analysis?
11. Has a reliability block diagram been prepared?
12. Have the environmental conditions for the item been clearly defined? How are the operating conditions for each element?
13. Have derating rules been appropriately applied?
14. Has the junction temperature of all semiconductor devices been kept no higher than 100°C?
15. Have drift, worst-case, and sneak path analyses been performed? What are the results?
16. Has the influence of on-off switching and of external interference (EMC) been considered?
17. Is it necessary to improve the reliability by introducing redundancy?
18. Has an FMEA/FMECA been performed, at least for the parts where redundancy appears? How? Are single points of failure present? Can nothing possible be done against them? Are there safety problems? Can liability problems be expected?
19. Does the predicted reliability of each element correspond to its allocated value?
20. Has the predicted reliability of the whole item been computed? With which π-factors? Does this value correspond to the target given in the item's specifications?
21. Are there elements with a limited useful life?
22. Are there components which require screening? Assemblies which require environmental stress screening (ESS)?
23. Can design or construction be further simplified?
24. Is failure detection, localization, and removal easy? Are hidden failures possible?
25. Have the aspects of user-friendliness and of the man-machine interface been considered?
26. Have reliability tests been planned? What does this test program include?
27. Have the aspects of producibility, testability, and reproducibility been considered?
28. Have the supply problems (second source, long-term deliveries) been solved?
29. Is the item sufficiently protected during transportation and storage?
30. Have the logistical aspects (standardization of maintenance tools, test equipment, spare parts, etc.) been adequately considered?

Table 2.9 Form sheets for detecting and investigating potential reliability weaknesses at assembly and equipment level

a) Assembly/subassembly design

Position	Component	Failure rate λ		Deviation from reliability design guidelines	Component selection and qualification	Problems during design, develop., manufact., test, use	El. test and screening
		Parameters	λ (FITs)				

b) Assembly/subassembly manufacturing

Item	Layout	Placing	Soldering	Cleaning	El. tests	Screening	Fault (defect, failure) analysis	Corrective actions	Transportation and storage

c) Prototype qualification tests

Item	Electrical tests	Environmental tests	Reliability tests	Fault (defect, failure) analysis	Corrective actions

d) Equipment level

Assembling	Test	Screening (ESS)	Fault (defect, failure) analysis	Corrective actions	Transportation and storage	Operation (field data)

3 Qualification Tests for Components and Assemblies

Components and materials, as well as externally procured assemblies, can have a major impact on the quality and reliability of the equipment and systems in which they are used. Their *selection* and *qualification* has to be considered on a case-by-case basis (with care in the case of new technologies or important redesigns). Besides cost and availability on the market, important selection criteria are the intended application (specific stresses), technology, quality, long-term behavior of relevant parameters, and reliability. A *qualification* includes a *characterization* at different stresses (thermal and electrical for electronic components), *environmental tests*, *reliability tests*, and *failure analysis*. This chapter deals with the *selection criteria* for electronic components (Section 3.1), *qualification tests* for complex integrated circuits (Section 3.2), *failure modes, mechanisms, and analysis* of electronic components (Section 3.3), and qualification tests for electronic assemblies (Section 3.4). *Screening procedures* as well as test and screening *strategies* are introduced in Chapter 8. *Design guidelines* are discussed in depth in Chapter 5.

3.1 Selection Criteria for Electronic Components

As stated in Section 2.2 (Eq. (2.16)), the failure rate of equipment or systems without redundancy is the *sum* of the failure rates of their elements. Thus, for large equipment or systems *without redundancy*, high reliability can only be achieved by selecting components and materials with sufficiently *low failure rates*. Useful *information for such a selection* is:

1. Intended application, in particular required function, environmental conditions, as well as reliability and safety targets.
2. Specific properties of the component or material considered (technological limitations, useful life, long term behavior of relevant parameters, etc.).
3. Possibility of accelerated tests.

4. Results of qualification tests on similar components or materials.
5. Experience from field operation.
6. Influence of derating and/or screening.
7. Potential design problems (sensitivity of performance parameters, interface problems, EMC, etc.).
8. Limitations due to standardization or logistical aspects.
9. Potential production problems (assembling, testing, handling, storing, etc.).
10. Purchasing considerations (cost, delivery time, second sources, long-term availability, quality level).

As many of the above requirements are conflicting, component selection often results in a *compromise*. The following is a brief discussion of the most important aspects in selecting electronic components.

3.1.1 Environment

Environmental conditions have a major impact on the functionality and reliability of electronic components, equipment, and systems. They are defined for relevant applications in international standards [3.9]. Such standards specify stress limits and test conditions. Among others for

> heat (steady-state, rate of temperature change), cold, humidity, precipitation (rain, snow, hail), radiation (solar, heat, ionizing), salt, sand, dust, noise, vibration (sinusoidal, random), shock, fall, acceleration.

Several combinations of stresses have also been defined, for example

> temperature and humidity, temperature and vibration, humidity and vibration.

However, not all stress combinations are relevant. By combining stresses, care must be taken to avoid the activation of failure mechanisms which would not appear in the field.

Environmental conditions at the *equipment or system level* are *given by the application*. They can range from severe, as in aerospace and defense fields (with extreme low and high ambient temperatures, 100% relative humidity, rapid thermal changes, vibration, shock, and high electromagnetic interference), to favorable, as in computer rooms (with forced cooling at constant temperature and no mechanical stress). International standards, e.g. IEC 60721 [3.9], can be used to fix representative environmental conditions for many applications. Table 3.1 gives *examples* for environmental test conditions for electronic/electromechanical equipment and systems. The stress conditions given in Table 3.1 have indicative purpose and have to be refined according to the specific application, to be cost and time effective.

Table 3.1 Examples for environmental test conditions for electronic/electromechanical equipment and systems (IEC 60068 [3.9])

Environmental condition	Stress profile	Induced failures
Dry heat	48 or 72 h at 55, 70 or 85°C: el. test, warm up (2°C / min), hold (80% of test time), power-on (20% of test time), el. test, cool down (1°C / min), el. test between 2 and 16 h	*Physical*: oxidation, structural changes, softening, drying out, viscosity reduction, expansion *Electrical*: drift parameters, noise, insulating resistance, opens, shorts
Damp heat (cycles)	2, 6, 12 or 24 times 24 h cycles between 25 and 55°C with rel. humidity between 90 and 96% at 55°C and over 95% at 25°C: el. test, warm up (3 h), hold (9 h), cool down (3 h), hold (9 h), dry with air, el. test between 6 and 16 h	*Physical*: corrosion, electrolysis, absorption, diffusion *Electrical*: drift parameters, insulating resistance, leakage currents, shorts
Low temperature	48 or 72 h at –25, –40 or –55°C: el. test, cool down (2°C / min), hold (80% of test time), power-on (20% of test time), el. test, warm up (1°C / min), electrical test between 6 and 16 h	*Physical*: ice formation, structural changes, hardening, brittleness, increase in viscosity, contraction *Electrical*: drift parameters,opens
Vibrations (random)	30 min random acceleration with rectangular spectrum 20 to 2000 Hz and an acceleration spectral density of 3, 10, or 30 $(\mathrm{m/s}^2)^2/\mathrm{Hz}$ ($0.03, 0.1,$ or $0.3\,g_n^2/\mathrm{Hz}$): el. test, stress, visual inspection, el. test	*Physical*: structural changes, fracture of fixings and housings, loosening of connections, fatigue *Electrical*: opens, shorts, contact problems, noise
Vibrations (sinusoidal)	30 min at 0.15 mm or 20 $\mathrm{m/s}^2$ ($2\,g_n$), 0.35 mm or 50 $\mathrm{m/s}^2$ ($5\,g_n$), or at 0.75 mm or 100 $\mathrm{m/s}^2$ ($10\,g_n$) (peak displacement below or peak acceleration above the crossover frequency $f_c \approx 60$ Hz) at the resonant frequencies and the same test duration for swept freq. (3 axes): el. test, resonance determination, stress at the resonant frequencies, stresses at swept freq. (10 to 500 Hz), visual inspection, el. test	
Mechanical shock (impact)	1000, 2000 or 4000 impacts (half sine curve 300 or 500 $\mathrm{m/s}^2$ (30 or 50 g_n) peak value and 6 ms duration in the main loading direction or distributed in the various impact directions: el. test, stress (1 to 3 impacts/s), inspection (shock absorber), visual inspection, el. test	*Physical*: structural changes, fracture of fixings and housings, loosening of connections, fatigue *Electrical*: opens, shorts, contact problems, noise
Free fall	26 free falls from 50 or 100 cm drop height distributed over all surfaces, corners and edges, with or without transport packaging: electrical test, fall onto a 5 cm thick wooden block (fir) on a 10 cm thick concrete base, visual insp., el. test	

$g_n = 10\,\mathrm{m/s}^2$, el. = electrical

At the *component level,* to the stresses caused by the equipment or system environmental conditions *add* those stresses produced by the component itself, due to its internal electrical or mechanical load. The *sum* of these stresses gives the *operating conditions* necessary to determine the *stress at component level* and the corresponding *failure rate*. For instance, the ambient temperature *inside* an electronic assembly can be just some few °C higher than the temperature of the cooling medium, if forced cooling is used, but can become more than 60°C higher than the ambient temperature if cooling is poor.

3.1.2 Performance Parameters

The required *performance parameters* at component level are defined by the intended application. Once these requirements are established, the necessary *derating* is determined taking into account the quantitative relationship between failure rate and stress factors (Sections 2.2.4 and 5.1.1). It must be noted that the use of *better* components does not necessarily imply better performance and/or reliability. For instance, a faster IC family can cause EMC problems, besides higher power consumption and chip temperature. In critical cases, component selection should not be based only on short data sheet information. Knowledge of parameter sensitivity can be mandatory for the application considered.

3.1.3 Technology

Technology is rapidly evolving for many electronic components, see Fig. 3.1 and Table 3.2 for some basic information. As each technology has its advantages and weaknesses with respect to performance parameters and/or reliability, it is necessary to have a set of rules which can help to select a technology. Such rules (*design guidelines* in Section 5.1) are evolving and have to be periodically refined.

Of particular importance for *integrated circuits* (ICs) is the selection of the packaging form and type.

For the *packaging form*, distinction is made between inserted and surface-mounted devices. *Inserted devices* offer the advantage of easy handling during the manufacture of PCBs and also of lower sensitivity to *manufacturing defects or deviations*. However, the number of pins is limited. *Surface mount devices* (SMD) allow a large number of pins (more than 196 for PQFP and BGA), are cost and space saving, and have better electrical performance because of the shortened and symmetrical bond wires. However, compared to inserted devices, they have greater junction to ambient *thermal resistance*, are more stressed during soldering, and solder joints have a much lower mechanical strength. Difficulties can be expected with pitch lower than 0.3 mm, in particular if thermal and/or mechanical stresses occur in field (Sections 3.4 and 8.3).

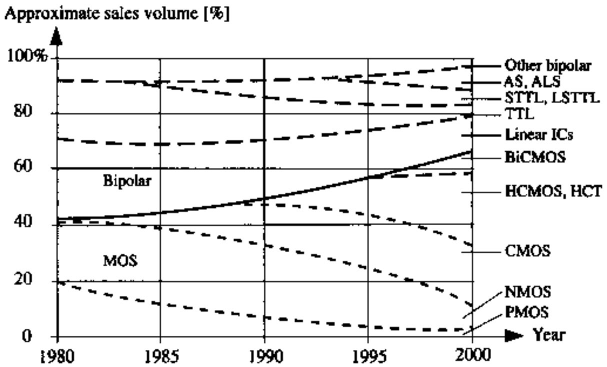


Figure 3.1 Basic IC technology evolution

Packaging types are subdivided into *hermetic* (ceramic, cerdip, metal can) and *nonhermetic* (plastic) packages. *Hermetic packages* should be preferred in applications with high humidity or in corrosive ambients, in any case if moisture condensation occurs on the package surface. Compared to plastic packages they offer lower thermal resistance between chip and case (Table 5.2), but are more expensive and sensitive to damage (microcracks) caused by inappropriate handling (mechanical shocks during testing or PCB production). *Plastic packages* are inexpensive, less sensitive to thermal or mechanical damage, but are permeable to *moisture* (other problems related to epoxy, such as ionic contamination and low glass-transition temperature, have been solved). However, better epoxy quality as well as new glassivation (passivation) based on silicon nitride leads to a much better protection against corrosion than formerly (Section 3.2.3, point 8).

If the results of qualification tests are good, the *use of ICs in plastic packages* can be allowed if one of the following conditions is satisfied:

1. Continuous operation, relative humidity ≤70%, noncorrosive or marginally corrosive environment, junction temperature ≤100°C, and equipment useful life less than 10 years.
2. Intermittent operation, relative humidity ≤60%, noncorrosive environment, no moisture condensation on the package, junction temperature ≤100°C, and equipment useful life less than 10 years.

For ICs with silicon nitride *glassivation* (*passivation*), the conditions stated in Point 1 above should also apply for the case of intermittent operation.

Table 3.2 Basic technological properties of electronic components

Component	Technology, Characteristics	Sensitive to	Application
Fixed resistors • Carbon film	A layer of carbon film deposited at high temperature on ceramic rods; ±5% usual; medium TC; relatively low drift (–1 to +4%); failure modes: opens, drift, rarely shorts; high noise; 1 Ω to 22 MΩ; low λ (0.2 to 0.4 FIT)	Load, temperature, overvoltage, frequency (> 50 MHz), moisture	Low power (≤1W), moderate temperatures (<85°C) and frequencies (≤ 50 MHz)
• Metal film	Evaporated NiCr film deposited on aluminum oxide ceramic; ±5% usual; low TC; low drift (±1%); failure modes: opens, drift, rarely shorts; low noise; 10 Ω to 2.4 MΩ; low λ (0.2 FIT)	Load, temperature, current peaks, ESD, moisture	Low power (≤ 0.5 W), high accuracy and stability, high freq. (≤ 500 MHz)
• Wire-wound	Usually NiCr wire wound on glass fiber substrate (sometimes ceramic); precision (±0.1%) or power (±5%); low TC; failure modes: opens, rarely shorts between adjacent windings; low noise; 0.1 Ω to 250 kΩ; medium λ (2 to 4 FIT)	Load, temperature, overvoltage, mechanical stress (wire < 25 µm), moisture	High power, high stability, low frequency (≤ 20 kHz)
• Thermistors (PTC, NTC)	PTC: Ceramic materials ($BaTiO_3$ or $SrTiO_3$ with metal salts), sintered at high temperatures showing strong increase of resistance (10^3 to 10^4) within 50°C; medium λ (5 to 50 FIT, large values for disk and rod packages) NTC: Rods pressed from metal oxides and sintered at high temperature with large negative TC ($TC \approx 1/T^2 \approx -3$ to $6\%/°C$ at 25°C); failure rate as for PTC	Current and voltage load, moisture	PTC: Temperature sensor, overload protection, etc.; NTC: Compensation, control, regulation, stabilization
Variable resistors • Cermet potentiometer, cermet trimmer	Metallic glazing (often ruthenium oxide) deposited as a thick film on ceramic rods and fired at about 800°C; usually ±10%; poor linearity (5%); medium TC; failure modes: localized wearout, drift, opens; relatively high noise (increases with age); 20 Ω to 2 MΩ; low to medium λ (10 to 50 FIT)	Load, current, fritting voltage (< 1.5 V), temperature, vibration, noise, dust, moisture, frequency (wire)	Should only be employed when there is a need for adjustment during operation, fixed resistors are to be preferred for calibration during testing, load capability proportional to the part of the resistor used
• Wire-wound potentiometer, wire-wound trimmer	CuNi or NiCr wire wound on ceramic rings or cylinders (spindle-operated potentiometers); normally ±10%; good linearity (1%); precision or power; low, nonlinear TC; low drift; failure modes: localized wearout, opens, relatively low noise; 10 Ω to 50 kΩ; medium λ (10 to 100 FIT)		
Capacitors • Plastic (KS, KP, KT, KC)	Wound capacitors with plastic film (K) of polystyrene (S), polypropylene (P), polyethylene-terephthalate (T) or polycarbonate (C) as dielectric and Al foil; very low loss factor (S, P, C); failure modes: shorts, opens, drift; pF to µF; low to medium λ (2 to 5 FIT)	Voltage stress, pulse stress (T, C), temperature (S, P), moisture* (S, P), cleaning agents (S)	Tight capacitance tolerances, high stability (S, P), low loss (S, P), well-defined temperature coefficient
• Metallized plastic (MKP, MKT, MKC, MKU)	Wound capacitors with metallized film (MK) of polypropylene (P), polyethylene-terephthalate (T), polycarbonate (C) or cellulose acetate (U); self-healing; low loss factor; failure modes: opens, shorts; nF to µF; low to medium λ (2 to 5 FIT)	Voltage stress, frequency (T, C, U), temperature (P), moisture* (P, U)	High capacitance values, low loss, relatively low frequencies (< 20 kHz for T, U)
• Metallized paper (MP, MKV)	Wound capacitors with metallized paper (MP) and in addition polypropylene film as dielectric (MKV); self-healing; low loss factor; failure modes: shorts, opens, drift; 0.1 µF to mF; low to medium λ (2 to 5 FIT)	Voltage stress and temperature (MP), moisture	Coupling, smoothing, blocking (MP), oscillator circuits, commutation, attenuation (MKV)
• Ceramic	Often manufactured as multi-layer capacitors with metallized ceramic layers by sintering at high temperature with controlled firing process (class 1: $\varepsilon_r < 200$, class 2: $\varepsilon_r \geq 200$); very low loss factor (class 1); temperature compensation (class 1); high resonance frequency: failure modes: shorts, drift, opens; pF to µF; low λ (0.3 to 3 FIT)	Voltage stress, temperature (even during soldering) moisture*, aging at high temperature (class 2)	Class 1: high stability, low loss, low aging; class 2: coupling, smoothing, buffering, etc.
• Tantalum (dry)	Manufactured from a porous, oxidized cylinder (sintered tantalum powder) as anode, with manganese dioxide as electrolyte and a metal case as cathode; polarized; medium frequency-dependent loss factor; failure modes: shorts, opens, drift; 0.1 µF to mF; medium λ (5 to 10 FIT, 20 to 40 FIT for bead)	Incorrect polarity, voltage stress, AC resistance (Z_0) of the el. circuit (new types less sensitive), temperature, frequency (>1kHz), moisture*	Relatively high capacitance per unit volume, high requirements with respect to reliability, $Z_0 \geq 1\,\Omega/V$
• Aluminum (wet)	Wound capacitors with oxidized Al foil (anode and dielectric) and conducting electrolyte (cathode); also available with two formed foils (nonpolarized); large, frequency and temperature dependent loss factor; failure modes: drift, shorts, opens; µF to 200 mF; medium to large λ (10 to 50 FIT); limited useful life (function of temperature and ripple)	Incorrect polarity (if polarized), voltage stress, temperature, cleaning agent (halogen), storage time, frequency (> 1 kHz), moisture*	Very high capacitance per unit volume, uncritical applications with respect to stability, relatively low ambient temperature (0 to 55°C)

Table 3.2 (cont.)

Table 3.2 (cont.)

Component	Technology, Characteristics	Sensitive to	Application
Diodes (Si) • General purpose	PN junction produced from high purity Si by diffusion or alloying; function based on the recombination of minority carriers in the depletion regions; failure modes: shorts, opens, drift; low λ (1 to 2 FIT for $\theta_J = 40°C$, 10 FIT for rectif. $\theta_J = 100°C$)	Forward current, reverse voltage, temperature, transients, moisture*	Signal diodes (analog, switch), rectifier, fast switching diodes (Schottky, avalanche)
• Zener	Heavily doped PN junction (charge carrier generation in strong electric field and rapid increase of the reverse current at low reverse voltages); failure modes: opens, drift, shorts; low to medium λ (2 to 4 FIT for voltage regulators ($\theta_J = 40°C$), 20 to 100 FIT for voltage ref. ($\theta_J = 100°C$)	Load, temperature, moisture*	Level control, voltage reference (allow for ±5% drift)
Transistors • Bipolar	PNP or NPN junctions manufactured using planar technology (diffusion or ion implantation); failure modes: shorts, opens, thermal fatigue for power transistors; low to medium λ (2 to 4 FIT for $\theta_J = 40°C$, 20 to 40 FIT for power transistors and $\theta_J = 100°C$)	Load, temperature, breakdown voltage (VBCEO, VBEBO), moisture*	Switch, amplifier, power stage (allow for ±20% drift, ±500% for ICBO)
• FET	Voltage controlled semiconductor resistance, with control via diode (JFET) or isolated layer (MOSFET); function based on majority carrier transport; N or P channel; depletion or enhancement type (MOSFET); failure modes: shorts, opens, drift; medium λ (3 to 10 FIT for $\theta_J = 40°C$, 30 to 60 FIT for power transistors and $\theta_J = 100°C$)	Load, temperature, breakdown voltage, ESD, radiation, moisture*	Switch (MOS) and amplifier (JFET) for high-resistance circuits (allow for ±20% drift)
Controlled rectifiers (Thyristors, triacs)	NPNP junctions with lightly doped inner zones (P, N), which can be triggered by a control pulse (thyristor), or a special antiparallel circuit consisting of two thyristors with a single firing circuit (triac); failure modes: drift, shorts, opens; large λ (25 to 100 FIT for $\theta_J = 100°C$)	Temperature, reverse voltage, rate of rise of voltage or current, commutation effects, moisture*	Controlled rectifier, overvoltage and overcurrent protection (allow for ±20% drift)
Opto-semiconductors (LED, IRED, photo-sensitive devices, opto-couplers, etc.)	Electrical/optical or optical/electrical converter made with photosensitive semiconductor components; transmitter (LED, IRED, laser diode etc.), receiver (photo-resistor, photo-transistor, solar cells etc.), opto-coupler, displays; failure modes: opens, drift, shorts; medium to large λ (2 to 100 FIT, $50\sqrt{\text{no. of pixels}}$ for LCD); limited useful life	Temperature, current stress, ESD, moisture*, mechanical stress	Displays, sensors, galvanic separation, noise rejection (allow for ±30% drift)

Table 3.2 (cont.)

Component	Technology, Characteristics	Sensitive to	Application
Digital ICs • Bipolar	Monolithic ICs with bipolar transistors (TTL, ECL, I^2L); important AS TTL (6 mW, 2 ns, 1.3 V) and ALS TTL (1 mW, 3 ns, 1.8 V); $V_{CC} = 4.5 - 5.5\,V$; $Z_{out} < 150\,\Omega$ for both states; low to medium λ (2 to 6 FIT for SSI/MSI, 30 to 300 FIT for LSI/VLSI)	Supply voltage, noise (> 1 V), temperature (0.5 eV), ESD, rise and fall times, moisture*	Fast logic (LS TTL 50 MHz, ECL 500 MHz) with uncritical power consump. rel. high cap. loading, $\theta_J < 175°C$ (< 200°C for SOI)
• MOS	Monolithic ICs with MOS transistors, mainly N channel depletion type (formerly also P channel); often TTL compatible and therefore $V_{DD} = 4.5 - 5.5\,V$ (100 µW, 10 ns); very high Z_{in}; medium Z_{out} (1 to 10 kΩ); medium to high λ (50 to 200 FIT)	ESD, noise (> 2 V), temperature (0.4 eV), rise and fall times, radiation, moisture*	Memories and microprocessors (up to 40 MHz), high source impedance, low capacitive loading
• CMOS	Monolithic ICs with complementary enhancement-type MOS transistors; often TTL compatible and therefore $V_{DD} = 4.5 - 5.5\,V$; power consumption ~ f (10 µW at 10 kHz, $V_{DD} = 5.5\,V$, $C_L = 15\,pF$); fast CMOS (HCMOS, HCT) for 2 to 6 V with 6 nsat 5 V and 20 µW at 10 kHz: large static noise immunity ($0.4\,V_{DD}$); very high Z_{in}; medium Z_{out} (0.5 to 5 kΩ); low to medium λ (2 to 6 FIT for SSI/MSI, 20 to 200 FIT for LSI/VLSI)	ESD, latch-up, temperature (0.4 eV), rise and fall times, noise (> $0.4\,V_{DD}$), moisture*	Low power consumption, high noise immunity, relatively low frequency (10 MHz CMOS, 50 MHz HCMOS), high source impedance, low cap. load, $\theta_J < 175°C$, for memories: < 125°C
• BiCMOS	Monolithic ICs with bipolar and CMOS devices; trend to 2 V supplies; combine the advantages of both bipolar and CMOS technologies	similar to CMOS	similar to CMOS but also for frequencies up to 1 GHz
Analog ICs • Operational amplifiers, comparators, voltage regulators, etc.	Monolithic ICs with bipolar and/or FET transistors for processing analog signals (operational amplifiers, special amplifiers, comparators, voltage regulators, etc.); up to about 200 transistors; often in metal packages; medium to high λ (3 to 300 FIT)	Temperature (0.6 eV), input voltage, load current, moisture*	Signal processing, voltage reg., low to medium power consump. (allow for ±20% drift), $\theta_J < 175°C$ (< 125°C for low power)
Hybrid ICs • Thick film, thin film	Combination of chip components (ICs, transistors, diodes, capacitors) on a thick film (5 – 20 µm) or thin film (0.2 – 0.4 µm) substrate with deposited resistors and connections; substrate area up to 10 cm^2; medium to high λ (usually determined by the chip components)	Manufacturing quality, temperature, mechanical stress, moisture*	Compact and reliable devices for avionics, instrumentation, etc. (allow for ±20% drift)

ESD = electrostatic discharge, TC = temperature coefficient, λ in $10^{-9}\,h^{-1}$ (FIT) for $\theta_A = 40°C$, $\pi_E = 1$, $\pi_Q = 1$, * nonhermetic packages

3.1.4 Manufacturing Quality

The *quality of manufacture* has a significant influence on electronic component reliability. However, information about *global* defective probabilities or agreed AQL values (also zero defects) are *not sufficient* to monitor the *reliability level* (AQL is nothing more than an agreed upper limit of the defective probability, generally at a risk $\alpha \approx 10\%$, see Section 7.1.3). Information about changes in the *defective probability* and the results of the corresponding *fault analysis* are important. For this, a direct *feedback* to the component manufacturer is generally more useful than an agreement on an AQL value.

3.1.5 Long-Term Behavior of Performance Parameters

The *long-term stability* of performance parameters is an important selection criterion for electronic components, allowing differentiation between good and poor manufacturers (Fig. 3.2). Verification of this behavior is generally undertaken with accelerated reliability tests (trends are often enough for many practical applications).

3.1.6 Reliability

The reliability of an electronic component can often be specified by its *failure rate* λ. Failure rate figures obtained from field data are valid if *intrinsic* failures can be *separated* from *extrinsic* ones, those figures given by component manufacturers are useful if computed with appropriate values for the *global activation energy* (0.4 to $0.6\,\text{eV}$ for ICs) and *confidence level* ($>60\%$ two sided or $>80\%$ one sided, see Section 7.1.1). Moreover, besides the numerical value of λ, the influence of the *stress factor* S is important as a selection criteria.

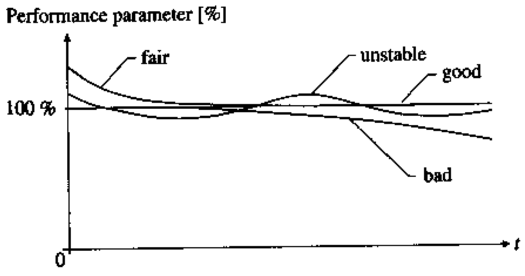


Figure 3.2 Long-term behavior of performance parameters

3.2 Qualification Tests for Complex Electronic Components

The purpose of a *qualification test* is to verify the *suitability* of a given item (material, component, assembly, equipment, system) for a stated application. Qualification tests are often a part of a *release procedure*. For instance, prototype release for a manufacturer and release for acceptance in a *Qualified Part List* (QPL) for a user. Such a test is generally necessary for new technologies or after important redesigns or changes in production processes. Additionally, periodic *requalification* of critical parameters is often necessary to monitor quality and reliability.

Electronic component qualification tests cover *characterization, environmental and special tests,* as well as *reliability tests.* They must be supported by intensive *failure analysis* to investigate *relevant failure mechanisms.* For a user, such a qualification test must consider:

1. Range of validity, narrow enough to be representative, but sufficiently large to cover the company needs and to repay test costs.
2. Characterization, to investigate the electrical performance parameters.
3. Environmental and special tests, to check technology limits.
4. Reliability tests, to gain information on the failure rate.
5. Failure analysis, to detect failure causes and investigate failure mechanisms.
6. Supply conditions, to define cost, delivery schedules, second sources, etc.
7. Final report and feedback to the manufacturer.

The extent of the above steps depends on the *importance* of the component being considered, the *effect* (consequence) of its failure in an equipment or system, and the *experience* previously gained with similar components and with the same manufacturer. National and international activities are moving toward agreements which should make a qualification test by the user unnecessary for many components [3.5, 3.17]. Procedures for environmental tests are often given in standards [3.9, 3.11].

A comprehensive qualification test procedure for ICs in *plastic packages* is given in Fig. 3.3. One recognizes the major steps (characterization, environmental tests, reliability tests, and failure analysis) of the above list. Environmental tests cover the thermal, climatic, and mechanical stresses expected in the application under consideration. The number of devices required for the reliability tests should be determined in order to *expect 3 to 6 failures during burn-in*. The procedure of Fig. 3.3 has been applied with the device-specific extensions (for instance data retention and programming cycles for nonvolatile memories, or modifications because of ceramic packages) to 12 memories each with 2 to 4 manufacturers for comparative investigations [3.2 (1993), 3.7, 3.15]. The cost for a qualification test based on Fig. 3.3 for 2 manufacturers (comparative studies) lies between US$ 30,000 and 60,000.

3.2.1 Electrical Test of Complex ICs

Electrical test of VLSI ICs is performed according to the following three steps:

1. Continuity test.
2. Functional and dynamic test (AC).
3. Test of DC parameters.

The *continuity test* checks whether every pin is connected to the chip. It consists in forcing a prescribed current (100 μA) into one pin after another (with all other pins grounded) and measuring the resulting voltage. For inputs with protection diodes and for normal outputs this voltage should lie between −0.1 and −1.5 V.

The *functional test* is performed together with the verification of the dynamic parameters. Figure 3.4 shows the principle of this test. The generator in Fig. 3.4 delivers one row after another of the *truth table* (or part of it) which has to be verified, with a frequency f_o. For a 40-pin IC, these are 40-bit words. Of these binary words, called *test vectors*, the inputs are applied to the *device under test* (DUT) and the expected outputs to a logical comparator. The actual outputs from the DUT and the expected outputs are compared at a time point selected with high accuracy by a strobe. Modern VLSI *automatic test equipment* (ATE) has test frequencies f_o up to 600 MHz and an overall precision of about 200 ps (resolution better than 30 ps). In a VLSI ATE not only the strobe but other pulses can be varied over a wide range. The *dynamic parameters* can be verified in this way. However, the direct measurement of a time delay or of a rise time is in general time consuming. The main problem with the functional test is that it is not possible to verify all the states and state sequences of a VLSI IC. To see this, consider for example that for an $n \times 1$ cell memory there are 2^n states and $n!$ possible address sequences, the corresponding *truth table* would contain $2^n \cdot n!$ rows, giving more than 10^{100} for $n = 64$. The procedure used in practical applications takes into account one or more of the following:

- *partitioning* the device into modules and testing each of them separately,
- finding out *regularities* in the truth table or given by technological properties,
- limiting the test to the part of the truth table which is important for the application under consideration.

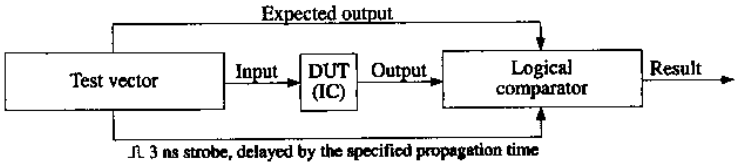


Figure 3.4 Principle of functional and AC testing for LSI and VLSI ICs

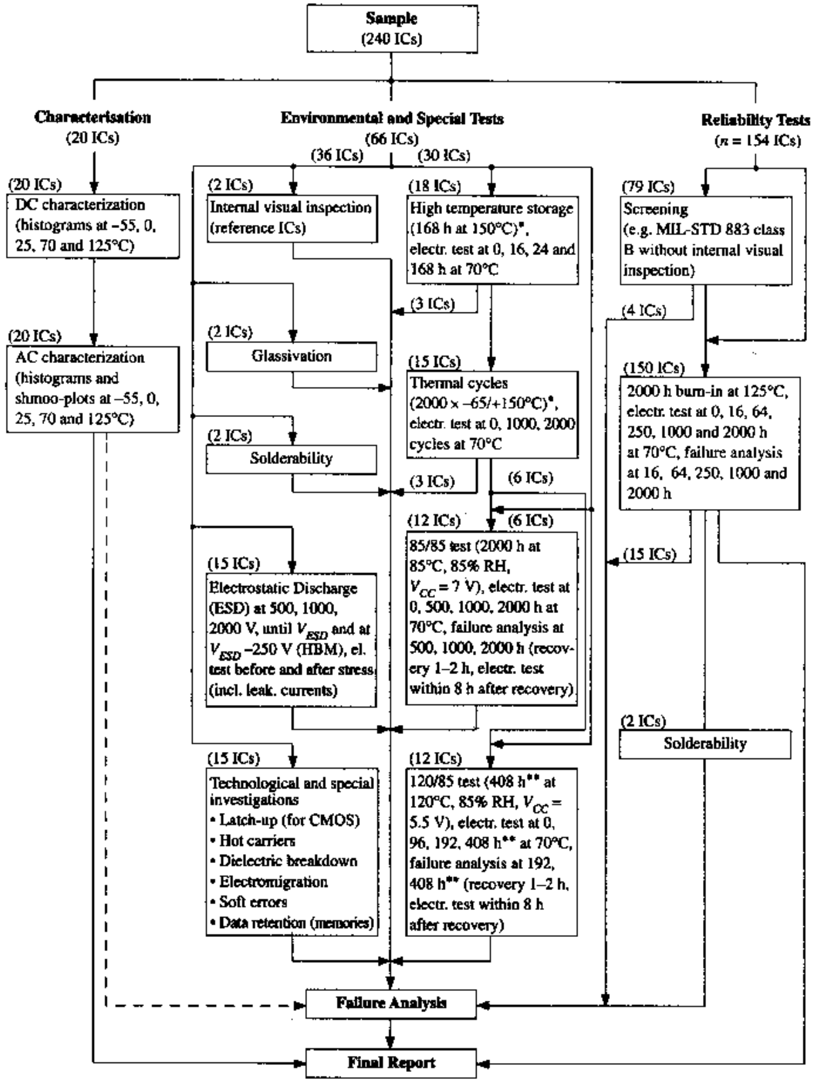


Figure 3.3 Qualification test procedure for complex ICs in plastic packages (industrial application, normal environmental conditions, 3 to 6 expected failures during the reliability test, i. e. $A\lambda \approx 2 \cdot 10^{-5}\,h^{-1}$, RH = relative humidity); *150°C by Epoxy resin and 175°C by Silicon resin, **1000 h by Si_3N_4 glassivation

The above limitations lead to the question of the *test coverage*, i.e. the percentage of faults which are detected by the test. A precise answer to this question can only be given in *some particular cases*, because information about the faults which *really* appear in a given IC is often lacking. *Fault models*, such as stuck-at-zero, stuck-at-one, or bridging are useful for PCB's testing, but generally of limited utility for test engineers at the component level.

The verification of *DC parameters* is simple. It is performed according to the manufacturer's specifications without restrictions (apart from very low input currents). For this purpose a *precision measurement unit* (PMU) is used to force a current and measure a voltage (V_{OH}, V_{OL}, etc.) or to force a voltage and measure a current (I_{IH}, I_{IL}, etc.). Before each step, the IC inputs and outputs are brought to the logical state necessary for the measurement.

For VLSI ICs, the electrical test should be performed at 70°C or at the highest specified operating temperature.

3.2.2 Characterization of Complex ICs

Characterization is a parametric, experimental analysis of the electrical properties of a given IC. Its purpose is to investigate the influence of different operating conditions such as supply voltage, temperature, frequency, and logic levels on the IC's behavior and to deliver a cost-effective test program for incoming inspection. For this reason a characterization is performed at 3 to 5 different temperatures and with a large number of different patterns.

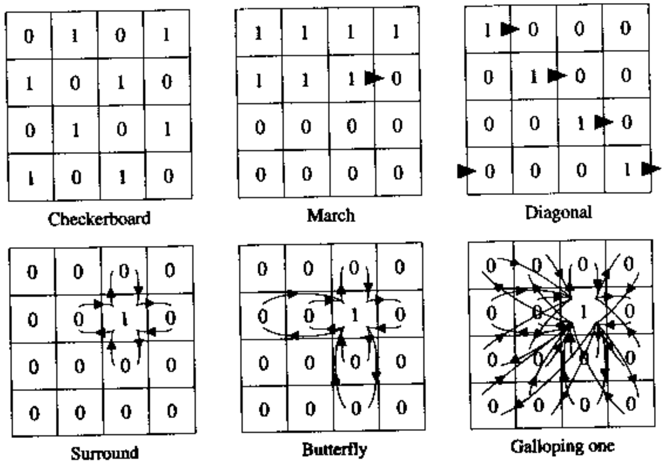


Figure 3.5 Test patterns for memories (see Table 3.3 for pattern sensitivity)

Table 3.3 Effectivity of various test patterns for detecting faults in SRAMs and approximate test times for a 100 ns 128 K×8 SRAM (tests on a Sentry S50, scrambling table with IDS5000 EBT)

Test pattern	Functional		Dyn. parameters		Number of test steps	Approx. test time[s]	
	D, H, S, O	C*	A, RA	C**		bit addr.	word addr.
Checkerboard	fair	poor	—	—	$4n$	—	0.05
March	good	poor	poor	—	$5n$	—	0.06
Diagonal	good	fair	poor	poor	$10n$	1	0.13
Surround	good	good	fair	fair	$26n-16\sqrt{n}$	27	0.34
Butterfly	good	good	good	fair	$8n^{3/2}+2n$	$8\cdot10^3$)	38
Galloping one	good	good	good	good	$4n^2+6n$	$4\cdot10^5$	$7\cdot10^3$

A=addressing, C=cap. coupling, D=decoder, H=stuck at 0 or at 1, O=open, S=short,
RA = read amplifier recovery time, * pattern dependent, ** pattern and level dependent

Referring to the functional and AC measurements, Figure 3.5 shows some basic *patterns for memories*. For the patterns of Fig. 3.5, Table 3.3 gives a qualitative indication of the corresponding pattern sensitivity for static random access memories (SRAMs), and the approximate test time for a 128K×8 SRAM. *Quantitative* evaluation of *pattern sensitivity* or of *test coverage* is seldom possible, in general, because of the limited validity of the *models for faults* available (Sections 4.2.1 and 5.2.2). As shown in Table 3.3, test time strongly depends on the pattern selected. As test times greater than 10s per pattern are also long in the context of a characterization (the same pattern will be repeated several thousands times, see e.g. Fig. 3.6), development of efficient test patterns is thus mandatory [3.2 (1989), 3.7, 3.15, 3.18]. For such investigations, knowledge of the relationship between address and physical location (scrambling) of the corresponding cell on the chip is important. If design information is not available, an *electron beam tester* (EBT) can often be used to establish the *scrambling table.*

An important evaluation tool during a characterization of complex ICs is the *shmoo plot*. It is the representation in an *x/y*-diagram of the operating region of an IC as a function of two parameters. As an example, Fig. 3.6 gives the shmoo plots for t_A versus V_{CC} of a 128K×8 SRAM for two patterns and two ambient temperatures [3.7]. To obtain the results shown in Fig. 3.6, the corresponding test pattern has been performed about 4000 times (2×29×61), each with a different combination of V_{CC} and t_A. If no fault is detected, an x or a • is plotted (defective cells are generally retested once, to confirm the fault). As shown in Fig. 3.6, a small (probably capacitive) coupling between nearby cells exists for this device, as a butterfly pattern is more sensitive than the diagonal pattern to this kind of fault. Statistical evaluation of shmoo plots is often done with *composite shmoo-plots* in which each record is labeled 0 through 9 (in 10% steps).

Table 3.4 DC parameters for a 40 pin CMOS ASIC with Schmitt trigger inputs (20 ICs)

		25°C			70°C		
V_{DD}		12 V	15 V	18 V	12 V	15 V	18 V
I_{DD} (μA)	min	310	410	560	260	340	470
	mean	331	435	588	270	358	504
	max	340	450	630	290	390	540
V_{0H} (V) ($I_{0H} = 2.4$ mA)	min	11.04	14.16	17.24	10.96	14.12	17.16
	mean	11.14	14.25	17.32	11.03	14.15	17.24
	max	11.20	14.33	17.40	11.12	14.20	17.32
V_{0L} (V) ($I_{0L} = 2.4$ mA)	min	0.40	0.36	0.32	0.44	0.24	0.32
	mean	0.47	0.42	0.38	0.52	0.45	0.41
	max	0.52	0.44	0.44	0.60	0.52	0.48
V_{Hyst} (V)	min	2.65	3.19	3.89	2.70	3.19	3.79
	mean	2.76	3.33	3.97	2.75	3.32	3.93
	max	2.85	3.44	4.09	2.85	3.44	4.04

From the above example one recognizes that in general only a *small part* of the possible states and state sequences can be tested. The definition of appropriate test patterns must thus pay attention to the specific device, its technology and regularities in the truth table, as well as to information about its application and experience with similar devices. A close *cooperation* between the test engineer and the user, and also if possible with the device designer and manufacturer, can help to reduce the amount of testing.

Measurement of DC parameters presents no difficulties for digital ICs (apart from small input currents in the pA range). Input and output pins are brought to the desired logic states and then a Precision Measurement Unit (PMU) is used to force a current and measure a voltage (V_{OH}, V_{OL}, V_{IL} etc.), or conversely to force a voltage and measure a current (I_{DD}, I_{OH}, I_{OL}, I_{IH}, I_{IL} etc.). Table 3.4 gives as an example some results for an application specific CMOS-IC (ASIC) with Schmitt-trigger inputs.

3.2.3 Environmental and Special Tests of Complex ICs

The aim of *environmental and special tests* is to submit a given IC to stresses which can be more severe than those encountered in field operation, in order to investigate *technological limits* and *failure mechanisms*. Such tests are often destructive. A failure analysis after each stress is important to evaluate *failure mechanisms* and to detect *degradation* (Section 3.3). The type and extent of environmental and special tests depend on the intended application (for Fig. 3.3, G_F according to Table 2.3) and on the specific characteristics of the component considered. The following is a description of the environmental and special tests given in Fig. 3.3:

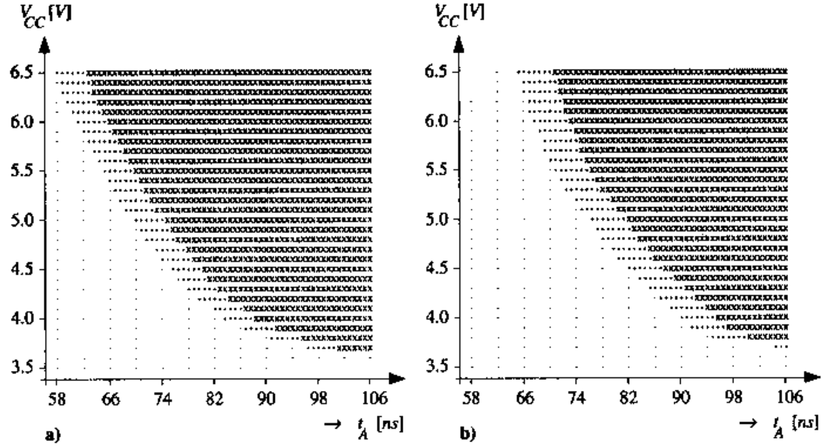


Figure 3.6 Shmoo plots of a 100 ns 128K×8 SRAM for test patterns **a)** Diagonal and **b)** Butterfly at two ambient temperatures 0°C (•) and 70°C (x)

1. *Internal Visual Inspection*: Two ICs are inspected and then kept as a reference for comparative investigation (check for damage after stresses). Before opening (using wet chemical or plasma etching), the ICs are x-rayed to locate the chip and to detect irregularities (package, bonding, die attach, etc.) or impurities. After opening, inspection is made with optical microscopes (conventional up to ×1,000 and/or stereo up to ×100). Improper placement of bonds, excessive height and looping of the bonding wires, contamination, etching, or metallization defects can be seen. Many of these deficiencies often have only a marginal effect on the reliability. Figure 3.7a shows a limiting case (mask misalignment). Figure 3.7b shows voids in the metallization of a 1M DRAM.
2. *Glassivation Test*: Glassivation (passivation) is the protective coating, usually *silicon dioxide* (PSG) and/or *silicon nitride*, placed on the entire (die) surface. For ICs in plastic packages it should ideally be free from *cracks* and *pinholes*. To check this, the chip is immersed for about 5 min in a 50°C warm mixture of nitric and phosphoric acid and then inspected with an optical microscope (e.g. as in MIL-STD-883 method 2021). *Cracks* occur in a silicon dioxide glassivation if the content of phosphorus is $< 2\%$. However, more than 4% phosphorus activates the formation of phosphoric acid. Recently, *silicon nitride* (often

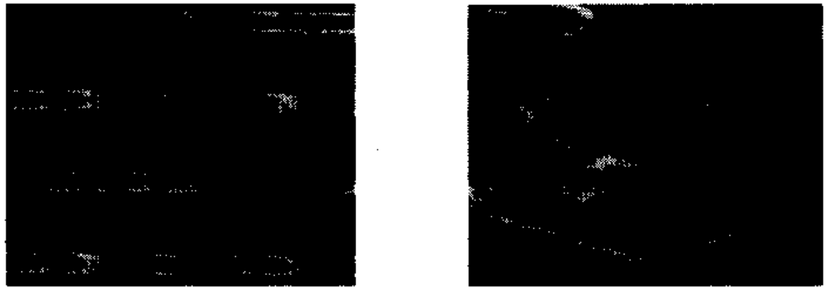
together with silicon dioxide in separate layers) glassivation has been introduced. Such a glassivation shows much more resistance to the penetration of moisture (see humidity tests in Point 8) and of ionic contamination.

3. *Solderability:* Solderability of tinned pins should no longer constitute a problem today, except after a very long storage time in a non-protected ambient or after a long burn-in or high-temperature storage. However, problems can arise with gold or silver plated pins, see Section 5.1.5.4. The solderability test is performed according to established standards (e.g. IEC 60068-2 or MIL-STD-883, [3.9, 3.11]) after the applicable conditioning, generally using the solder bath or the meniscograph method.
4. *Electrostatic Discharge (ESD)*: Electrostatic discharges during *handling*, *assembling*, and *testing* of electronic components and populated printed circuit boards (PCBs) can destroy or damage sensitive components, particularly *semiconductor devices*. All ICs families are sensitive to ESD. Such devices have in general *protection circuits*, passive and more recently active (better protection by a factor 2). To determine the *ESD immunity*, i.e. the voltage value at which damage occurs, different pulse shapes (models) and procedures to perform the test exist. For semiconductor devices, the *human body model* (HBM) and the *charged device model* (CDM) are the most widely used (the CDM seems to apply better than the HBM in reproducing some of the damage observed in field applications, see Section 5.1.4 for further details). Based on the experiences gained in qualifying 12 memory types according to Fig. 3.3 [3.2 (1993), 3.7] the following procedure can be suggested for the HBM:
   (i) 9 ICs divided into 3 equal groups are tested at 500, 1000, and 2000 V, respectively. Taking note of the results obtained during these preliminary tests, 3 new ICs are stressed with steps of 250 V up to the voltage at which damage occurs (V_{ESD}). 3 further ICs are then tested at $V_{ESD} - 250\,\text{V}$ to confirm that no damage occurs.
   (ii) The test consists of 3 positive and 3 negative pulses applied to each pin within 30 s. Pulses are generated by discharging a 100 pF capacitor through a 1.5 kΩ resistor placed in series to the capacitor (HBM), with wiring inductance < 10 μH. Pulses are between pin and ground, unused pins are left open.
   (iii) Before and after each test, leakage currents (when possible with the limits ±1 pA for open and ±200 nA for short) and all electrical characteristics are measured (electrical test as after any other environmental test).

   Experience shows that an electrostatic discharge often occurs between 1000 and 3000 V. The model parameters of 100 pF and 1.5 kΩ for the HBM are average values measured with humans (80 to 500 pF, 50 to 5000 Ω, 2 kV on synthetic floor and 0.8 kV on a antistatic floor with a relative humidity of about 50%). Measures to protect against ESD are discussed in Section 5.1.4.

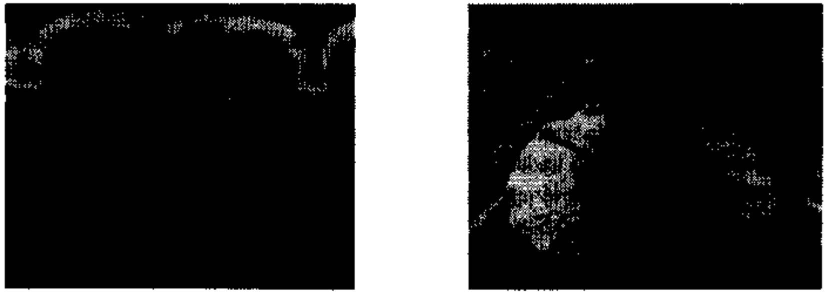
a) Alignment error at a contact window (SEM, ×10,000)

d) Silver dendrites near an Au bond ball (SEM, ×800)



b) Opens in the metallization of a 1 M DRAM bit line, due to particles present during the photolithographic process (SEM, ×2,500)

e) Electromigration in a 16K Schottky TTL PROM after 7 years field operation (SEM, ×500)



c) Cross section through two trench-capacitor cells of a 4 M DRAM (SEM, ×5,000)

f) Bondwire damage (delamination) in a plastic-packaged device after 500 × −50 / +150°C thermal cycles (SEM, ×500)

Figure 3.7 Failure analyses on ICs (Reliability Laboratory at the ETH Zurich)

5. *Technological Characterization*: Technological investigations are performed to check technological and process parameters with respect to *adequacy* and *maturity*. The extent of these investigations can range from a simple check (see Fig. 3.7c for an example) to a comprehensive analysis, because of detected weaknesses (e.g. misalignment, cracks, hidden particles, etc.). Investigations on *technological characterizations* are in progress, e.g. [3.31 to 3.65, 3.71 to 3.88]. The following is a short description of the most important *technological characterization methods:*
   - *Latch-up* is a condition in which an IC latches into a nonoperative state drawing an excessive current (often a short between power supply and ground), and can only be returned to an operating condition through removal and reapplication of the power supply. It is typical for CMOS structures, but can also occur in other technologies where a PNPN structure appears. Latch-up is primarily induced by voltage overstresses (on signals or on power supply lines) or by radiation. Modern devices often have a relatively high latch-up immunity (up to 200 mA injection current). A verification of latch-up sensitivity can become necessary for some special devices (ASICs for example). Latch-up tests stimulate voltage overstresses on signal and power supply lines as well as power-on/power-off sequences.
   - *Hot Carriers* arise in micron and submicron MOSFETs as a consequence of the *high electric fields* (10^4 to 10^5 V/cm) in transistor channels. Carriers may gain sufficient kinetic energy (some eV, compared to 0.02 eV in thermal equilibrium) to surmount the potential barrier at the oxide interface. The injection of carriers into the gate oxide is generally followed by electron-hole pairs creation by impact ionization in the channel (phonon or lattice impurity), and causes an increasing degradation of the transistor parameters, in particular an increase with time of the threshold voltage V_{TH} which can be measured in NMOS transistors. Effects on VLSI and ULSI-ICs are an increase of switching times (access times in RAMs for example), possible data retention problems (soft writing in EPROMs) and in general an increase of noise. Degradation through hot carriers is accelerated by increasing the drain voltage and lowering the temperature (negative activation energy of about −0.03 eV). The test is generally performed under dynamic conditions, at high power supply voltages (7 to 9 V) and at low temperatures (−20 to −70°C).
   - *Time-Dependent Dielectric Breakdown* occurs in thin gate oxide layers (< 100 nm) as a result of *very high electric fields* (up to 10^7 V/cm). Voltage stress on dielectric thin layers causes a continuous carrier injection (Fowler-Nordheim tunneling, hot carriers, etc.) into the isolation layer. As soon as the critical threshold is reached, breakdown takes place (often suddenly). The effects of gate oxide breakdowns are increased leakage currents or shorts between gate and substrate. The development in time of this failure mechanism depends on the oxide defects and process parameters. Particularly sensitive are memories ≥ 4M. An *Arrhenius model* with *activation energy* $E_a \approx 0.3$ eV can be used for the temperature, while for the voltage an exponential relationship is possible (the time dependence is considered by assuming a lognormal distribution of the failure-free operating time). Time-dependent dielectric breakdown tests are generally performed on special test structures (often capacitors).

   - *Electromigration* is the migration of metal atoms, and also of Si at the Al/Si interface, as a result of *very high current densities*, see Fig. 3.7e for an example of a 16K TTL PROM after 7 years of field operation. Earlier limited to ECL, electromigration also occurs today with other technologies (because of scaling). The median t_{50} of the failure-free operating time as a function of the current density and temperature can be obtained from the empirical model given by Black [3.44], $t_{50} = B\, j^{-n}\, e^{E_a/kT}$, where $E_a = 0.55$ eV for pure Al (0.75 eV for Al-Cu alloy), $n = 2$, and B is a process-dependent constant. Electromigration tests are generally performed at wafer level on test structures. Measures to avoid electromigration are the optimization of grain structure (bamboo structures), the use of Al-Cu alloys for the metallization and of compressive glassivations, as well as the introduction of multilayer metallizations.
   - *Soft errors* can be caused by the process or the chip design, as well as by process deviations. Key parameters are MOSFET threshold voltages, oxide thickness, doping concentrations, and line resistances. If for instance the post-implant of a silicon layer has been improperly designed, its conductivity might become too low. In this case, the word lines of a DRAM could suffer from signal reductions and at the end of the word line soft errors could be observed on some cells. As a further example, if logical circuits with different signal levels are unshielded and arranged close to the border of a cell array, stray coupling may destroy the information of cells located close to the circuit (chip design problem). Finally, process deviations can cause soft errors. For instance, signal levels can be degraded when metal lines are locally reduced to less than half of their width by the influence of dirt particles. The characterization of soft errors is difficult in general. At the chip level, an electron beam tester allows the measurement of signals within the chip circuitry. At the wafer level, single test structures located in the space between the chips (kerf) can be used to measure and characterize important parameters independently of the chip circuitry. These structures can usually be contacted by needles, so that a well-equipped bench setup with high-resolution I-V and C-V measurement instrumentation would be a suitable characterization tool.
   - *Data Retention* and *Program/Erase Cycles* are important for nonvolatile memories (EPROM, EEPROM, FLASH). A test for data retention generally consists of storage (bake) at high temperature (2000 h at 125°C for plastic packages and 500 h at 250°C for ceramic packages) with an electrical test at

70°C at 0, 250, 500, 1000, and 2000h (often using a checkerboard pattern with measurement of t_{AA} and of the margin voltage). A careful experimental investigation of EPROM data retention at temperatures higher than 250°C has shown a large deviation from the charge loss predicted by the thermionic model for temperatures higher than 300°C (about a factor 10 greater loss at 400°C [3.7]). Typical values for program/erase cycles during a qualification test are 100 for EPROMs and 10,000 for EEPROMs and Flash memories.

6. *High-Temperature Storage*: The purpose of high-temperature storage is the stabilization of thermodynamic equilibrium, and thus of the IC's electrical parameters. Failure mechanisms related to surface problems (contamination, oxidation, contacts, charge induced failures) are in particular activated. To perform the test, the ICs are placed on a metal tray (pins on the tray to avoid thermal voltage stresses) in an oven at 150°C for 200h. Should solderability be a problem, a protective atmosphere (N_2) can be used. Experience shows that for a mature technology (design and production processes), high temperature storage produces only a few failures (see also Section 8.2.2).
7. *Thermal Cycles*: The purpose of thermal cycles is to test the IC's ability to support rapid temperature changes. This activates failure mechanisms related to mechanical stresses caused by mismatch in the expansion coefficients of the materials used, as well as wearout because of fatigue, see Fig. 3.7f for an example. Thermal cycles are generally performed from air to air in a two-chamber oven (transfer from low to high temperature chamber and vice versa using a lift). To perform the test, the ICs are placed on a metal tray (pin on the tray to avoid thermal voltage stresses) and subjected to 2,000 thermal cycles from −65 to +150°C. Dwell time at the temperature extremes should be ≥ 10min (after thermal equilibrium of the IC has been reached within ± 5°C), transition time less than 1min. Should solderability be a problem, a protective atmosphere (N_2) can be used. Experience shows that for a mature technology (design and production processes), failures should not appear before several thousand thermal cycles (lower figures, for power devices).
8. *Humidity or Damp Heat Test*, 85/85 and *pressure cooker*: The aim of humidity tests is to investigate the influence of *moisture* on the chip surface, in particular corrosion. The following two procedures are often used:
   (i) Atmospheric pressure, 85 ± 2°C and 85 ± 5% rel. humidity (*85/85 Test*) for 168 to 5,000h.
   (ii) Pressurized steam, 110 ± 2°C or 120 ± 2°C or 130 ± 2°C and 85 ± 5% rel. humidity (*pressure-cooker test* or *highly accelerated stress test* (HAST)) for 24 to 408h (1,000h for silicon nitride glassivation).

   In both cases, a *voltage bias* is applied during exposure in such a way that power consumption is as low as possible, while the voltage is kept as high as possible (*reverse bias* with adjacent metallization lines alternatively polarized high and low, 1h *on* / 3h *off* intermittently if power consumption is greater than 0.01W). For a detailed procedure one may refer to IEC 60749 [3.9]. In the procedure of Fig. 3.3, both 85/85 and HAST tests are performed in order to correlate results and establish (empirically) a conversion factor. Of great importance for applications is the relation between the *failure rates* at elevated temperature and humidity (e.g. 85/85 or 120/85) and at field operating condi-tions (e.g. 35/60). A large number of models have been proposed in the literature to empirically fit the *acceleration factor* A associated with the 85/85 test

$$A = \frac{\text{mean time to failure at 85/85 } (\theta_2 / RH_2)}{\text{mean time to failure at lower stress } (\theta_1 / RH_1)}. \tag{3.1}$$

The most important of these models are

$$A = (\frac{RH_2}{RH_1})^{1/3} e^{\frac{E_a}{k}(\frac{1}{T_1} - \frac{1}{T_2})}, \tag{3.2}$$

$$A = e^{E_a [C_1(\theta_2 - \theta_1) + C_2(RH_2 - RH_1)]}, \tag{3.3}$$

$$A = e^{[\frac{E_a}{k}(\frac{1}{T_1} - \frac{1}{T_2}) + C_3(RH_2^2 - RH_1^2)]}, \tag{3.4}$$

$$A = e^{[\frac{E_a}{k}(\frac{1}{T_1} - \frac{1}{T_2}) + C_4(\frac{1}{RH_1} - \frac{1}{RH_2})]}, \tag{3.5}$$

$$A = e^{[\frac{1}{k}(\frac{E_a(RH_1)}{T_1} - \frac{E_A(RH_2)}{T_2}) + (RH_2 - RH_1)]}. \tag{3.6}$$

In Eqs. (3.2) to (3.6), E_a is the *activation energy*, k the Boltzmann constant ($8.6 \cdot 10^{-5}$ eV/K), θ the temperature in °C, T the absolute temperature (K), RH the relative humidity, and C_1 to C_4 are constants. Equations (3.2) to (3.6) are based on the *Eyring model* (Eq. (7.59)), the influence of the temperature and the humidity is multiplicative in Eqs. (3.3) to (3.5). Eq. (3.2) has the same structure as in the case of electromigration (Eqs. (7.60), (7.61)). In all models, the technological parameters (type, thickness, and quality of the glassivation, kind of epoxy, type of metallization, etc.) appear only indirectly in the activation energy E_a or in the constants C_1 to C_4. Relationships for HAST are less known, because it is not certain that the *failure mechanisms* are not changed. From the above considerations, 85/85 and HAST tests can be used as *accelerated tests* to assess the effect of damp heat combined with bias on ICs, however by accepting the numerical uncertainty in computing the acceleration factor. As a *global value* for the acceleration factor referred to operating field conditions of 35°C and 60% RH, one can assume for PSG a value between 100 and 150 for the 85/85 test and between 500 and 1,500 for the 120/85 test. To assure 10 years field operation at 35°C and 60% RH, PSG-ICs should thus pass without evident corrosion damage about 1,000h at 85/85 or 150h at 120/85. Practical results

[3.7] show that *silicon-nitride glassivation* offers a much greater resistance to moisture than PSG (factor of 5 to 10).

Also related to the effects of humidity is metal migration in the presence of reactive chemicals and voltage bias, leading to the formation of conductive paths (*dendrites*) between electrodes, see an example in Fig. 3.7d. A further problem related to *plastic packaged ICs* is that of bonding a *gold wire* to an *aluminum* contact surface. Because of the different interdiffusion constants of gold and aluminum, an inhomogeneous *intermetallic layer* (Kirkendall voids) appears at high temperature and/or in presence of contaminants, considerably reducing the electrical and mechanical properties of the bond. Voids grow into the gold surface like a plague, from which the name *purple plague* derives. Purple plague was an important reliability problem in the sixties. It propagates exponentially at temperatures greater than about 180°C. Although almost generally solved (bond temperature, Al-alloy, metallization thickness, wire diameter, etc.), verification after high temperature storage and thermal cycles is a part of a qualification test, especially for ASICs and devices in small-scale production.

Table 3.5 Failure modes of electronic components (indicative distribution in %)

Component		Shorts	Opens	Drift	Functional
Digital bipolar ICs		$50^{*\Delta}$	30^{*}	—	20
Digital MOS ICs		20^{Δ}	60^{*}	—	20
Linear ICs		—	25^{+}	—	75^{++}
Bipolar transistors		75	25	—	—
Field effect transistors (FET)		60	10	10	20
Diodes	general purpose	50	30	20	—
	Zener	20	40	40	—
Thyristors		40	10	—	$50^{\lozenge}$
Optoelectronic devices		10	50	40	—
Resistors, fixed		≈ 0	60	40	—
Resistors, variable		≈ 0	30	30	$40^{\#}$
Capacitors	foil	50	40	10	—
	ceramic	50	20	30	—
	Ta (solid)	60	30	10	—
	Al (wet)	30	30	40	—
Coils		40	40	10	10
Relays		20	—	—	$80^{\dagger}$
Quartz crystals		—	80	20	—

* input and output half each, $^{\Delta}$ short to V_{CC} or to GND half each, $^{+}$ no output, $^{++}$ improper output, $^{\lozenge}$ fallen off, $^{\#}$ localized wearout, † 60 fail to trip / 40 spurious trip

3.2.4 Reliability Tests

The aim of a *reliability test* for electronic components is to obtain information about the

- failure rate,
- long-term behavior of critical parameters,
- effectiveness of screening to be performed at the incoming inspection.

The test consists in general of a *dynamic burn-in*, with electrical measurements and *failure analyses* at appropriate time points (Fig. 3.3), also including some components which have not failed to check for degradation. The number of devices under test can be estimated from the *predicted failure rate* (Section 2.2.4) and the *acceleration factor* (Eq. (7.56)) in order to expect *3 to 6 failures* during burn-in ($n = k / (\lambda A t)$). Half of the devices are submitted to a *screening* (Section 8.2.2) in order to better isolate *early failures*. Statistical data analyses are given in Section 7.2 and Appendix A8.

3.3 Failure Modes, Failure Mechanisms, and Failure Analysis of Electronic Components

3.3.1 Failure Modes of Electronic Components

A *failure mode* is the *symptom* (local effect) by which a failure is observed. Typical failure modes are *opens*, *shorts*, *drift*, or *functional faults* for electronic components, and brittle rupture, creep, or cracking for mechanical components. *Average values* for the relative frequency of failure modes in electronic components are given in Table 3.5. From Table 3.5, resistors, crystals, and optoelectronic devices fail mainly as opens, shorts are more frequent in transistors, diodes, and capacitors, while functional faults occur mainly in relays, linear ICs, and thyristors. The values given in Table 3.5 have indicative purpose and have to be supplemented by results from *one's own experience* in qualification tests, field data, or failure analysis.

The different failure modes of *hardware*, often influenced also by the specific application, cause difficulties in investigating the *effect* (consequence) of failure and thus in the concrete realization of *redundancy* (series if short, parallel if open). For critical situations it can become necessary to use *quad redundancy* (Section 2.3.6). Quad redundancy is the simplest *fault tolerant structure* which can accept at least one failure (short or open) of any one of the 4 elements involved in the redundancy.

3.3.2 Failure Mechanisms of Electronic Components

A *failure mechanism* is the physical, chemical, or other process resulting in a failure. A large number of failure mechanisms have been investigated in the literature, e.g. [3.31 to 3.65] and [3.71 to 3.88]. For some of them, appropriate physical explanations have been found. For many others the *models* are *empirical* and often of limited validity. Evaluation of models for failure mechanisms should be developed in two steps: (i) verify the *physical validity* of the model; (ii) give its *analytical formulation* with the appropriate set of parameters to *fit the model to the data*. In any case, *experimental verification* of the model should be performed with at least a second, independent experiment. The *limits of the model* should be clearly indicated, to avoid misuse. The two most important models for discussions on failure mechanisms, the *Arrhenius model* and the *Eyring model* are introduced in Section 7.4 with *accelerated tests* (Eqs. (7.56) and (7.58) to (7.59)). Models to describe the influence of temperature and humidity in damp heat tests are discussed in Section 3.2.2 (Eqs. (3.2) to (3.6)). Table 3.6 summarizes the most important failure mechanisms for ICs, specifying influencing factors and the approximate distribution of the failure mechanisms for plastic-packaged ICs in industrial applications (G_F in Table 2.3). The percentage of misuse and mishandling failures can vary over a large range (20 to 80%) depending on the know-how of the design engineer using the device as well as on that of the equipment manufacturer and of the end user. For ULSI-ICs one can expect that the percentage of failure mechanisms related to *oxide breakdown* and *hot carriers* will grow in the future.

3.3.3 Failure Analysis of Electronic Components

The aim of a *failure analysis* is to investigate the *failure mechanisms* and find out the possible *failure causes*. A procedure for failure analysis of ICs (from a user's point of view) is shown in Fig. 3.8. It is based on the following steps and can be terminated as soon as the necessary information has been obtained:

1. *Failure detection and description*: A careful description of the failure as observed in situ and of the surrounding circumstances (operating conditions of the IC at the time of occurrence of the failure) is important. Also necessary are detailed information on the IC itself (type, manufacturer, manufacturing data, etc.), on the electrical circuit in which it was used, on the operating time, and if possible on the tests to which the IC was submitted previous to the final use (evaluation of possible damage, e.g. ESD). In a few cases the failure analysis procedure can be terminated, if evident mishandling or misuse failure can be confirmed.
2. *Nondestructive analysis*: The nondestructive analysis begins with an *external visual inspection* (mechanical damage, cracks, corrosion, burns, overheating, etc),

Table 3.6 Basic failure mechanisms of integrated circuits in plastic packages

Failure mechanism	Short description	Causes	Acceleration factors	%
Bonding **• Purple plague**	Formation of an intermetallic layer at the interface between wire (Au) and metallization (A1) causing a brittle region (voids in Au due to diffusion) which can provoke bond lifting	Different interdiffusion constants of Au and A1, Bonding temperature, contamination, metallization too thick	Temperature >180°C ($E_a = 0.7$ to 1.1 eV)	5
• Fatigue	Mechanical fatigue of bonding wires or bonding pads because of thermomechanical stress (also because of vibrations at the resonance frequency for hermetically sealed devices)	Different expansion coefficients of the materials in contact, for hermetically sealed devices also wire resonance	Thermal cycles with $\Delta\theta > 150°C$ (for hermetic devices, vibrations at resonance freq. ≤ 20 kHz)	
Surface **• Charge spreading (leakage currents, inversion)**	Charge spread laterally from the metallization or along the isolation interface, resulting in an inversion layer outside the active region which can provide (among other) a conduction path between two diffusion regions	Contamination with Na^+, K^+, etc., oxide layer too thin (MOS), package material	E, θ_J (E_a=0.5 to 1.2 eV, up to 2 eV for linear ICs)	5
Metallization **• Corrosion** **• Metal migration**	Electrochemical or galvanic reaction in the presence of humidity and ionic contamination (P, Na, Cl etc.), critical for PSG (SiO_2) glassivation with more than 4% P (<2% P gives cracks); migration of metal atoms in the presence of reactive chemicals, water, and bias, leading to conductive paths (dendrites) between electrodes	Humidity, voltage, contamination (Na^+, Cl^-, K^+), cracks or pinholes in the glassivation; humidity, voltage, migrating metals (Au, Ag, Pd, Cu, Pb, Sn), contaminant (encapsulant)	RH, E, θ_J ($E_a = 0.5$ to 0.7 eV)	5
• Electromigration	Migration of metal atoms (also of Si at contacts) in the direction of the electron flow, creating voids or opens in the structure	Current density ($>10^6$ A/cm^2), temperature gradient, anomalies in the metallization	j^n, θ_J ($n = 2$, E_a=0.55 to 0.75 eV, 1 eV for large-grain Al)	
Oxide **• Time-dependent dielec. breakdown (TDDB)** **• Ion migration (parasitic transistors, inversion)**	Breakdown of thin oxide layers occurring suddenly when sufficient charge has been injected to trigger a runaway proc.; carrier injection in the gate oxide because of E and θ_J; creation of charges in the SiO_2/Si-interface	High voltages, thin oxides, oxide defects; contamination with alkaline ions, pinholes, oxide or diffusion defects	E; θ_J (E_a=0.2 to 0.4 eV for oxide with defects, E_a = 0.5 to 0.6 eV for intrinsic oxide)	10
Others **• Intermetallic compound** **• hot carriers** **• α-particles** **• Latch-up, etc.**	Formation of intermetallic layer between metallization (Al) and substrate (Si); injection of electrons because of high E; generation of electron-hole pairs by α-particles (DRAMs); activation of PNPN paths	Mask defects, overheating, pure Al; dimensions, diffusion profiles, E; package material, external radiation; PNPN paths	E, θ_J, thermal cycles ($\Delta\theta > 200°C$); E; external radiation, voltage overstress	10
Misuse / Mishandling	Electrical, thermal, mechanical, or climatic overstress	Application, design, handling, test	———	65
				100

E = electric field, RH = relative humidity, j = current density, θ_J = junction temperature, % = indicative distribution in percent followed by an *x-ray inspection* (evident internal fault or damage), and a careful *electrical test* (Section 8.2.1). For ICs in hermetic packages, it can also be necessary to perform a seal test and if possible a dew-point test. The result of the nondestructive analysis is a careful description of the *external failure mode* and a first information about possible failure causes and failure mechanisms. In the cases of *evident failure causes* (mishandling or misuse failure), the failure analysis procedure can be terminated.

3. *Semidestructive analysis*: The semidestructive analysis begins by *opening the package*, mechanically for hermetic packages and with wet chemical (or plasma etching) for plastic ICs. A careful *internal visual* check is then performed with optical microscopes, conventional up to 1000× or stereo up to 100×. This evaluation includes opens, shorts, state of the glassivation/passivation, bonding, damage due to ESD, corrosion, cracks in the metallization, electromigration, particles, etc. If the IC is still operating (at least partially), other procedures can be used to *localize* more accurately the fault on the die. Among these are the *electron beam tester* (EBT or other voltage contrast techniques), *liquid crystals* (LC), *infrared thermography* (IRT), *emission microscopy* (EMMI), or one of the methods to detect irregular recombination centers, like *electron beam induced current* (EBIC) or *optical beam induced current* (OBIC). For further investigations it is then necessary to use a *scanning electron microscope* (SEM). The result of the semidestructive analysis is a careful description of the *internal failure mode* and first information about possible failure causes and failure mechanisms. In the case of *evident failure causes*, the failure analysis procedure can be terminated.
4. *Destructive Analysis*: A destructive analysis is performed if the previous investigations yield unsatisfactory results and there is a *realistic chance* of success through further analyses. After removal of the glassivation and other layers (as necessary) an inspection is carried out with a *scanning electron microscope* supported by a material investigation (e.g. *EDX spectrometry*). Analyses are then continued using methods of microanalysis (electron microprobe, ion probe, diffraction, etc.) and performing *microsections*. The destructive analysis is the last possibility to recognize the *original failure mode* and the *failure mechanisms* involved. However, it cannot guarantee success, even with skilled personnel and suitable analysis equipment.
5. *Failure mechanism analysis*: This step implies a correct interpretation of the results from steps 1 through 4. Additional investigations have to be made in some cases, but questions related to failure mechanisms can still remain open. In general, *feedback to the manufacturer* at this stage is mandatory.
6. *Corrective actions*: Depending on the identified failure causes, appropriate corrective actions should be started. These have to be discussed with the IC manufacturer as well as with the equipment designer, manufacturer, or user depending on the failure causes which have been identified.

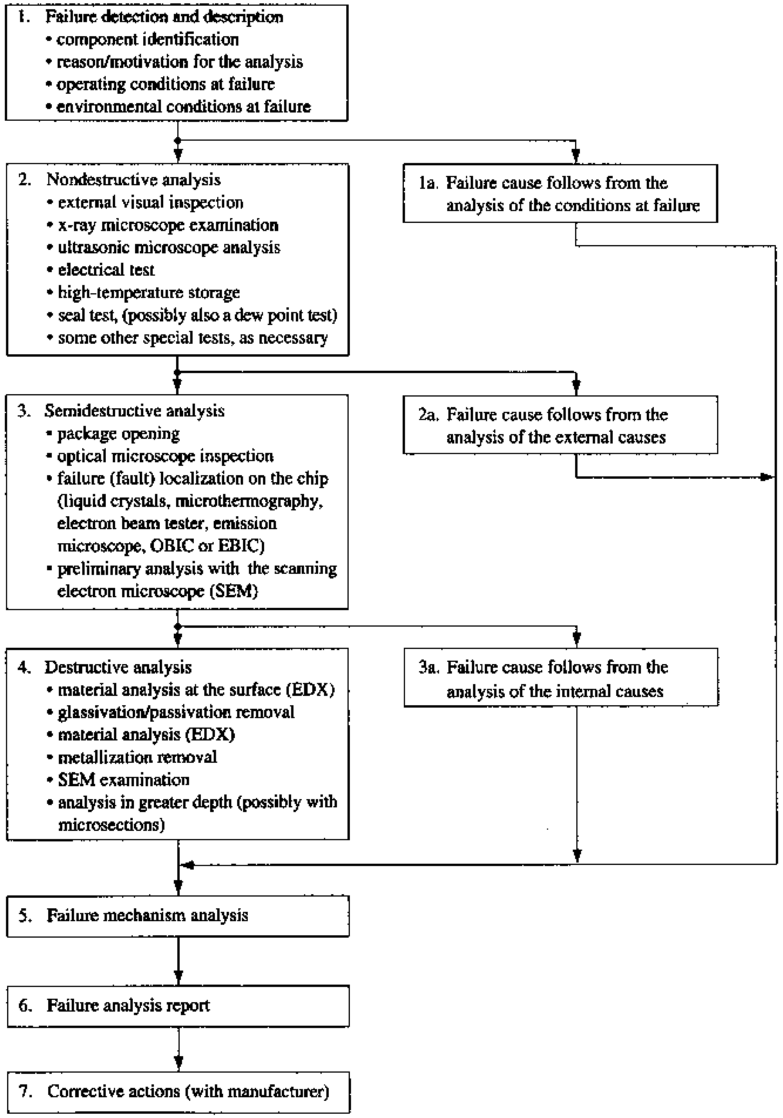


Figure 3.8 Procedure for failure analysis of electronic components (ICs as an example)

7. *Final report*: All relevant results of the steps 1 to 6 above and the agreed correcrective actions must be included in a (short and clear) final report.

The failure analysis procedure described in this section for ICs can be adapted to other components and extended to cover populated printed circuit boards (PCBs) as well as subassemblies or assemblies.

3.4 Qualification Tests for Electronic Assemblies

As outlined in Section 3.2 for components, *qualification tests* have the purpose of verifying the suitability of a given item (electronic assemblies in this section) for a stated application. Such a qualification involves *performance*, *environmental*, and *reliability* tests, and has to be supported by a careful *failures (faults) analysis*. To be efficient, it should be performed on *prototypes* which are *representative* for the production line in order to check not only the design but also the *production processes and procedures*. Results of qualification tests are an important input to the *critical design review* (Table A3.3). This section deals with some important aspects for the *qualification tests of electronic assemblies*, for instance *populated printed circuit boards (PCBs)*.

The aim of the *performance test* is similar to that of the *characterization* discussed in Section 3.2.2 for complex ICs. It is an experimental analysis of the electrical properties of the given assembly, with the purpose of investigating the influence of the most *relevant electrical parameters* on the behavior of the assembly at different *ambient temperatures* and power supply conditions (see Section 8.3 for some considerations on electrical tests of PCBs).

Environmental tests have the purpose of submitting the given assembly to stresses which can be more severe than those encountered in the field, in order to investigate *technological limits* and *failure mechanisms* (see Section 3.2.3 for complex ICs). The following procedure, based on the experience of a large number of equipment [3.76], can be recommended for assemblies of standard technology used in *high reliability* (or safety) applications (total ≥ 10 assemblies):

1. Electrical behavior at extreme temperatures with functional monitoring, 100h at −40°C, 0°C, and at +80°C (2 assemblies).
2. 4,000 thermal cycles (−40/+120°C with functional monitoring, < 5°C/min or > 20°C/min within the components according to the field application, ≥ 10min dwell time at −40°C and ≥ 5min at +120°C after the thermal equilibrium has been reached within ± 5°C (≥ 3 assemblies, metallographic analysis after 2,000 and 4,000 cycles).
3. Random vibrations at low temp., 1h with 2 to $6 g_{rms}$ at −20°C (2 assemblies).

4. EMC and ESD tests (2 assemblies).
5. Humidity tests, 240h 85/85 test (1 assembly).

Experience shows [3.76] that electronic equipment often behaves well even under extreme environmental conditions (operation at +120°C and −60°C, thermal cycles −40/+120°C with up to 60°C/min within the components, humidity test 85/85, cycles of 4h 95/95 followed by 4h at −20°C, random vibrations 20 − 500Hz at $4 g_{rms}$ and −20°C, ESD/EMC with pulses up to 15kV). However, problems related to crack propagation *in solder joints* appear, and metallographic investigations on more than 1,000 microsections confirm that *cracks* in solder joints are often initiated by production flaws, see Fig. 3.9 for some examples.

Many of the production flaws with *inserted components* can be avoided and cause only *minor* reliability problems. Voids, which make up the major portion of the solder defects observed, can be eliminated by a better plating of the through-holes. An optimization of the production process (reduced surface roughness of the walls, optimized plating parameters) generally solves the problem of voids. Since even voids up to 50% of the solder volume do not severely reduce the reliability of solder joints for inserted components, it is preferable to *avoid rework*. Poor wetting of the leads or the excessive formation of brittle intermetallic layers are *major* potential reliability problems for solder joints. These kinds of defects must be avoided through a better production process.

More critical are *surface mount devices* (SMD), for which a detectable *crack propagation* in solder joints often begins after some few thousand thermal cycles. Recent extensive investigations [3.79, 3.80, 3.88] show that crack propagtion is almost independent of pitch, at least down to a pitch of 0.3mm. Experimental results indicate an increase in the reliability of solder joints of IC's with shrinking pitches, due to the increasing flexibility of the leads. A new model to describe the *viscoplastic behaviour* of SMT solder joints has been developed in [3.88]. This model outlines the strong impact of *deformation energy* on *damage evolution* and assumes that cracks begin in a locally restricted recristallized area within the joint and propagate in a stripe along the main stress. The faster the *deformation rate* (the higher the termal gradient) and the lower the temperature, the faster damage accumulates in the solder joint. Basically, two different *deformation mechanisms* are present, *grain boundary sliding* at rather low thermal gradient and *dislocation climbing* at higher thermal gradient. Hence attention must be paid in defining environmental and reliability tests or screening procedures for assemblies in SMT (Section 8.3). In such a test or screening it is mandatory to activate only the failure mechanism which would also be activated in the field. Because of the elastic behaviour of the components and PCB, the *dwell time during thermal cycles* also plays an important role. The dwell time must be long enough to allow *relaxation* of the stresses and depends on temperature, temperature swing, and materials stiffness. As for the thermal gradient, it is difficult to give general rules [3.79 (1997 & 1999)].

Reliability tests at the assembly and higher integration level have as a primary purpose the detection of all *systematic failures* (Section 8.5) and a rough estimation of the *failure rate*. Precise information on the failure rate shape is seldom possible from qualification tests, because of cost and time limits. However, if such tests are *possible*, the following procedure can be used (total ≥ 8 assemblies):

1. 4,000 h dynamic burn-in at 80°C ambient temperature (≥ 2 assemblies, functional monitoring, intermediate electrical tests at 24, 96, 240, 1,000, and 4,000 h).
2. 5,000 thermal cycles −20/+100°C with < 5°C/min for applications with slow heat up and > 20°C/min for rapid heat up, dwell time ≥ 10 min at −20°C and ≥ 5 min at 100°C after the thermal equilibrium has been reached within ± 5°C (≥ 3 assemblies, metallographic analysis after 1,000, 2,000, and 5,000 cycles; crack propagation can be estimated using a Coffin-Manson relationship of the form $N = A\varepsilon^n$ with $\varepsilon = (\alpha_B - \alpha_C) l \Delta\theta / d$ [3.88, 3.79], the parameter A has to be determined with tests at different temperature swings).
3. 5,000 thermal cycles 0/+80°C, with temperature gradient as in point 2 above, combined with random vibrations $1 g_{rms}$, 20 − 500 Hz (≥ 3 assemblies, metallographic analysis after 1,000, 2,000 and 5,000 cycles).

Thermal cycles with random vibrations highly activate the major failure mechanism at the assembly level, i.e. *crack propagation in solder joints*. If such a stress occurs in the field, insertion technology should be preferred to SMT. Figure 3.10 shows a comparative investigation of crack propagation [3.79 (1993)].

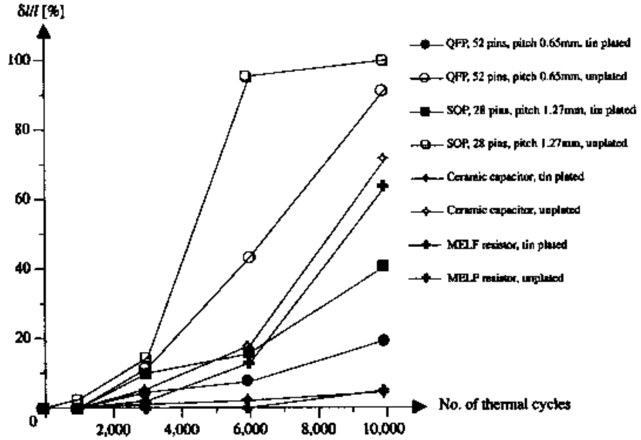
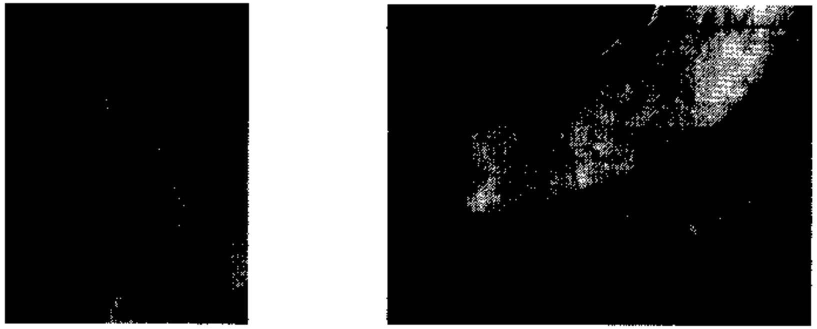
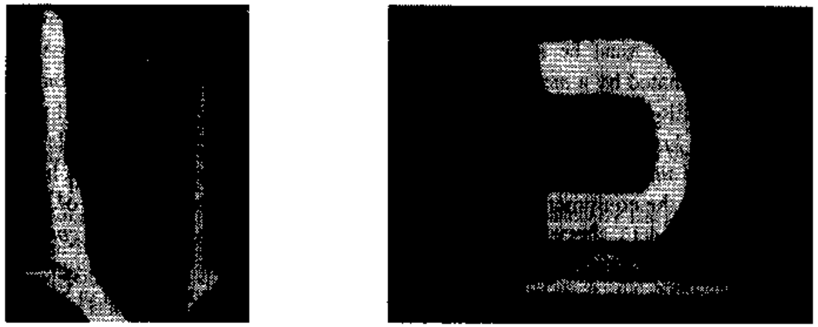


Figure 3.10 Crack propagation in different SMD solder joints as a function of the number of thermal cycles ($\delta l / l$ = crack length in % of the solder joint length, mean over 20 values, thermal cycles −20/+100°C with 60°C/min inside the solder joint [3.79 (1993)])



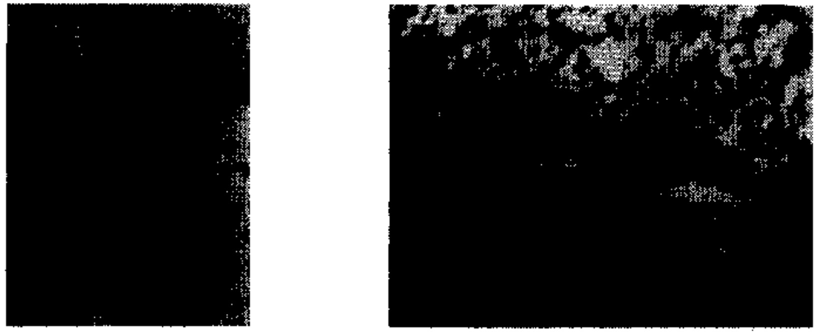
a) Void caused by an s-shaped pin gassing out in the area A (×20)

d) A row of voids along the pin of an SOP package (×60)



b) Flaw caused by the insertion of the insulation of a resistor network (×20)

e) Soldering defect in a surface mounted resistor, area A (×30)



c) Defect in the copper plating of a hole in a multilayer printed board (×50)

f) Detail A of Fig. 3.9e (×500)

Figure 3.9 Examples of production flaws responsible for the initiation of cracks in solder joints **a)** to **c)** inserted devices, **d)** to **f)** SMD (Reliability Laboratory at the ETH Zurich)

4 Maintainability Analysis

At the equipment and system level, *maintainability* has a great influence on *reliability* and *availability*. This is true if *redundancy* has been implemented and redundant parts are repaired on line, i.e. *without interruption* of operation. Maintainability represents thus an important parameter in the optimization of *availability* and *life-cycle cost*. Achieving high maintainability, requires appropriate activities which must be started *early in the design and development phase*, and be coordinated by a *maintenance concept*. To this belong *faults detection* and *isolation* (built-in tests), *partitioning* of the equipment or system into (almost) independent *last repairable units* (spare parts at equipment or system level), and *logistical support*, including after-sales service. A maintenance concept has to be *tailored* to the equipment or system considered. After the introduction of basic terms (Section 4.1), this chapter deals with a *maintenance concept* for *complex equipment and systems*, and presents then methods and tools for *maintainability calculations*. Models for *spare part provisioning* are considered in depth in Section 4.5. *Design guidelines* for maintainability are given in Section 5.2.

4.1 Maintenance, Maintainability

Maintenance defines all those *activities* performed on an item to *retain* it in or to *restore* it to a specified state. Maintenance is thus subdivided into *preventive maintenance*, carried out at predetermined intervals and according to prescribed procedures, to reduce the probability of failures or the degradation of the functionality of an item, and *corrective maintenance*, initiated after fault recognition and intended to bring the item into a state in which it can again perform the required function (Fig. 4.1). Corrective maintenance is also known as *repair* and can include any or all of the following steps: localization, isolation, disassembly, exchange, reassembly, alignment, checkout. The aim of preventive maintenance must also be to detect and repair *hidden failures*, i.e. failures in redundant elements. The time elapsed from the recognition of a failure until the final test after failure elimination, including all logistical delays (waiting for spare parts or for tools) is the *repair time*. Often, ideal *logistical support* with no *logistical delay* is assumed.

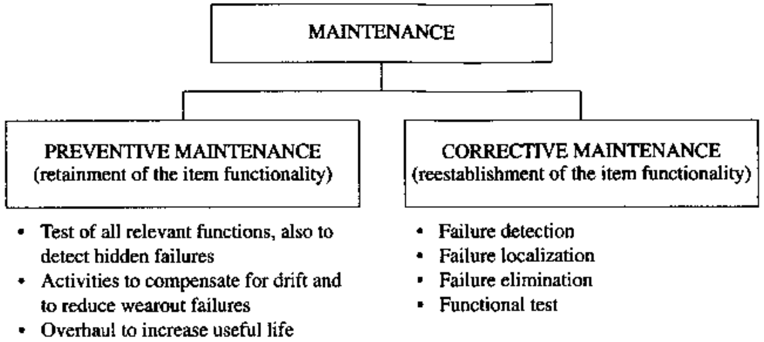


Figure 4.1 Maintenance tasks (failure could also be replaced by fault, thus including both defects and failures)

Maintainability is a *characteristic* of an item, expressed by the *probability* that *preventive maintenance* (serviceability) or *repair* (repairability) of the item will be performed within a stated time interval by given *procedures and resources* (number and skill level of the personnel, spare parts, test facilities, etc.). If τ' and τ'' are the (random) times required to carry out a repair or for a preventive maintenance, respectively, then

$$\text{Repairability} = \Pr\{\tau' \le t\} \qquad \text{and} \qquad \text{Serviceability} = \Pr\{\tau'' \le t\}. \qquad (4.1)$$

For a rough characterization, the expected values (means) of τ' and τ''

$$E[\tau'] = MTTR = \text{mean time to repair}$$

$$E[\tau''] = MTTPM = \text{mean time to preventive maintenance}$$

can often be used. Assuming t as a parameter, Eq. (4.1) gives the *distribution functions* of τ' and τ''. These distribution functions characterize the *repairability* and the *serviceability* of the item considered, respectively. Experience shows that τ' and τ'' often exhibit a *lognormal distribution* as defined by Eq. (A6.110). The typical shape of the corresponding *density* is shown in Fig. 4.2. A characteristic of the lognormal density is the sudden increase after a period of time in which its value is practically zero, and the relatively fast decrease after reaching the

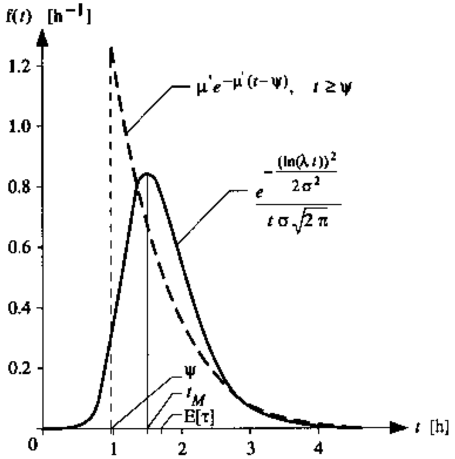


Figure 4.2 Density of the lognormal distribution function for $\lambda = 0.6\,\text{h}^{-1}$ and $\sigma = 0.3$ (dashed is the approximation given by a shifted exponential distribution with same mean)

maximum (modal value t_M). This shape can be accepted, taking into consideration the main terms of a repair time (Fig. 4.1). However, work with a lognormal distribution can become time consuming. In practical applications it is therefore useful to distinguish between one of the following two situations:

1. Investigation of *maintenance times*, often under the assumption of ideal logistical support; in this case, the appropriate distribution function must be considered, see Sections 7.3 and 7.5 for some examples with a lognormal distribution.
2. Investigation of the *reliability and availability of repairable equipment or systems*; the exact shape of the repair time distribution has often less influence on the final reliability and availability figures, as long as the *MTTR* is *unchanged* and *MTTR* << *MTTF* holds (Examples 6.7 and 6.8). In this case, the actual repair time distribution function can be *approximated* by an exponential function *with same mean* (ev. shifted at the origin for better results).

An *approximation* according to Point 2 is shown dashed in Fig. 4.2. For $\sigma/\lambda \ll 0.6$

$$\psi = t_M - \sqrt{\text{Var}[\tau']} = \frac{e^{\sigma^2} - \sqrt{e^{2\sigma^2} - e^{\sigma^2}}}{\lambda}$$

can be taken for the shift. The parameter μ' of the shifted exponential distribution function is then obtained by making the two mean values equal (Example 6.8)

$$\frac{e^{\sigma^2/2}}{\lambda} = \psi + \frac{1}{\mu'} \quad \rightarrow \quad \mu' = \frac{\lambda}{e^{\sigma^2/2} - \lambda\psi}. \tag{4.2}$$

As in the case of the failure rate $\lambda(t)$, for a statistical evaluation of repair times (τ') it would be preferable to omit data attributable to *systematic failures*. For the remaining data, a *repair rate* $\mu(t)$ can be obtained from the distribution function

$$\text{G}(t) = \Pr\{\tau' \leq t\},$$

with density $\text{g}(t) = d\text{G}(t)/dt$, as per Eq. (A6.27)

$$\mu(t) = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{t < \tau' \leq t + \delta t \mid \tau' > t\} = \frac{\text{g}(t)}{1 - \text{G}(t)}. \tag{4.3}$$

In evaluating the maintainability *achieved in the field*, the influence of the *logistical support* must be considered. *MTTR requirements* are discussed in Appendix A3.1. *MTTR estimation* and *demonstration* is given in Section 7.3.

4.2 Maintenance Concept

Like reliability, maintainability must be *built into* equipment and systems during the *design and development phase*. This is true in particular because maintainability cannot be easily predicted, and a maintainability improvement often requires important changes in the layout or construction of a given item. For these reasons, attaining a prescribed maintainability in complex equipment or systems can generally only be achieved by planning and realizing a *maintenance concept*. Such a concept deals with the following aspects:

1. Fault detection and isolation, including functional test after repair.
2. Partitioning of the equipment or system into (independent) last repairable units (LRU, i.e. spare parts at equipment or system level).
3. User documentation, including operating and maintenance manuals.
4. Training of operating and maintenance personnel.
5. Logistical support for the user, starting with after-sales service.

This section introduces the above points for the case of complex equipment or systems.

4.2.1 Fault Detection and Isolation

For complex equipment and systems, *detection* of partial failures, i.e. of failures of redundant parts (*hidden failures*) is generally performed by a *status test*, initiated by the operating personnel or by the *operation monitoring* (running in background). Properties, advantages, and disadvantages of both methods are summarized in Table 4.1. The choice between a (simple) *status test* or a (costly) *operating monitoring* is performed by considering reliability, availability, or safety aspects at system level.

The goal of *fault isolation* (diagnostic) is to localize faults (failures and defects) down to the *last repairable unit* (LRU), i.e. to the part which is considered as a *spare part* at the equipment or system level. LRUs are often designated as *line replaceable units*. They are generally assemblies (PCBs) or units which for repair purposes are considered as an *entity* and replaced on a *plug-out/plug-in basis* to reduce repair times. The repair of LRUs is often performed by specialized personnel and repaired LRUs are stored for *reuse*. Fault isolation should be performed using *built-in test* (BIT) facilities, if necessary supported by *built-in test equipment* (BITE). Use of external special tools should be avoided, however *check lists* and portable test equipment can be useful to limit the amount of built-in facilities.

Table 4.1 Automatic and semiautomatic fault detection

	Status Test		Operation Monitoring
	Rough (quick test)	Complete (functional test)	
Properties	• Testing of all important functions, if necessary with the help of external test equipment • Initiated by the operating personnel, then runs automatically	• Periodic testing of all important functions • Initiated by the operating personnel, then runs automatically or semi-autom. (possibly without external stimulation or test equipment)	• Monitoring of all important functions and automatic display of complete and partial faults • Performed with built-in means (BIT/BITE)
Advantages	• Lower cost • Allows fast checking of the functional conditions	• Gives a clear status of the functional conditions of the item considered • Allows fault isolation down to an assembly (LRU)	• Runs automatically on-line, i.e. in background
Drawbacks	• Limited fault isolation	• Relatively expensive • Runs generally off-line (i.e. not in background)	• Expensive

LRU = last repairable unit, BIT = built-in test, BITE = built-in test equipment

Fault detection and *fault isolation* are closely related and should be considered together using *common* hardware and/or software. A high degree of automation should be striven for, and test results should be automatically recorded. A one-to-one correspondence between test messages and content of the *user documentation* (operating and maintenance manuals) must be assured.

Built-in tests (BIT) should be able to identify *hidden faults*, i.e. faults (defects or failures) of redundant parts and, as far as possible, also of *software defects*. This ability is generally characterized by the following *testability* parameters:

- degree of fault detection,
- degree of fault isolation,
- correctness of the fault isolation,
- test duration.

The first two parameters can be expressed by a *probability*. *Distinction* between *failures* and *defects* is important. As a measure of the *correctness* of the fault isolation capability, one can use the ratio between the number of correctly localized faults and the number of localization tests performed. This figure, similar to that of *test coverage*, must often remain on an empirical basis, because of the lack of information about the defects and failures really present or possible in the item considered. For the test duration, it is generally sufficient to work with mean values. *Fault mode* analysis methods (FMEA/FMECA, FTA, cause-and-effect charts, etc.) as introduced in Section 2.6 can be used to check the effectiveness of built-in facilities.

Built-in test facilities, and in particular built-in test equipment, must be defined taking into consideration not only price/performance aspects but also their *impact* on the *reliability* and *availability* of the equipment or system in which they are used. BITs can often be integrated into the equipment or system considered. However, specially conceived BITE is generally more efficient than standard solutions. For such a selection, the following aspects are important:

1. *Simplicity*: Test sequences, procedures, and documentation should be as easy as possible.
2. *Standardization*: The greatest possible standardization should be striven for, in the hardware and software.
3. *Reliability*: Built-in facilities should have a failure rate of at least one order of magnitude lower than that of the equipment or system in which they are used; their failure should not influence the item's operation (FMEA/FMECA).
4. *Maintenance*: The maintenance of BIT/BITE must be simple and should not interfere with that of the equipment or system; the user should be connected to the *field data change service* of the manufacturer.

For some applications, it is important that fault isolation (or at least part of the diagnostic) can be *remotely controlled*. Such a requirement can often be easily satisfied, if stated early in the design phase. A further development of the above considerations can lead to maintenance concepts which allow an automatic or semiautomatic *reconfiguration* of an item after a failure.

4.2.2 Equipment and System Partitioning

The consequent *partitioning* of complex equipment and systems into almost independent *last repairable* (or line replaceable) *units* (LRUs) is essential for good maintainability (a typical LRU for complex equipment or systems is a populated printed circuit board (PCB)). Partitioning must be performed *early in the design phase*, because of its impact on the layout and construction of the equipment or system considered. LRUs should constitute *functional units* and have *clearly defined interfaces* with other LRUs. Ideally LRUs should allow a *modular construction* of the equipment or system, i.e. constitute almost autonomous units which can be tested independently from every other one.

Related to the above aspects are those of *accessibility, adjustment*, and *exchangeability*. Accessibility should be especially easy for LRUs with *limited useful life*, high failure rate, or wearout. The use of digital techniques largely reduces the need for *adjustment* (alignment). Quite generally, hardware adjustment in the field should be avoided. *Exchangeability* can be a problem for equipment and systems with long *useful life*. *Spare part reservation* and aspects of *obsolescence* can in such cases become mandatory (Section 4.5).

4.2.3 User Documentation

User (or product) documentation for complex equipment and systems can include all of the following Manuals or Handbooks

- General Description
- Operating Manual
- Preventive Maintenance (Service) Manual
- Corrective Maintenance (Repair) Manual
- Illustrated Spare Parts Catalog
- Logistical Support.

It is important that the contents of the user documentation is *consistent* with the hardware and software status of the item considered. Emphasis must be placed on a clear and concise presentation, with block diagrams, flow charts, and check lists. The language should be easily understandable to non-specialized personnel. Procedures should be self sufficient and contain checkpoints to prevent the skipping of important steps.

4.2.4 Training of Operating and Maintenance Personnel

Suitably equipped, well trained, and motivated maintenance personnel are an important prerequisite to achieve short maintenance times and to avoid *human errors*. Training must be comprehensive enough to cover present needs. However, for a complex system it should be periodically updated to cover technological changes introduced in the system and to remotivate the operating and maintenance personnel.

4.2.5 User Logistical Support

For complex equipment or systems, customers (users) generally expect from the manufacturer a *logistical support* during the useful life of the item under consideration. This can range from support on an *on-call basis* up to a *maintenance contract* with manufacturer's personnel located at the user site. One important point in such a logistical support is the definition of *responsibilities*. For this reason, maintenance is often subdivided into four levels, see Table 4.2 for an example in the defense area. The *first level* concerns simple maintenance work such as the status test, fault detection, and fault isolation down to the equipment level. This task is generally performed by *operating personnel*. At the *second level*, fault localization is refined, the defective LRU is replaced by a good one, and the functional test is performed. For this task *first line maintenance personnel* is often required. At the *third level*, faulty LRUs are repaired by *maintenance*

Table 4.2 Maintenance levels in the defense area

	logistical level	Location	Carried out by	Tasks
Advanced maintenance service	Level 1	Field	Operating personnel	• Simple maintenance work • Status test • Fault detection • Fault isolation down to equipment level
	Level 2	Cover	First line maintenance personnel	• Preventive maintenance • Fault isolation down to LRU level • First line repair (LRU replacement) • Functional test
Back-up maintenance service	Level 3	Depot	Maintenance personnel	• Difficult maintenance • Repair of LRUs
	Level 4	Arsenal or Industry	Specialists from arsenal or industry	• Reconditioning work • Important changes or modifications

LRU = last repairable unit (spare part at system level) *personnel* and stored for reuse. The *fourth level* generally relates to *overhaul* or *revision* (essentially for large mechanical parts subjected to wear, erosion, scoring, etc.) and is often performed at the manufacturer's site by *specialized personnel.*

For large mechanical systems, maintenance can account for 30% of the operating costs. A careful optimization of these costs may be necessary in many cases. The part contributed by preventive maintenance is more or less deterministic. For the corrective maintenance, cost equations weighted by probabilities of occurrence can be established from considerations similar as those given in Sections 1.2.9 and 8.4, see also Section 4.5 for some aspects of spare-part provisioning.

Table 4.3 Catalog of questions for the preparation of project-specific checklists for the evaluation of maintainability aspects in preliminary design reviews (see Tab.2.8 & Appendix A4 for other aspects)

1. Has the equipment or system been conceived with modularity in mind? Are the modules functionally independent and separately testable?
2. Has a concept for fault detection and isolation been planned and realized? Is fault detection automatic or semiautomatic? Which kind of faults are detected? How does fault isolation work? Is isolation down to LRUs possible? How large are the values for of fault detection and fault isolation (coverage)?
3. Can redundant elements be repaired on-line (without interruption of the operation at the system level)?
4. Are enough test points provided? Are they clearly marked and easily accessible? Do they have pull-up/pull-down resistors?
5. Have hardware adjustments (or alignments) been reduced to a minimum? Are the adjustable elements clearly marked and easily accessible? Is the adjustment uncritical?
6. Has the amount of external test equipment been kept to a minimum?
7. Has the standardization of components, materials, and maintenance tools been considered?
8. Are last repairable units (LRUs) identical with spare parts? Can they be easily tested? Is a spare part provisioning concept available?
9. Are all elements with limited useful life clearly marked and easily accessible?
10. Are access flaps (and doors) easy to open (without special tools) and self-latching? Are they of sufficient size? Have plug-in unit guide rails self-blocking devices? Can a standardized extender for PCBs be used?
11. Have indirect connectors been used? Is the plugging-out/plugging-in of PCBs (or LRUs) easy? Have reserve contacts been provided? Are power supplies and ground distributed across different contacts?
12. Have wires and cables been conveniently placed? Also with regard to maintenance?
13. Are sensitive elements sufficiently protected against mishandling during maintenance?
14. Can preventive maintenance be performed on-line? Does preventive maintenance also allow the detection of hidden failures?
15. Can the item (possibly the system) be considered as-good-as-new after a maintenance action?
16. Is the operating console well conceived with respect to human factors? Have man-machine aspects been sufficiently considered?
17. Have all safety aspects for operating and maintenance personnel been considered? Also in the case of failure (FMEA/FMECA, FTA, etc.)?

4.3 Maintainability Aspects in Design Reviews

Design reviews are important to point out, discuss, and eliminate *design weaknesses.* As a part of *configuration management,* they are described in Table A3.2. To be effective, design reviews must be supported by *project-specific checklists.* Table 4.3 gives a catalog of questions which can be used to generate project-specific checklists for maintainability aspects in design reviews (see Table 2.8 for reliability and Appendix A4 for other aspects).

4.4 Predicted Maintainability

Knowing the reliability structure of a system and the reliability and maintainability of its elements, it is theoretically possible to calculate the maintainability of the system considered as a one-item structure (e.g. calculating the reliability function and the point availability at system level and extracting $g(t)$ as the density of the repair time at the system level using Eq. (6.18)). However, such a calculation soon becomes laborious for arbitrary systems (Chapter 6). For many practical applications it is often sufficient to know the *mean time to repair* at the system level $MTTR_S$ (expected value of the repair time at the system level) as a function of the system reliability structure, and of the mean time to failure $MTTF_i$ and mean time to repair $MTTR_i$ of its elements. Such a calculation is discussed in Section 4.4.1. Section 4.4.2 deals then with the calculation of the *mean time to preventive maintenance* at system level $MTTPM_S$. The method used in Sections 4.4.1 and 4.4.2 is easy to understand and delivers mathematically exact results for $MTTR_S$ and $MTTPM_S$. Use of statistical methods to estimate or demonstrate a maintainability or a *MTTR* are discussed in Sections 7.2.1, 7.3 and 7.5.

4.4.1 Calculation of $MTTR_S$

Let us first consider a *system without redundancy,* with elements $E_1, \ldots, E_n$ in series as given in Fig. 6.4. $MTTF_i$ and $MTTR_i$ are the *mean time to failure* and the *mean time to repair* of element E_i, respectively $(i = 1, \ldots, n)$. Assume now that each element works the for same *cumulative operating time* T and let T be *arbitrarily large*. In this case, the expected value (mean) of the number of failures of element E_i during T is given by (Eq. (A7.27))

$$\frac{T}{MTTF_i}.$$

The mean of the total repair time necessary to restore the $T/MTTF_i$ failures follows then from

$$MTTR_i \frac{T}{MTTF_i}.$$

For the whole system, there will be in *mean*

$$\sum_{i=1}^{n} \frac{T}{MTTF_i} \tag{4.4}$$

failures and a *mean* total repair time of

$$\sum_{i=1}^{n} MTTR_i \frac{T}{MTTF_i}. \tag{4.5}$$

From Eqs. (4.4) and (4.5) it follows then for the *mean time to repair* at the system level $MTTR_S$ the final value

$$MTTR_S = \frac{\sum_{i=1}^{n} \frac{MTTR_i}{MTTF_i}}{\sum_{i=1}^{n} \frac{1}{MTTF_i}}. \tag{4.6}$$

Equation (4.6) gives the *mathematically exact* value for the mean system repair time $MTTR_S$ under the assumption that at system down (during a repair) no further failures can occur and that switching is ideal (no influence on the reliability). From Eq. (4.6) one can easily verify that

$$MTTR_S = MTTR$$

when $MTTR_1 = \ldots = MTTR_n = MTTR$,

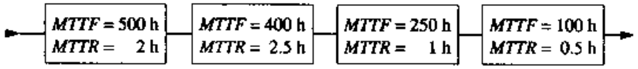
and

$$MTTR_S = \frac{1}{n} \sum_{i=1}^{n} MTTR_i$$

when $MTTF_1 = \ldots = MTTF_n$.

Example 4.1

Compute the mean time to repair at system level $MTTR_S$ for the following system.



How large is the mean of the *total system down time* during the interval $(0, t]$ for $t \to \infty$?

Solution

From Eq. (4.6) it follows that

$$MTTR_S = \frac{\frac{2\,\mathrm{h}}{500\,\mathrm{h}} + \frac{2.5\,\mathrm{h}}{400\,\mathrm{h}} + \frac{1\,\mathrm{h}}{250\,\mathrm{h}} + \frac{0.5\,\mathrm{h}}{100\,\mathrm{h}}}{\frac{1}{500\,\mathrm{h}} + \frac{1}{400\,\mathrm{h}} + \frac{1}{250\,\mathrm{h}} + \frac{1}{100\,\mathrm{h}}} = \frac{0.01925}{0.0185\,\mathrm{h}^{-1}} \approx 1.04\,\mathrm{h}.$$

The mean down time at the system level is also 1.04 h then for a *system without redundancy* it holds that down time = repair time. The *mean operating time* at the system level in the interval $(0, t]$ can be obtained from the expression for the average availability AA_S (Eqs. (6.23), (6.24), (6.48), and (6.49))

$$\lim_{t \to \infty} \mathrm{E}[\text{total operating time in } (0, t]] = t \cdot AA_S = t \frac{MTTF_S}{MTTF_S + MTTR_S}.$$

From this, the mean of the *total system down time* during $(0, t]$ for $t \to \infty$ follows then from

$$\lim_{t \to \infty} \mathrm{E}[\text{total system down time in } (0, t]] = t(1 - AA_S) = t \frac{MTTR_S}{MTTF_S + MTTR_S}.$$

Numerical computation then leads to

$$t \frac{MTTR_S}{MTTF_S + MTTR_S} \approx t \frac{MTTR_S}{MTTF_S} = t \cdot 0.0185 \cdot 1.04 \approx 0.019\,t.$$

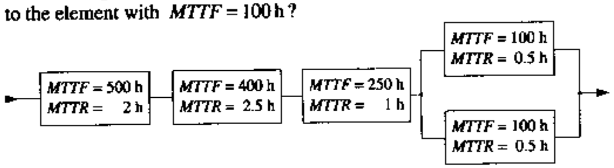
If every element exhibits a constant failure rate λ_i, then $MTTF_i = 1/\lambda_i$ and Eq. (4.6) yields

$$MTTR_S = \frac{\sum_{i=1}^{n} \lambda_i MTTR_i}{\sum_{i=1}^{n} \lambda_i} = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_S} MTTR_i, \qquad \text{with} \qquad \lambda_S = \sum_{i=1}^{n} \lambda_i. \tag{4.7}$$

Equations (4.6) and (4.7) also hold approximately for systems with redundancy. However, in this case, a distinction at system level between *repair time* and *down time* is necessary. If the system contains only *active redundancy*, the *mean time* *to repair* at the system level $MTTR_S$ is given by Eq. (4.6) or (4.7) by summing over *all elements* of the system, as if they were in series. By assuming that failures of redundant elements are repaired without interruption of operation at the system level, Eq. (4.6) or (4.7) can be also used to obtain an *approximate value* of the *mean down time* at the system level by summing only over *all elements without redundancy* (series elements), see Example 4.2.

Example 4.2

How does the $MTTR_S$ of the system in Example 4.1 change, if an active redundancy is introduced to the element with $MTTF = 100\,\text{h}$?



Under the assumption that the redundancy is repaired without interruption of operation at the system level, is there a difference between the *mean time to repair* and the *mean down time* at the system level?

Solution

Because of the assumed active redundancy, the operating elements and the reserve elements show the same mean number of failures. The mean system repair time follows then from Eq. (4.6) by summing over all system elements, yielding

$$MTTR_S = \frac{\frac{2\,\text{h}}{500\,\text{h}} + \frac{2.5\,\text{h}}{400\,\text{h}} + \frac{1\,\text{h}}{250\,\text{h}} + \frac{0.5\,\text{h}}{100\,\text{h}} + \frac{0.5\,\text{h}}{100\,\text{h}}}{\frac{1}{500\,\text{h}} + \frac{1}{400\,\text{h}} + \frac{1}{250\,\text{h}} + \frac{1}{100\,\text{h}} + \frac{1}{100\,\text{h}}} = \frac{0.02425}{0.0285\,\text{h}^{-1}} \approx 0.85\,\text{h}\,.$$

However, the system down time differs now from the system repair time. Assuming for the redundancy an availability equal to one (for constant failure rate $\lambda = 1/MTTF$, constant repair rate $\mu = 1/MTTR$, and one repair crew, Table 6.6 gives for the 1-out-of-2 active redundancy $PA = AA = \mu(2\lambda + \mu)/(2\lambda(\lambda + \mu) + \mu^2)$ yielding $AA \approx 0.99995$), the system down time is defined by the elements in series on the reliability block diagram, thus

$$\text{mean down time at system level} \approx \frac{\frac{2\,\text{h}}{500\,\text{h}} + \frac{2.5\,\text{h}}{400\,\text{h}} + \frac{1\,\text{h}}{250\,\text{h}}}{\frac{1}{500\,\text{h}} + \frac{1}{400\,\text{h}} + \frac{1}{250\,\text{h}}} = \frac{0.01425}{0.0085\,\text{h}^{-1}} \approx 1.68\,\text{h}\,.$$

Similarly to Example 4.1, the mean of the system down time during the interval $(0, t]$ follows then from

$$\lim_{t \to \infty} \mathrm{E}[\text{total down time in } (0, t]] = t(1 - AA_s) \approx t\,\frac{MTTR_S}{MTTF_S} = t \cdot 0.0085 \cdot 1.68 \approx 0.014\,t\,.$$

4.4.2 Calculation of *MTTPM$_S$*

Based on the results of Section 4.4.1, the calculation of the *mean time to preventive maintenance* at system level $MTTPM_S$ can be performed for the following two situations:

1. Preventive maintenance is carried out at once for the *entire system*, one element after the other. If the system consists of elements $E_1, \ldots, E_n$ (arbitrarily grouped on the reliability block diagram) and the mean time to preventive maintenance of element E_i is $MTTPM_i$, then

$$MTTPM_S = \sum_{i=1}^{n} MTTPM_i\,. \tag{4.8}$$

2. Every element E_i of the system is serviced for preventive maintenance *independently* of all other elements and has a mean time to preventive maintenance $MTTPM_i$. In this case, Eq. (4.6) can be used with $MTBPM_i$ instead of $MTTF_i$ and $MTTPM_i$ instead of $MTTR_i$, where $MTBPM_i$ is the *mean time between preventive maintenance* for the element E_i.

Case 2 only has a practical significance when preventive maintenance can be performed without interruption of the operation at the system level. If the preventive maintenance is periodically performed at times T_{PM}, $2T_{PM}$, …, Eq. (4.8) delivers $n \cdot MTTPM_S$ for the case 2 above.

4.5 Basic Models for Spare Part Provisioning

Spare part provisioning is important for systems with long *useful life* or when short repair times or a considerable independence from the manufacturer is required (spare part is also used here for *last repairable unit* (LRU)). Basically, a distinction is made between centralized and decentralized logistical support. Also it is important to take into account whether spare parts are repairable or not.

4.5.1 Centralized Logistical Support, Nonrepairable Spare Parts

In *centralized logistical support*, spare parts are stocked at *one place*. The basic problem can in this case be formulated as follows:

At time $t = 0$, the first spare part is put into operation, it fails at time $t = \tau_1$ and is replaced (in a negligible time) by a second spare part which fails at time $t = \tau_1 + \tau_2$ and so forth; to be determined is the number n of spare parts which must be stocked in order that the requirement for parts during the cumulative operating time T is met with a prescribed probability γ.

To answer this question, the *smallest integer* n must be found for which

$$\Pr\{\tau_1 + \ldots + \tau_n > T\} \geq \gamma . \tag{4.9}$$

In general, $\tau_1, \ldots, \tau_n$ are assumed to be independent positive random variables with the same distribution function F(t), density f(t), and mean $\mathrm{E}[\tau_i] = \mathrm{E}[\tau] = MTTF$. If the number of spare parts is computed from

$$n = \frac{T}{MTTF}, \tag{4.10}$$

the requirement can only be covered (for T large) with a probability of 0.5. Thus, more than $T / MTTF$ spare parts are necessary to meet the requirement with a given probability $\gamma > 0.5$.

According to Eq. (A7.13), the probability as per Eq. (4.9) can be expressed by the ($n-1$)th convolution of the distribution function F(t) with itself, i.e.

$$\Pr\{\tau_1 + \ldots + \tau_n > T\} = 1 - \mathrm{F}_n(T)$$

with

$$\mathrm{F}_1(T) = \mathrm{F}(T) \qquad \text{and} \qquad \mathrm{F}_n(T) = \int_0^T \mathrm{F}_{n-1}(T-x)\mathrm{f}(x)dx . \tag{4.11}$$

Of the distribution functions F(t) used in reliability theory, a simple form for the function $\mathrm{F}_n(t)$ exists only for the *exponential*, *gamma*, and *normal* distribution functions (yielding a Poisson, gamma, and normal distribution, respectively). In particular, $\mathrm{F}(t) = 1 - e^{-\lambda t}$ leads to (Eq. (A7.38))

$$\Pr\{\sum_{i=1}^{n} \tau_i > T\} = \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t}. \tag{4.12}$$

The important case of the *Weibull distribution* $\mathrm{F}(t) = 1 - e^{-(\lambda t)^\beta}$ must be solved numerically. Figure 4.3 shows the results with γ and β as parameters [4.2].

For large values of n, an *approximate solution* for a wide class of distribution functions F(t) can be obtained using the *central limit theorem*. From Eq. (A6.148) if follows that (for $\mathrm{Var}[\tau] < \infty$)

$$\lim_{n \to \infty} \Pr\{\frac{\sum_{i=1}^{n}(\tau_i - \mathrm{E}[\tau])}{\sqrt{n \mathrm{Var}[\tau]}} > x\} = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{y^2}{2}} dy \tag{4.13}$$

and thus, using $x\sqrt{n \mathrm{Var}[\tau]} + n\mathrm{E}[\tau] = T$,

$$\lim_{n \to \infty} \Pr\{\sum_{i=1}^{n} \tau_i > T\} = \frac{1}{\sqrt{2\pi}} \int_{\frac{T - n\mathrm{E}[\tau]}{\sqrt{n \mathrm{Var}[\tau]}}}^{\infty} e^{-\frac{y^2}{2}} dy = \gamma . \tag{4.14}$$

Setting $(T - n\mathrm{E}[\tau]) / \sqrt{n \mathrm{Var}[\tau]} = -d$ it follows then that

$$n = [\frac{d\kappa}{2} + \sqrt{(\frac{d\kappa}{2})^2 + \frac{T}{\mathrm{E}[\tau]}}]^2, \qquad \text{with} \qquad \kappa = \frac{\sqrt{\mathrm{Var}[\tau]}}{\mathrm{E}[\tau]}, \tag{4.15}$$

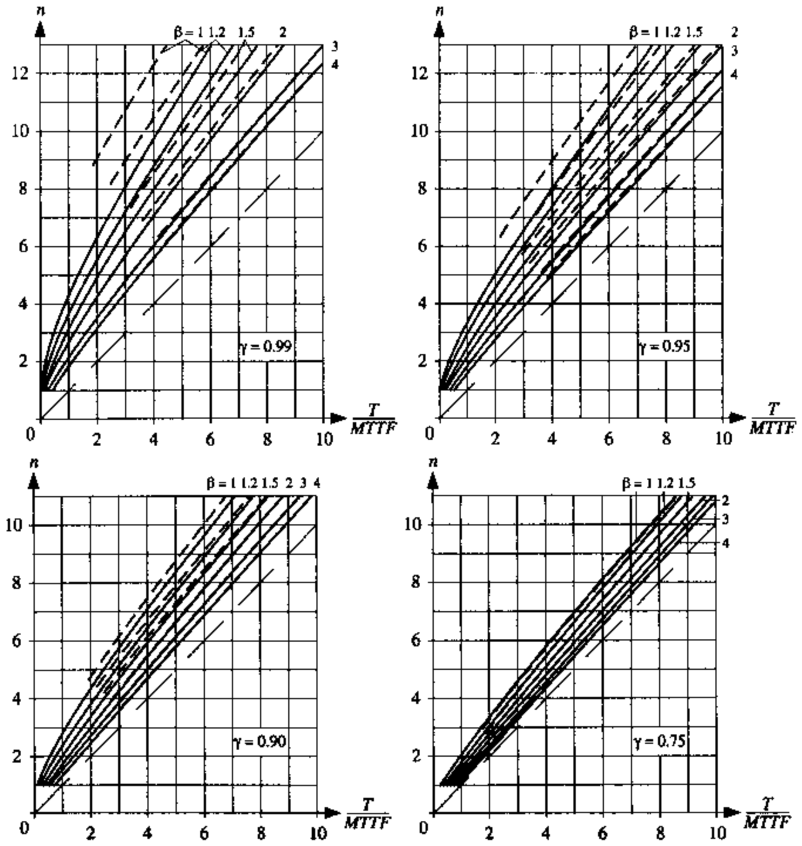


Figure 4.3 Number of spare parts n which are necessary to cover a total cumulative operating time T with a probability γ, i.e. $\Pr\{\tau_1 + \ldots + \tau_n > T\} \geq \gamma$ with $\Pr\{\tau_i \leq t\} = 1 - e^{-(\lambda t)^\beta}$ and $MTTF = \Gamma(1 + 1/\beta)/\lambda$ (the results given by the central limit theorem are dashed)

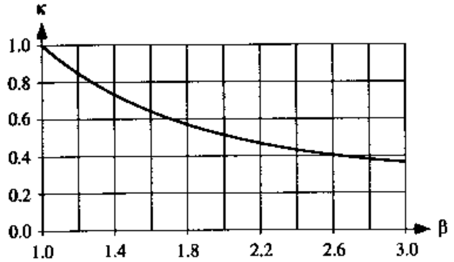


Figure 4.4 Coefficient of variation for the Weibull distribution $(1 \le \beta \le 3)$

where d is the γ quantile of the *standard normal distribution* $\Phi(d) = \gamma$, as $\Phi(-d) = 1 - \Phi(d) = 1 - \gamma$. From Table A9.1 the following values can be taken

$\gamma =$	0.99	0.95	0.90	0.75	0.5
$d =$	2.33	1.64	1.28	0.67	0

Equation (4.15) gives for $\gamma < 0.95$ a good *approximation* of the number of spare parts n, down to low values of n. The quantity $\kappa = \sqrt{\mathrm{Var}[\tau]}\,/\,\mathrm{E}[\tau]$ is the *coefficient of variation*. $\kappa = 1$ for the exponential distribution and

$$\kappa = \sqrt{\frac{\Gamma(1+2/\beta)}{(\Gamma(1+1/\beta))^2} - 1} \tag{4.16}$$

for the Weibull distribution, see Fig. 4.4.

For the case of a Weibull distribution with $\beta \ge 1$, approximate values for n obtained using the central limit theorem (Eq. (4.15)) are shown *dashed* in Fig. 4.3. Deviation from the exact value is ≤ 1 for $\gamma < 0.95$ and $n \ge 5$, this deviation drops off rapidly for increasing values of β ($\mathrm{F}_n(t)$ already approaches a normal distribution for small n). From Eq. (4.14) one recognizes that for $\gamma = 0.5$, $T - n\mathrm{E}[\tau] = 0$ and thus, for n large, $n = T/\mathrm{E}[\tau]$.

Let us now consider the case in which the same spare part (last repairable unit) *occurs k times* in the system under consideration. For $\mathrm{F}(t) = 1 - e^{-\lambda t}$, one can use Eqs. (4.12) and (4.15) with

$$\lambda' = k\lambda \tag{4.17}$$

instead of λ, and $\mathrm{E}[\tau] = 1/k\lambda$ instead of $1/\lambda$, respectively ($\kappa = 1$). This is true because the *sum of independent Poisson processes* is again a Poisson process (Example 7.7, Eq. (7.26)). The situation presented here also corresponds to the case in which k systems use the same spare part (one per system) and *storage is centralized.*

Example 4.3

A spare part with a constant failure rate $\lambda = 10^{-3}\,\mathrm{h}^{-1}$ is used three times in a system $(k = 3)$. Determine the minimum number of spare parts which must be stored to cover a cumulative operating time $T = 10{,}000\,\mathrm{h}$ with a probability $\gamma \ge 0.90$.

Solution

Considering that $kT\lambda = 30$, the exact solution is given by the smallest integer n for which

$$\sum_{i=0}^{n-1} \frac{30^i}{i!} e^{-30} \ge 0.9.$$

From Table A9.2 it follows, for $q = 1 - 0.9 = 0.1$ and $t_{\nu,q} = 2 \cdot 30 = 60$, the value $\nu = 2n \approx 75.2$ (lin. interpolation); thus, $\nu = 76$ and therefore $n = 38$. If 3 parts are operating in the system, 35 spare parts have to be stored. The same result is obtained with Fig.7.3 ($m = 30$ and $\gamma = 0.9$, yielding $c = 37 = n - 1$). The approximate solution according to Eq. (4.15), with $\kappa = 1$ and $d = 1.28$, yields also $n = 38$ $([0.64 + \sqrt{0.64^2 + 30}\,]^2 \approx 37.9)$.

4.5.2 Decentralized Logistical Support, Nonrepairable Spare Parts

For users who have the same equipment or system located in different places, spare parts are often stored *decentralized*, in the extreme case separately at each location (possible reasons for such a strategy may be considerations of logistical time for transportation, independence of the different locations, etc.). If there are l systems, each with a given spare part, and the storage of spare parts is decentralized at each system (or location), then a first step could be to store, with each system, the same number of spare parts n obtained by solving Eq. (4.9) or for n large Eq. (4.15). In this case, the total number of spare parts would be nl. This number of spare parts, which would be sufficient to meet, with a probability $> \gamma$ (often $>> \gamma$) the needs of the l systems with *centralized storage*, would now in general be too small to meet all the individual needs at each location. In fact, assuming that failures at each location are independent, and that with n spare parts the probability of meeting the needs at any location individually is γ, then the probability of meeting the need at all locations is γ^l. Thus, to meet the need at the l locations with a probability γ

$$n_{dec} = l\, n_l \tag{4.18}$$

spare parts are required, where n_l is calculated for each location individually with $\gamma_l = \sqrt[l]{\gamma}$. To make a comparison between a centralized and a decentralized logistical support, let us assume for the spare parts a constant failure rate λ and $\lambda T >> d^2/4$ $(\lambda T > 20d^2/4)$. In this case Eq. (4.15) leads to

$$n \approx \lambda T + d\sqrt{\lambda T}, \qquad \lambda T >> d^2/4, \qquad \text{probability } \gamma. \tag{4.19}$$

For *centralized logistical support*, Eqs. (4.17) and (4.19) yield

$$n_{cen} \approx l\lambda T + d\sqrt{l\lambda T}, \qquad \lambda T >> d^2/4, \qquad \text{probability } \gamma. \tag{4.20}$$

For *decentralized logistical support*, Eqs. (4.18) and (4.19) yield

$$n_{dec} \approx l(\lambda T + d_l\sqrt{\lambda T}), \qquad \lambda T >> d_l^2/4, \qquad \text{probability } \gamma, \tag{4.21}$$

where d_l is obtained as for Eq. (4.14) with $\gamma_l = \sqrt[l]{\gamma}$ instead of γ (for example, $d = 1.64$ for $\gamma = 0.95$ and $d_l = 2.57$ for $l = 10$ i.e. $\gamma_l \approx 0.9949$, see Table A9.1). From the above considerations it follows that for $\lambda T >> d_l^2/4$

$$\frac{n_{dec}}{n_{cen}} \approx \frac{1 + d_l/\sqrt{\lambda T}}{1 + d/\sqrt{l\lambda T}}, \qquad \text{with } \Phi(d) = \gamma \text{ and } \Phi(d_l) = \sqrt[l]{\gamma} \quad \text{(Table A9.1)}. \tag{4.22}$$

Example 4.4

Let $\lambda = 10^{-4}\,\text{h}^{-1}$ be the constant failure rate of a spare part in a given system. The user has 6 locations ($l = 6$) and would like to achieve a cumulative operating time $T = 50{,}000\,\text{h}$ at each location with a probability $\gamma \geq 0.95$. How many spare parts could be saved if the user would store all spare parts at the same location ?

Solution

According to Fig. 4.3 ($T/MTTF = 5$ and $\gamma_l = \sqrt[6]{0.95} \approx 0.99$), from Fig 7.3 ($m = 5$ and $\gamma_l \approx 0.99$, yielding $c = 11 = n-1$), or from a χ^2-Table like Table A9.2 ($t_{\nu,q} = 10$ and $q = 1 - 0.99 = 0.01$, yielding $\nu = 24 = 2n$) each user would need $n = 12$ spare parts (14 using Eq. (4.15) with $d = 2.33$); thus, $n_{dec} = 6 \cdot 12 = 72$. Combining the storage ($l = 6$), one obtains $n_{cen} = 40$, from Fig. 7.3 ($m = 30$ and $\gamma = 0.95$, yielding $c = 39 = n-1$) or Table A9.2 ($t_{\nu,q} = 60$ and $q = 0.05$, yielding $\nu = 80 = 2n$); $n_{cen} = 41$ using Eq. (4.15) with $\lambda T = 30$ and $d = 1.64$.

4.5.3 Repairable Spare Parts

In Sections 4.5.1 and 4.5.2 it was assumed that the spare parts (last repairable units) were *nonrepairable*, i.e. that a new spare part was necessary at each failure. In many cases, spare parts can be repaired and then stored for *reuse*. Calculation of the number of spare parts which should be stored can be performed in a way similar to the investigation of a *k-out-of-n standby redundancy*, where k is the number of identical spare parts used in the system (same as in Eq. (4.17)) and n is the smallest integer *to be determined* such that the requirement is met with a prescribed probability γ. Following two cases should be distinguished:

1. γ is the probability that a request for a spare *at an arbitrary time point* can be met without time delay; in this case, γ can be considered as the *point availability* PA_S (in steady state to simplify the investigation) and n is the smallest integer such that $PA_S \geq \gamma$ for a given γ.

---

2. γ is the probability that any request for a spare part *during the time interval* $(0, t]$ will be met without time delay; in this case, γ can be considered as the *reliability function* $R_{S0}(t)$ and n is the smallest integer such that $R_{S0}(t) \geq \gamma$ for given (fixed) γ and t.

If the *spare parts* have a *constant failure rate* λ and a *constant repair rate* μ, *birth-and-death processes* can be used (Section A7.5.4), see also Section 6.5.1 for the situation in which only one spare part at a time can be repaired and no further failures are considered when a request for a spare part cannot be met. As in Sections 6.3 to 6.6, the results obtained with these assumptions apply with a good approximation if $k\lambda << \mu$ holds. These assumptions will be considered here to simplify the investigation. For case 1 above, Eqs. (6.138), with $\lambda_r \equiv 0$, and (6.140) yield

$$PA_S = \sum_{j=0}^{n-k} P_j = 1 - P_{n-k+1} \geq \gamma \tag{4.23}$$

with

$$P_j = \frac{\pi_j}{\sum\limits_{i=0}^{n-k+1} \pi_i} \qquad \text{and} \qquad \pi_i = (\frac{k\lambda}{\mu})^i, \quad i = 0, \ldots, n-k+1. \tag{4.24}$$

The problem is thus to find, for a given k and γ, the smallest integer n which satisfies Eq. (4.23). Often $n = k+1$ (*one spare part*) or $n = k+2$ (*two spare parts*) will be *sufficient*. In these cases the results of Table 6.8 can be used, yielding

$$PA_{S1} = \frac{k\lambda\mu + \mu^2}{k^2\lambda^2 + k\lambda\mu + \mu^2} \approx 1 - (\frac{k\lambda}{\mu})^2, \qquad \text{for } n-k=1 \quad \text{(1 spare part)}, \tag{4.25}$$

$$PA_{S2} = \frac{k^2\lambda^2\mu + k\lambda\mu^2 + \mu^3}{k^3\lambda^3 + k^2\lambda^2\mu + k\lambda\mu^2 + \mu^3} \approx 1 - (\frac{k\lambda}{\mu})^3, \qquad \text{for } n-k=2 \quad \text{(2 spare parts)}. \tag{4.26}$$

The reliability function is obtained from Eq. (6.144) for $n = k+1$ and Eq. (6.145) for $n = k+2$, respectively. Considering the quantities ν_j as per Eq. (6.138), with $\lambda_r \equiv 0$ (standby redundancy), yields

$$R_{S0_1}(t) \approx e^{-t/MTTF_{S0_1}}, \qquad \text{with } MTTF_{S0_1} \approx \frac{\mu}{(k\lambda)^2}, \qquad \text{for } n-k=1, \tag{4.27}$$

$$R_{S0_2}(t) \approx e^{-t/MTTF_{S0_2}}, \qquad \text{with } MTTF_{S0_2} \approx \frac{\mu^2}{(k\lambda)^3}, \qquad \text{for } n-k=2. \tag{4.28}$$

Example 4.5

A system contains $k = 100$ identical spare parts with a constant failure rate $\lambda = 10^{-5}\,\mathrm{h}^{-1}$ and which can be repaired with a constant repair rate $\mu = 10^{-1}\,\mathrm{h}^{-1}$. (*i*) Determine the number of spare parts which must be stored in order to meet without any time delay and with a probability $\gamma \geq 0.99$ a request for a spare part at an arbitrary time point t (consider the steady-state only, one repair crew, and no further failure when a request for a spare part cannot be met). (*ii*) If one spare part is stored ($n = k + 1$), how large is the probability that any request for a spare part during the time interval $(0, 10^4\,\mathrm{h}]$ will be met without any time delay?

Solution

(*i*) Taking $n = k + 1 = 101$, Eq. (4.25) yields

$$PA_S = \frac{100 \cdot 10^{-5} \cdot 10^{-1} + 10^{-2}}{10^4 \cdot 10^{-10} + 100 \cdot 10^{-5} \cdot 10^{-1} + 10^{-2}} = \frac{1.01}{1.0101} \approx 1 - (\frac{100 \cdot 10^{-5}}{10^{-1}}) \approx 0.9999.$$

Thus only one spare part must be stored.

(*ii*) For $n = k + 1$, Eq. (4.28) yields $\mathrm{R}_{S0_1}(t) \approx e^{-0.0000098t}$ and thus $R_{S0}(10^4) \approx e^{-0.098} \approx 0.91$.

Assuming, for *comparative investigations*, that each spare part can be repaired *independently* from each other (*n*-*k*+1 repair crews instead of 1 repair crew), the results of Section A7.5.4, with $\nu_i = k\lambda$, $i = 0, \ldots, n-k$ and $\theta_i = i\mu$, $i = 1, \ldots, n-k+1$, yield

$$PA_{S_1} \approx 1 - (k\lambda/\mu)^2/2 \qquad \text{and} \qquad PA_{S_2} \approx 1 - (k\lambda/\mu)^3/3. \tag{4.29}$$

For the reliability function it holds that (ν_i as before and $\theta_i = i\mu$, $i = 1, \ldots, n-k$)

$$R_{S0_1}(t) \approx e^{-t(k\lambda)^2/\mu} \qquad \text{and} \qquad R_{S0_2}(t) \approx e^{-t(k\lambda)^3/2\mu^2}. \tag{4.30}$$

4.5.4 Cost Considerations

The investigations of Sections 4.5.1 to 4.5.3, in particular those of Section 4.5.2 on *decentralized logistical support*, can be extended to cover the more general case of systems with *different spare parts*. In many practical situations, spare parts provisioning has to be considered as a parameter in the *optimization* between performance, reliability, availability, logistical support, and *cost*, taking care also of *obsolescence* aspects. In some cases, one parameter is given (for example cost) and the best logistical structure is sought to maximizes system availability or minimizes system down time. Even assuming constant failure and repair rates, analytical solution of such problems is time consuming, see e.g. [4.22] for a computer program.

---

5 Design Guidelines for Reliability, Maintainability, and Software Quality

Reliability, maintainability, and software quality *have to be built into an equipment or system* during the design and development phase. This has to be supported by *analytical investigations* (Chapters 2, 4, and 6) as well as by *design guidelines*. Adherence to such guidelines limits the influence of those aspects or effects which can invalidate the models assumed for analytical investigations, and contributes greatly to build in reliability, maintainability, and software quality. This chapter gives a comprehensive list of design guidelines for reliability, maintainability, and software quality of equipment and systems, as used in industry.

5.1 Design Guidelines for Reliability

Reliability analysis in the design and development phase (Chapter 2) gives an estimate of an item's true reliability, based on some assumptions regarding data used, interface problems, dependence between components, compatibility between materials, environmental influences, transients, EMC, ESD, etc., as well as on the quality of manufacture and the user's skill level. To deal exhaustively with all these aspects is often difficult. However, the following design guidelines can help to improve the inherent reliability of complex equipment and systems.

5.1.1 Derating

Thermal and electrical stresses largely influence the failure rate of electronic components. *Derating* these stresses is mandatory to improve the inherent reliability of equipment and systems. Table 5.1 gives recommended *stress factors* S (Eq. (2.1)) to be used for an ambient temperature θ_A up to 40°C. For $\theta_A > 40$°C,

Table 5.1 Recommended derating values for electronic components at ambient temperature $\theta_A \leq 40°\mathrm{C}$

Component	Power	Voltage	Current	Internal Temperature	Frequency
Resistors					
• Fixed	0.6			0.8	
• Variable	0.6			0.7	
• Thermistors	0.4			0.7	
Capacitors					
• Film, Ceramic		0.5		0.5	
• Ta (solid)		0.5		0.5	
• Al (wet)		0.8		0.5	
Diodes					
• Gen. purpose		0.5^*	0.6	0.7	
• Zener	0.6			0.7	
Transistors		0.5^*	0.7	0.7	$0.1 f_T$
Thyristors, Triacs		0.6^*	0.6	0.7	
Optoel. devices		0.5^{**}	0.5	0.8	
ICs					
• Linear		0.7	0.8^+	0.7^x	0.9
• Voltage reg.			0.7^+	0.7^x	
• Digital bipolar			0.8^+	0.7^x	
• Digital MOS			0.8^+	0.7^x	0.9
Coils, Transf.	0.5				
Switches, Relays			$0.4–0.7^{++}$	0.7	0.5
Connectors		0.7	0.6	0.8	0.5

*breakdown voltage, **isolation voltage (0.7 for U_{in}),
$^+$sink current, $^{++}$low values for inductive loads, ^{x} $\theta_J \leq 100°\mathrm{C}$

a further reduction of S is necessary (in general, linearly up to the limit temperature, as shown in Fig. 2.3). Too low value of S ($S < 0.1$) can also cause problems. $S = 0.1$ can be used in many cases to compute the failure rate in a standby or *dormant state*.

5.1.2 Cooling

As a general rule, the *junction temperature* θ_J of semiconductor devices should be kept as near as possible to the *ambient temperature* θ_A of the equipment or system

in which they are used. For a good design, $\theta_J < 100°\mathrm{C}$ is recommended. In a steady-state situation, i.e. with constant power dissipation P, the following relationships

$$\theta_J = \theta_A + R_{JA}\,P \tag{5.1}$$

or

$$\theta_J = \theta_A + (R_{JC} + R_{CS} + R_{SA})\,P \tag{5.2}$$

can be established and used to define the thermal resistances

R_{JA}	junction – ambient	R_{JC}	junction – case
R_{CS}	case – surface	R_{SA}	surface – ambient,

where *surface* is used for *heat sink.*

Example 5.1

Determine the thermal resistance R_{SA} of a heat sink by assuming $P = 400\,\mathrm{mW}$, $\theta_J = 70°\mathrm{C}$, and $R_{JC} + R_{CS} = 35°\mathrm{C/W}$.

Solution

From Eq. (5.2) it follows that

$$R_{SA} = \frac{\theta_J - \theta_A}{P} - R_{JC} - R_{CS} \qquad \text{and thus} \qquad R_{SA} = \frac{30°\mathrm{C}}{0.4\,\mathrm{W}} - 35°\mathrm{C/W} = 40°\mathrm{C/W}.$$

For many practical applications, thermal resistances can be assumed to be independent of the temperature. However, R_{JC} generally depends on the *package* used (lead frame, packaging form and type), R_{CS} varies with the kind and thickness of *thermal compound* between the device package and the heat sink (or device support), and R_{SA} is a function of the heat-sink dimensions and form as well as of the type of cooling used (free convection, forced air, liquid-cooled plate, etc.). Typical *thermal resistance values* R_{JC} and R_{JA} for free convection in ambient air without heat sinks are given in Table 5.2. The values of Table 5.2 are *indicative* only and have to be replaced with specific values for exact computations.

Cooling problems should not only be considered locally at the component level, but be integrated into a *thermal design concept*. In defining the layout of an assembly, care must be taken in placing high power dissipation parts *away* from temperature sensitive components like wet Al wet capacitors and some optoelectronic devices (the *useful life* is reduced by a factor of 2 for a $10-20°\mathrm{C}$ increase of the ambient temperature). In placing the assemblies in a rack, the cooling flow should be directed from the parts with low toward those with high power dissipation.

Table 5.2 Typical thermal resistance values for semiconductor component packages

Package form	Package type	R_{JC} [°C/W]	R_{JA} [°C/W]**
DIL	Plastic	10 - 40*	30 - 100*
DIL	Ceramic/Cerdip	7 - 20*	30 - 100*
PGA	Ceramic	6 - 10*	20 - 40*
SOL, SOM, SOP	Plastic (SMT)	20 - 60*	70 - 240*
PLCC	Plastic	10 - 20*	30 - 70*
QFP	Plastic	15 - 25*	30 - 80*
TO	Plastic	70 - 140	200 - 400
TO	Metal	2 - 5	—

JC = junction to case, JA = junction to ambient, *lower values for ≥ 64 pins (144 for PGA & QFP), **free convection at 0.15 m/s (factor 2 lower for forced cooling at 4 m/s)

5.1.3 Moisture

For electronic components in nonhermetic packages, *moisture* can cause *drift* and activate various failure mechanisms such as *corrosion* and *electrolysis* (see Section 3.2.2, Point 8 for considerations on ICs). Critical in these cases is not so much the water itself, but the impurities and gases dissolved in it. If high relative humidity can occur, care must be taken to avoid the formation of galvanic couples as well as condensation or ice formation on the component packages or on conductive parts.

As stated in Section 3.1.3, the *use of ICs in plastic packages* can be allowed if one of the following conditions is satisfied:

1. Continuous operation, relative humidity ≤70%, noncorrosive or marginally corrosive environment, junction temperature ≤100°C, and equipment useful life less than 10 years.
2. Intermittent operation, relative humidity ≤60%, noncorrosive environment, no moisture condensation on the package, junction temperature ≤100°C, and equipment useful life less than 10 years.

For ICs with silicon nitride *glassivation*, intermittent operation holds also for Point 1.

Drying materials should be avoided, in particular if chlorine compounds are present. *Conformal coating* on the basis of acrylic, polyurethane, epoxy, silicone or fluorocarbon resin $25-125\mu m$ thick, filling with gel, or encapsulation in epoxy or similar resins are currently used (attention must be given to thermomechanical stresses during hardening). The use of *hermetic enclosures* for assemblies or equipment should be avoided if condensation cannot be excluded. Indicators for the effects of moisture are an increase of leakage currents or a decrease of insulation resistances.

5.1.4 Electromagnetic Compatibility, ESD Protection

Electromagnetic compatibility (EMC) is the ability of an item to function properly in its intended electromagnetic environment without introducing unacceptable electromagnetic noise (disturbances) into that environment. EMC has thus two aspects, *susceptibility* and *emission*. Agreed susceptibility and emission levels are given in international standards such as IEC 61000 [3.9]. *Electrostatic discharge* (ESD) protection can be considered as a part of an *electromagnetic immunity concept*, which is particularly important for semiconductor devices (Section 3.2.3). Causes of noise (disturbances) in electronic equipment and systems are

- switching and transient phenomena,
- electrostatic discharges,
- stationary electromagnetic fields.

Noise coupling can be

- conductive (galvanic),
- through common impedances,
- by radiated electromagnetic fields.

In the context of ESD or EMC, noise often appears as electrical pulses with rise times in the range 0.1 to 10 kV/ns, peak values of 0.1 to 10 kV, and energies of 0.1 to 10^3 mJ (high values for equipment). EMC aspects, and in particular *ESD protection*, have to be considered early in the design of equipment and systems. The following *design guidelines* can help to avoid problems:

1. For high speed logic circuits ($f \geq 10$ MHz) use a whole *plane* (layer of a *multilayer*) or at least a tight grid for ground and for the power supply, to minimize inductance and to ensure a distributed decoupling capacitance (4 layers as signal/V_{CC}/ground/signal or better 6 layers as shield/signal/V_{CC}/ground/signal/ shield).
2. For low frequency digital circuits, analog circuits, and power circuits use a *single-point ground* concept, and wire all different grounds separately to a *common ground point* at system level (across antiparallel suppressor diodes).
3. Use *low inductance decoupling capacitors* (generally 10 nF ceramic capacitors, placed where spikes may occur, i.e. at every IC for fast logic and bus drivers, every 4 ICs for HCMOS) and a 1μF metallized paper (or a 10μF electrolytic) capacitor per board; in the case of a highly pulsed load, locate the voltage regulator on the same board as the logic circuits.
4. Avoid logic which is *faster* than necessary and ICs with widely *different rise times*; adhere to required rise times (10 ns for AS-TTL and ACMOS, up to 150 ns for HCT and HCMOS) and use Schmitt trigger inputs if necessary; a good design should work properly with rise and fall times t_r, $t_f \leq 20$ ns, as is generally required for clock pulses.

5. Pay attention to *dynamic stresses* (particularly of *breakdown voltages* on semiconductor devices) as well as of *switching phenomena* on inductors or capacitors; implement noise reduction measures near the noise source (preferably with *Zener diodes* or *suppressor diodes*).
6. *Match* signal lines whose length is greater than $v \cdot t_r$, also when using differential transmission (often possible with a series resistor at the source or a parallel resistor at the sink, v=signal propagation speed $\approx c/\sqrt{\varepsilon_r \mu_r}$); for HCMOS also use a 1 to 2 kΩ pull-up resistor and a pull-down resistor equal to the line impedance Z_0, in series with a capacitor of about 200 pF per meter of line.
7. Capture *induced noise* at the beginning and at the end of long signal lines using *parallel suppressors* (suppressor diodes), series protectors (ferrite beads) or series/parallel networks (RC), in that order, taking into account the required rise and fall times.
8. Use *twisted pairs* for signal and return lines (one twist per centimeter); ground the *return line at one end* and the *shield at both ends* for magnetic shielding (at more points to shield against electric fields); provide a closed (360°) contact with the shield for the ground line; clock leads should have adjacent ground returns; for clock signals leaving a board consider the use of fiber optics, coax, tri-leads, or twisted pairs in that order.
9. Avoid *apertures in shielded enclosures* (many small holes disturb less than a single aperture having the same area); use *magnetic material* to shield against low-frequency magnetic fields and materials with good *surface conductivity* against electric fields, plane waves, and high frequency magnetic fields (above 10 MHz, *absorption loss* predominates and shield thickness is determined more for its mechanical rather than for its electrical characteristics); *filter or trap all cables* entering or leaving a shielded enclosure (filters and cable shields should make very low inductivity contacts to the enclosure); RF parts of analog or mixed signal equipment should be appropriately shielded (air core inductors have greater emission but less reception capability than magnetic core inductors); all signal lines entering or leaving a circuit should be investigated for common-mode emission; minimize *common-mode currents*.
10. Implement *ESD current-flow paths* with *multipoint grounds* at least for plug-in populated printed circuit boards (PCBs), e.g. with guard rings, *ESD* networks, or suppressor diodes, making sure in particular that all signal lines entering or leaving a PCB are sufficiently *ESD* protected (360° contact with the shield if shielded cables are used, latched and strobed inputs, etc.); ground to *chassis ground* all exposed metal, if necessary use secondary shields between sensitive parts and chassis; design *keyboards* and other *operating parts* to be immune to *ESD*.

5.1.5 Components and Assemblies

5.1.5.1 Component Selection

1. Pay attention to *all* specification limits given by the manufacturer and to company-specific rules, in particular to dynamic parameters and breakdown limits.
2. Limit the number of entries in the *list of preferred parts* (LPP) and whenever possible ensure a *second source* procurement; if obsolescence problems are possible (very long warranty or operation time), take care of this aspect in the LPP and/or in the design/layout of the equipment or system considered.
3. Use *non-qualified* parts and components only after checking the technology and reliability *risks* involved (the *learning phase* at the manufacturer's plant can take more than 6 months); in the case of critical applications, intensify the *feedback* to the manufacturer and plan an appropriate *incoming inspection with screening*.

5.1.5.2 Component Use

1. Tie *unused logic inputs* to the power supply or ground, usually through *pull-up/pull-down resistors* (100 kΩ for CMOS), also to improve testability; pull-up/pull-down resistors are also recommended for inputs driven by three-state outputs; unused outputs remain basically open.
2. Protect all *CMOS terminals* from or to a *connector* with a 100 kΩ *pull-up/pull-down resistor* and a 1 to 10 kΩ *series resistor* (latch-up) for an *input*, or an appropriate *series resistor* for an *output* (add diodes if V_{in} and V_{out} cannot be limited between $-0.3\,V$ and $V_{DD}+0.3\,V$); observe *power-up* and *power-down sequences*, make sure that the ground and power supply are *applied before* and *disconnected after* the signals.
3. Analyze the *thermal stress* (internal operating temperature) of each part and component carefully, placing dissipating devices away from temperature-sensitive ones and adequately cooling components with high power dissipation (failure rates double generally for a temperature increase of 10 − 20°C); for semiconductor devices, design for a *junction temperature* $\theta_J \le 100°C$ (if possible keep $\theta_J < 80°C$).
4. Pay attention to *transients*, especially in connection with *breakdown voltages* of transistors ($V_{BEO} \le 5\,V$, stress factor $S < 0.5$ for V_{CE}, V_{GS}, and V_{DS}).
5. Derate *power devices* more than signal devices (stress factor $S < 0.4$ if more than 10^5 power cycles occur during the useful life, $N \leqq 10^7 e^{-0.05\,\Delta\theta_J}$).
6. Avoid *special diodes* (tunnel, step-recovery, pin, varactor, which are 2 to 20 times less reliable than normal Si diodes); *Zener diodes* are about one half as reliable as Si switching diodes, their stress factor should be > 0.1.

7. Allow a ±30% drift of the coupling factor for *optocouplers* during operation; regard optocouplers and *LEDs* as having a limited useful life (generally $>10^6$ h for $\theta_J < 40°C$ and $<10^5$ h for $\theta_J > 80°C$), design for $\theta_J \le 70°C$ (if possible keep $\theta_J < 40°C$); pay attention to optocoupler insulation voltage.
8. Observe operating temperature, voltage stress (DC and AC), and technological suitability of *capacitors* for a given application: *foil capacitors* have a reduced impulse handling capability; *Al wet capacitors* have a limited useful life (which halves for every 10°C increase in temperature), a large series inductivity, and a moderately high series resistance; for *solid Ta capacitors* the AC impedance of the circuit as viewed from the capacitor terminals should not be too small (the failure rate is an order of magnitude higher with $0.1\Omega/V$ than with $2\Omega/V$, although new types are less sensitive); use a $10-100$nF *ceramic capacitor* parallel to each electrolytic capacitor; avoid electrolytic capacitors $< 1\mu F$.
9. Cover *EPROM windows* with metallized foils, also when stored.
10. Avoid the use of *variable resistors* in final designs (50 to 100 times less reliable than fixed resistors); for *power resistors*, check the internal operating temperature as well as the voltage stress.

5.1.5.3 PCB and Assembly Design

1. Design all *power supplies* to handle *permanent short* circuits and monitor for under/overvoltage (protection diode across the voltage regulator to avoid $V_{out} > V_{in}$ at power shutdown); use a 10 to 100nF *decoupling ceramic capacitor* parallel to each electrolyte capacitor.
2. Clearly define, and implement, *interfaces* between *different logic families*.
3. Establish *timing diagrams* using *worst-case conditions*, also taking the effects of *glitches* into consideration.
4. Pay attention to *inductive and capacitive coupling* in parallel signal leads ($0.5-1\mu H/m$, $50-100\,pF/m$); place signal leads *near to ground returns* and away from power supply leads, in particular for *clocks*; for high-speed circuits, investigate the requirement for *wave matching* (parallel resistor at the sink, series resistor at the source); introduce *guard rings* or *ground tracks* to limit coupling effects.
5. Place all *input/output drivers* close together and near to the connectors, away from clock circuitry and power supply lines (inputs should be latched and strobed).
6. Protect PCBs against damage through *insertion or removal under power* (use appropriate connectors).
7. For PCBs employing *surface mount technology* (SMT), make sure that the component spacing is not smaller than 0.5mm and that the lead width and spacing are not smaller than 0.25mm; test pads and solder-stop pads should be provided; for large leadless ceramic ICs, use an appropriate lead frame (problems in SMT arise with soldering, heat removal, mismatch of expansion coefficients, pitch dimensions, pin alignment, cleaning, and contamination); pitch < 0.3 mm can give production problems.

8. Observe the *power-up and power-down sequences*, especially in the case of different power supplies (no signals should be applied to unpowered devices).
9. Make sure that the *mechanical fixing* of power devices is appropriate, in particular of those with high power dissipation; avoid having current carrying contacts under thermomechanical stress.
10. The *testability* of PCBs and assemblies should be considered early in the design of the layout (number and dimension of test points, pull-up/pull down resistors, activation/deactivation of three-state outputs, see also Section 5.2); manually extend the capability of CAD tools if necessary.

5.1.5.4 PCB and Assembly Manufacturing

1. Keep the workplaces for assembling, *soldering*, and testing *conductive*, in particular ground tools and personnel with 1MΩ resistors; avoid touching the active parts of components during assembling; use soldering irons with transformers and grounded tips.
2. When using *automatic placing machines* for inserted devices, verify that only the parts of pins free from insulation penetrate into the soldering holes (critical in particular are some resistor networks, capacitors, and relays) and that IC pins are not bent or pressed into the soldering holes (hindering degassing during soldering); for *surface mount devices* (SMD), make sure that the correct quantity of solder material is deposited, and that the stand-off height between the component body and the printed circuit surface is not less than 0.25mm (pitch < 0.3 mm can give production problems).
3. Control the *soldering temperature profile* carefully; for wave soldering choose the best compromise between soldering time and soldering temperature (about 3s at 245°C) as well as an appropriate *preheating* (about 60s to reach 100°C); check the solder bath periodically and make sure that there is sufficient distance between the solder joints and the package for temperature sensitive devices; for *surface mount technology* (SMT) give preference to *IR reflow soldering* and provide good *solder-stop pads* (vapor-phase can be preferred for substrates with metal core or PCBs with high component density); avoid having inserted and surface mounted devices (SMD) on the same (two-sided) PCB (thermal shock on the SMD with consequent crack formation and possibly ingress of flux to the active part of the component, particularly for ceramic capacitors greater than 100 nF and large plastic ICs).
4. Avoid soldering *gold-plated pins;* if not possible, tin-plate the pins in order to reduce the Au concentration to < 4% in the solder joint (intermetallic layers) and < 0.5% in the solder bath (contamination), $0.2\mu m <$ Au thickness $< 0.5\mu m$.

5. Avoid having more than *one heating* process that reaches the soldering temperature, and hence any kind of *rework*; for temperature sensitive devices, consider the possibility of adequate protection during soldering (support, cooling ring, etc.).
6. For high reliability applications, *wash* PCBs and assemblies after soldering (possibly with deionized water ($< 5\mu S/cm$), but in any case with halogen-free liquids); periodically check the *washing liquid* for contamination; use ultrasonic cleaning only when resonance problems in components are excluded.
7. Avoid any kind of *electrical overstress* when testing components, PCBs or assemblies; avoid removal and insertion under power.

5.1.5.5 Storage and Transportation

1. Keep the *storage temperature* between 5 and 30°C and the *relative humidity* between 40 and 60%; avoid dust, corrosive atmospheres, and mechanical stresses (particularly for electromechanical components); use *hermetically sealed containers* for high-humidity environments only.
2. Limit the storage time by implementing *first-in/first-out* rules (storage time should be no longer than two years, just-in-time shipping is often only possible for a stable production line).
3. Ensure *antistatic storage and transportation* of all ESD sensitive electronic components, in particular semiconductor devices (use metallized, unplasticized bags, avoid PVC for bags).
4. Transport PCBs and assemblies in *antistatic containers* and with all connectors shorted.

5.1.6 Particular Guidelines for IC Design and Manufacturing

1. Reduce *latch-up sensitivity* by increasing critical distances, changing local doping, or introducing vertical thick-oxide isolation.
2. Avoid significant *voltage drops* along resistive leads (polysilicon) by increasing line conductivity and/or dimensions or by using *multilayer metallizations*.
3. Give sufficient size to the *contact windows* and avoid large contact depth and thus sharp edges (slopes); ensure material compatibility, in particular with respect to metallization layers.
4. Take into account *chemical compatibility* between materials and tools used in sequential processes; limit the use of *planarization processes* to uncritical metallization line distances; employ preferably *stable processes* (low-risk processes) which allow a reasonable parameter deviation; control carefully the *wafer raw material* (CZ/FZ material, crystal orientation, O_2 concentration, etc.).

5.2 Design Guidelines for Maintainability

As pointed out in Section 4.2, maintainability must be built into equipment or systems. This has generally to be performed project-specific. However, a certain number of *design guidelines for maintainability* apply quite generally. These will be discussed in this section for the case of complex electronic equipment and systems with high maintainability requirements.

5.2.1 General Guidelines

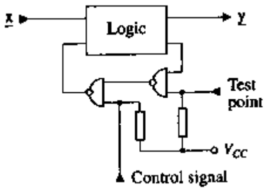
1. Plan and implement a concept for automatic *fault detection* and automatic or semiautomatic *fault isolation* down to the *last repairable unit* (LRU) level, including *hidden failures* and *software defects*, as far as possible.
2. *Partition* the equipment or system into last repairable units (LRUs) and apply techniques of *modular construction*, starting from the *functional structure*; make modules functionally *independent* and electrically as well as mechanically separable; develop easily *replaceable* LRUs which can be tested with commonly available test equipment.
3. Aim for the greatest possible *standardization* of parts, tools, and testing equipment; keep the need for external testing facilities to a minimum.
4. Conceive operation and maintenance *procedures* to be *as simple as possible*, also taking into account personnel safety, describe them in appropriate manuals.
5. Consider *environmental conditions* (thermal, climatic, mechanical) in field operation as well as during transportation and storage.

5.2.2 Testability

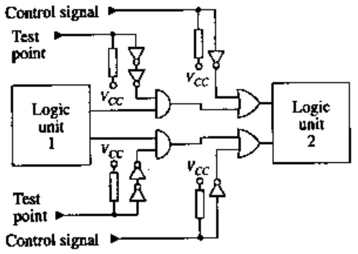
Testability includes the degrees of failure detection and isolation, the correctness of test results, and the test duration. High testability can generally be achieved by improving *observability* (the possibility to check internal signals at the outputs) and *controllability* (the possibility to modify internal signals from the inputs). Of the following design guidelines, the first five are specifically valid for assemblies and the last five are also valid for ICs (ASICs).

1. Avoid *asynchronous logic* (asynchronous signals should be latched and strobed at the inputs).
2. Simplify *logical expressions* as far as possible.
3. Improve testability of connection paths and simple circuitry using ICs with *boundary-scan* (IEEE STD 1149, [4.9]).
4. Separate *analog and digital circuit* paths, as well as circuitry with different supply voltages; make power supplies mechanically separable.

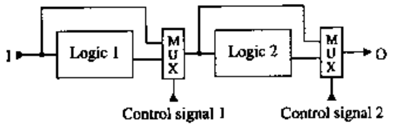
5. Make *feedback paths* separable



6. Realize *modules* as self-contained as possible, with small sequential depth, electrically separable and individually testable,

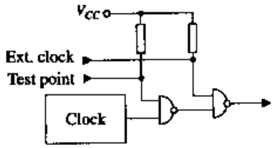


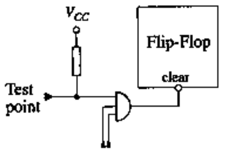
with gates



with MUXs

7. Allow for *external initialization* of sequential logic





8. Develop and introduce *built-in self-test* (BIST); introduce *test modi* also for the detection of hidden failures.
9. Provide enough *test points* (at a minimum on functional-unit inputs and outputs as well as on bus lines) and support them with pull-up/pull-down resistors, provide access for a probe, taking into account the capacitive load (resistive in the case of DC measurements).
10. Make use of a *scan path* to reduce test time; the basic idea of a scan path is shown on the right-hand side of Fig. 5.1, the test procedure with a scan path is as follows ($n = 3$ in Fig. 5.1):
    1. Activate the MUX control signal (connect Z to B).
    2. Scan-in with n clock pulses an appropriate n-bit test pattern, this pattern appears in parallel at the FF outputs and can be read serially with $n-1$ additional clock pulses (repeat this step to completely test MUXs and FFs).

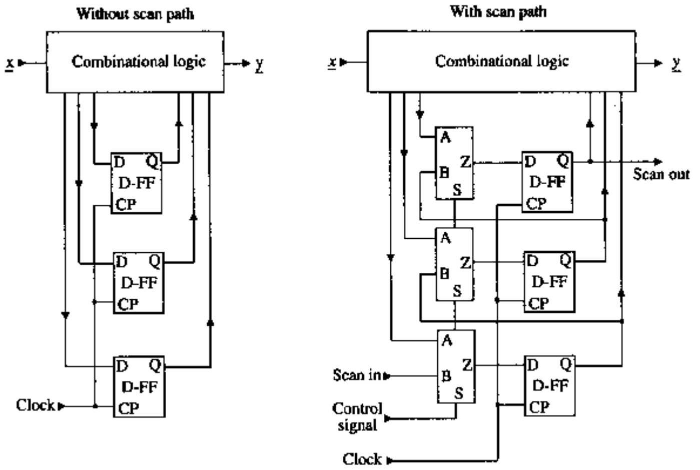


Figure 5.1 Basic structure of a synchronous sequential circuit, without a scan path on the left-hand side and with a scan path on the right-hand side

3. Scan-in with n clock pulses a first test pattern for the combinational logic (feedback part) and apply an appropriate pattern also to the input $\underline{x}$ (both patterns are applied to the combinational circuit and generate corresponding results which appear at the output $\underline{y}$ and at the inputs A of the MUXs).
4. Verify the results at the output $\underline{y}$.
5. Deactivate the MUX control signal (connect Z to A).
6. Give one clock pulse (feedback results from the combinational circuit appear parallel at the FF outputs).
7. Activate the MUX control signal (connect Z to B).
8. Scan-out with $n-1$ clock pulses and verify the results, at the same time a second test pattern for the combinational circuit can be scanned-in.
9. Repeat steps 3 – 8 up to a satisfactory test of the combinational part of the circuit (see e.g. [4.12, 4.18] for test algorithms specially developed for combinational circuits).

5.2.3 Accessibility, Exchangeability

1. Provide *self-latching* access flaps of sufficient size; avoid the need for *special tools* (one-way screws, Allen screws, etc.); use clamp fastening.
2. Plan *accessibility* by considering the frequency of maintenance tasks.

3. Use preferably *indirect plug connectors*; distribute power supply and ground over several contacts (20% of the contacts should be reserved for power supply and ground); plan to have *reserve contacts*; avoid any external mechanical stress on connectors, define (if possible) only one kind of extender for PCBs and plan its use.
4. Provide for *speedy replaceability* by means of plug-out/plug-in techniques.
5. Prevent *faulty installation or connection* (of PCBs in particular) through mechanical keying.

5.2.4 Operation, Adjustment

1. Use high *standardization* in selecting operational tools and make any labeling simple and clear.
2. Consider *human aspects* in the layout of *operating consoles* and in defining operating and maintenance procedures.
3. Order all steps of a procedure in a *logical sequence* and document these steps by a visual feedback.
4. Describe system status, detected fault, or action to be accomplished concisely in *full text*.
5. Avoid any form of hardware *adjustment* (or alignment) in the field, if necessary, carefully describe the relevant procedure.

5.3 Design Guidelines for Software Quality

Software plays an increasingly important role in equipment and systems, both in terms of technical relevance and of development costs (often higher than 50% even for small systems). Unlike hardware, software does not go through a *production phase*. Also, software cannot break or wear out. However, it can *fail* to satisfy its required function because of *defects* which manifest themselves while the system is operating (*dynamic defects*). A *fault* in the software is thus still caused by a *defect*, even if it appears *randomly* in time. Software problems are thus basically *quality problems* and should be solved with *quality assurance tools* (configuration management, testing, and quality data reporting systems). This section *introduces* some tools for *software quality assurance*, with particular emphasis on design guidelines and *preventive actions*. Because of their utility in *debugging* complex software packages, models for *software quality growth* are also briefly discussed.

Table 5.3 Software life-cycle phases

Phase	Objectives / Tasks	Inputs	Outputs
Concept	• Problem definition • Feasibility check	• Problem definition • Constraints on computer size, programming languages, I/O, etc.	• System specifications for functional (what) and performance (how) aspects • Proposal for the definition phase
Definition	• Investigation of alternative solutions • Interface definitions • Second feasibility check	• System specifications • Proposal for the definition phase	• Revised system specifications • Interface specifications • Updated estimation of costs and schedule • Feedback from users • Proposal for the design, coding, and testing phase
Design, Coding, Testing	• Setup of detailed specifications • Software design • Coding • Test of each module • Verification of compliance with module specifications (design reviews) • Data acquisition	• Revised system specifications • Interface specifications • Proposal for the design, coding, and testing phase	• Definitive flowcharts, data flow diagrams, and data analysis diagrams • Test procedures • Completed and tested software modules • Tested I/O facilities • Proposal for the integration, validation, and installation phase
Integration, Validation, Installation	• Integration and validation of the software • Verification of compliance with system specs (design reviews) • Setup of the definitive documentation	• Completed and tested software modules • Tested I/O facilities • Proposal for the integration, validation, and installation phase	• Completed and tested software • Complete definitive documentation
Operation, Maintenance	• Use/application of the software • Maintenance (corrective and perfective)	• Completed and tested software • Complete definitive documentation	

A first difference between hardware and software appears in the *life-cycle phases* (Table 5.3). In contrast to Fig. 1.6, the production phase does not appear in the software life-cycle phases, since software can be copied without errors. A *partition* of the software life-cycle into clearly defined phases, each of them closed with an extensive *design review*, is mandatory for software quality assurance. A second basic distinction between hardware and software is given by *the quality attributes* or

Table 5.4 Important.software quality attributes and characteristics

Attribute	Definition
Compatibility	Degree to which two or more software modules or packages can perform their required functions while sharing the same hardware or software environment
Completeness	Degree to which a software module or package possesses the functions necessary and sufficient to satisfy user needs
Consistency	Degree of uniformity, standardization, and freedom from contradiction within the documentation or parts of a software package
Defect Freedom (Reliability)	Degree to which a software package can execute its required function without causing system failures
Defect Tolerance (Robustness)	Degree to which a software module or package can function correctly in the presence of invalid inputs or highly stressed environmental conditions
Documentation	Totality of documents necessary to describe, design, test, install, and maintain a software package
Efficiency	Degree to which a software module or package performs its required function with minimum consumption of resources (hardware and/or software)
Flexibility	Degree to which a software module or package can be modified for use in applications or environments other than those for which it was designed
Integrity	Degree to which a software package prevents unauthorized access to or modification of computer programs or data
Maintainability	Degree to which a software module or package can be easily modified to correct faults, improve the performance, or other attributes
Portability	Degree to which a software package can be transferred from one hardware or software environment to another
Reusability	Degree to which a software module can be used in another program
Simplicity	Degree to which a software module or package has been conceived and implemented in a straightforward and easily understandable way
Testability	Degree to which a software module or package facilitates the establishment of test criteria and the performance of tests to determine whether those criteria have been met
Usability	Degree to which a user can learn to operate, prepare inputs for, and interpret outputs of a software package

Software module is used here also for *software element*

characteristics (Table 5.4). The definitions of Table 5.4 extend those given in Appendix A1 and consider established standards [5.46]. Not all quality attributes of Table 5.4 can be *fulfilled* at the same time. In general, a priority list must be established and consequently followed *by all engineers* involved in a project. A further difficulty is the *quantitative evaluation* (assessment) of software quality attributes, i.e. the definition of *software quality metrics*.

From the above considerations, *software quality* can be defined as the degree to which a software package possesses a *stated combination of quality attributes* (characteristics). If completed by an appropriate set of software quality metrics, this allows an *objective* assessment of the quality level achieved. Since only a limited number of quality attributes can be fully satisfied by a specific software package, the main purpose of software quality assurance is to *maximize the common part of the quality attributes needed, specified, and realized.* To reach this target, specific activities have to be performed during *all software life-cycle* phases. Many of these activities can be derived from hardware quality assurance tasks, in particular regarding *defect prevention*, *configuration management*, and *testing* (including *defect correction*). However, *auditing* software quality assurance activities in a project should be more intensive and with a shorter feedback than for hardware (Fig. 5.2). Quite generally, it would be good to overcome the *traditional separation* of hardware and software and focus efforts on *defect prevention.*

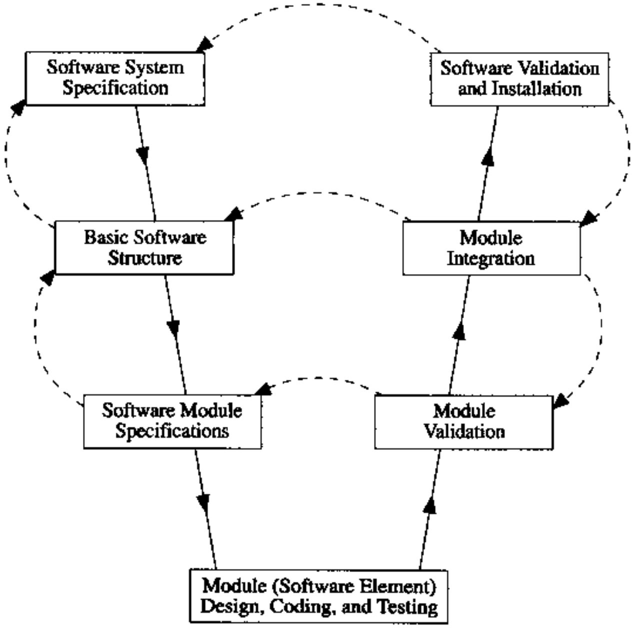


Figure 5.2 Procedure for software development (top-down design and bottom-up integration with vertical and horizontal control loops)

5.3.1 Guidelines for Software Defect Prevention

Defects can be introduced in different ways and at different points along the life cycle phases of software. The following are some *causes for defects*:

1. During the concept and definition phase
   - misunderstandings in the problem definition,
   - constraints on CPU performance, memory size, computing time, I/O facilities or others,
   - inaccurate *interface specifications,*
   - too little attention to user needs and/or skills.
2. During the design, coding, and testing phase
   - inaccuracies in the *detailed specifications,*
   - misinterpretation of the detailed specifications,
   - inconsistencies in procedures or algorithms,
   - timing problems,
   - data conversion errors,
   - faulty initialization,
   - complex software structuring, or large dependence between software modules.
3. During the integration, validation, and installation phase
   - too large an *interaction between software modules,*
   - errors during software corrections or modifications,
   - unclear or incomplete documentation,
   - changes in the hardware or software environment,
   - exceeding important resources (dynamic memory, disk, etc.).

Defects are thus caused in general by *human errors,* software developer or user. Their detection and removal become more expensive as the software life cycle progresses (often by a factor of 10 between each of the four main phases of Table 5.3, as in Fig. 8.2 for hardware manufacturing). Considering also that many defects are detected only by a particular combination of data and system states, and can therefore remain undiscovered for a long time after the software installation, the necessity for *defect prevention* through an appropriate software quality assurance becomes obvious. Following *design guidelines* can be useful:

1. Fix *written procedures/rules* and follow them during software development (such rules specify quality attributes, giving them the *necessary priority,* as well as quality assurance procedures and can be project-specific).
2. Formulate *detailed specifications and interfaces* as carefully as possible, such specifications/interfaces should exist before coding begins.
3. Give priority to *object oriented programming.*
4. Use well-behaved *high-level programming languages,* assembler only when a problem cannot be solved in any other way; use established Computer Aided Software Engineering (CASE) for program development and testing.

5. Partition software into *independent software modules* (modules should be individually testable, developed *top-down,* and integrated *bottom-up*).
6. Take into account all constraints given by *I/O facilities.*
7. Develop software able to *protect itself* and its data; plan for automatic testing and validation of data.
8. Consider aspects of *testing/testability* as *early* as possible in the development phase; increase testability through the use of definition languages (Vienna, RTRL, PSL, IORL).
9. Improve *understandability* and readability of software by introducing appropriate *comments.*
10. Document software carefully and carry out sufficient *configuration management,* particularly with respect to *design reviews* (Table 5.5).

Software for on-line systems (product and embedded software) should further be conceived to be *tolerant* (as far as possible) on *hardware failures* and to allow a *system reconfiguration,* particularly in the context of a *fail-safe concept* (hardware and software involved in fail-safe procedures should be periodically checked).

For equipment and systems exhibiting high reliability or safety requirements, software should be conceived and developed to be *defect tolerant,* i.e. to be able to continue operation despite the presence of software defects. For this purpose, *redundancy* considerations are necessary, in the *time domain* (protocol with retransmission, cyclic redundancy check, assertions, exception handling, etc.), in the *space domain* (error correcting codes, parallel processes, etc.), or a combination of both. Moreover, if the interaction between hardware and software in the realization of the required function at the system level is large (embedded software), redundancy considerations should also be extended to cover hardware defects and *failures,* i.e. to make the system *fault tolerant* (Section 2.3.7). In this context, effort should be devoted to the investigation of the *causes and effects* (criticality) of hardware and software *faults* from a *system level* point of view, including hardware, software, human factors, and logistical support.

5.3.2 Configuration Management

Configuration management is an important *quality assurance tool* during the design and development of complex equipment and systems, *both for hardware and software.* Applicable methods and procedures are outlined in Section 1.5 and discussed in Appendices A3 and A4. Many of these methods and procedures can be adapted to software, and this has also been done in some standards (IEEE Std 828-1998, 1028-1997, 1042-1987, [5.46 or A2.7]. Of particular importance for software are *design reviews* as described in Table 5.5 (see Table A3.2 for hardware aspects) and *configuration control* (management of changes and modifications).

5.3.3 Guidelines for Software Testing

Planning for *software testing* is generally a difficult task, as even small programs can have an extremely large number of states which makes a complete test impossible. A *test strategy* is then necessary. The problem is also known for hardware, for which special design guidelines to increase *testability* have been developed (Section 5.2). The most important rule, which applies to both hardware and software, is the *partitioning* of the item (hardware or software) into *independent modules* which can be individually tested and integrated bottom-up to constitute the system. Many rules can be project-specific. The following *design guidelines* can be useful in establishing a test strategy for software used in complex equipment and systems:

1. Plan *software* tests *early* in the design and coding phases, and integrate them step by step into a *test strategy.*
2. Use appropriate *tools* (debugger, coverage-analyzer, test generators, etc.).
3. Perform tests first at the *module level*, exercising all instructions, branches and logic paths.
4. Integrate and test successively the different modules *bottom-up* to the system level.
5. Test carefully all *suspected paths* (with potential defects) and software parts whose incorrect running could cause major system failures.
6. *Account for all defects* which have been discovered with indication of running time, software and hardware environments at the occurrence time (state, parameter set, hardware facilities, etc.), changes introduced, and *debugging* effort.
7. Test the complete software carefully in its *final* hardware and software environment.

Testing is the only practical possibility to find (and eliminate) *defects*. It includes *debug tests* (generally performed early in the design phase using breakpoints, desk checking, dumps, inspections, reversible executions, single-step operation, or traces) and *run tests*. Although costly (often up to 50% of the software development costs), tests cannot guarantee *freedom from defects*. A balanced distribution of the efforts between *preventive actions* (defect prevention) and *testing* must thus be found for each project.

5.3.4 Software Quality Growth Models

Since the beginning of the seventies, a large number of models have been proposed to describe the occurrence of *software defects* during operation of complex equipment or systems. Such an occurrence can generate a *failure at system level* and appears often *randomly distributed* in time. For this reason, modeling has been done in a similar way as for hardware failures, i.e. by introducing the concept of *software failure rate*. Such an approach may be valid to investigate *software quality growth* during *software validation* and installation, similarly to the reliability growth models developed in the sixties for hardware (Section 8.4). From the above consideration, the main target should be the *development of software free from defects* and thus focus the effort on *defect prevention* rather than on *defect modeling*. However, because of their use in investigating *software quality growth*, this section introduces briefly the *basic models* known for software defect modeling.

Table 5.5 Software design reviews (IEEE Std 1028-1988 [5.46])

	Type	Objective
Evaluation	Management Review	Provide recommendations for the following • activities progress, based on an evaluation of product development status • changing project direction or identifying the need for alternate planning • adequate allocation of resources through global control of the project
	Technical Review	Evaluate a specific software element and provide management with evidence that • the software element conforms to its specifications • the design (or maintenance) of the software element is being done according to plans, standards, and guidelines applicable for the project • changes to the software element are properly implemented and affect only those system areas identified by change specifications
Verification	Software Inspection	Detect and identify software element defects, in particular • verify that every software element satisfies its specifications • verify that every software element conforms to applicable standards • identify deviations from standards and specifications • evaluate software engineering data (e.g. defect and effort data)
	Walkthrough	Find defects, omissions, and contradictions in the software elements and consider alternative implementations (long associated with code examination, this process is also applicable to other aspects, e.g. architectural design, detailed design, test plans/procedures, and change control procedures)

Software Element is used here also for Software Module

1. Between consecutive occurrence points of a software defect, the "*failure rate*" is only a function of the *number of defects* still present in the software under consideration. This model leads to a *death process* and is known as the *Jeliniski-Moranda model*. If at $t=0$ the software contains n defects, the probability $P_i(t) = \Pr\{i$ defects have been removed up to the time $t \mid n$ defects were present at $t=0\}$ can be computed recursively from (see Fig. A7.9 with $\nu_0 = n\lambda$, $\nu_i = (n-i)\lambda$ and $\theta_i = 0$ for $i = 1, \ldots, n$)

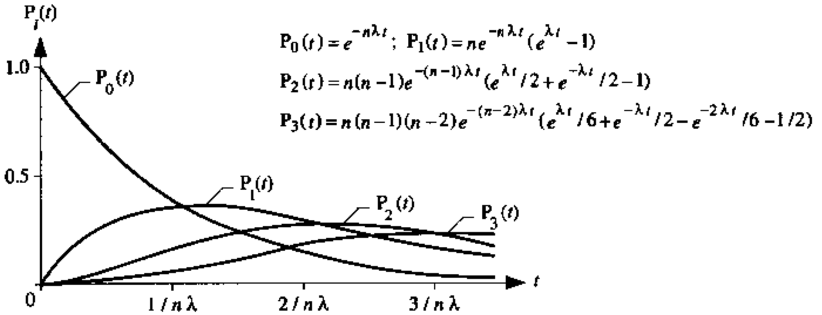


Figure 5.3 $P_i(t) = \Pr\{i$ defects have been removed up to the time $t \mid n$ defects were present at $t = 0\}$ for $i = 0$ to 3 and $n = 10$ (the time interval between consecutive occurrence points of a defect is exponentially distributed with parameter $\lambda_i = (n-i)\lambda$)

$$P_0(t) = e^{-n\lambda t}, \qquad P_i(t) = \int_0^t (n-i+1)\lambda e^{-(n-i)\lambda x} P_{i-1}(t-x)\,dx, \qquad i = 1, \ldots, n, \qquad (5.3)$$

or directly as

$$P_i(t) = \binom{n}{i}(1-e^{-\lambda t})^i e^{-(n-i)\lambda t}, \qquad i = 1, \ldots, n. \qquad (5.4)$$

Figure 5.3 shows $P_0(t)$ to $P_3(t)$ for $n = 10$. This model can be easily extended to cover the case in which the parameter λ also depends on the number of defects still present in the software.

2. Between consecutive occurrence points of a software defect, the "*failure rate*" is a function of the *number of defects* still present in the software and of the *time elapsed since the last occurrence point* of a defect. This model generalizes Model 1 above and can be investigated using *semi-Markov processes* (Appendix A7.6).
3. The *flow of occurrence* of software defects constitutes a *nonhomogeneous Poisson process* (Appendix A7.2.5). This model has been extensively investigated in the literature, together with reliability growth models for hardware, with different assumptions about the form of the process *intensity* (Section 8.5).
4. The *flow of occurrence* of software defects constitutes an arbitrary *point process*. This model is very general but difficult to verify because of the lack of reliable field data.

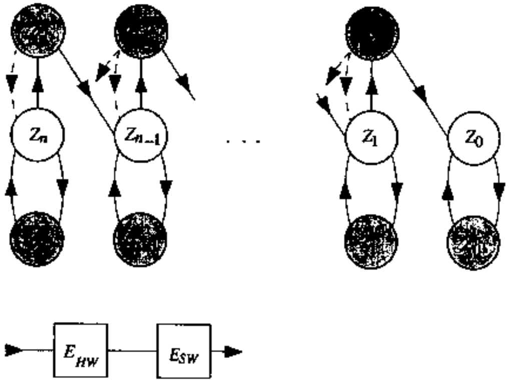


Figure 5.4 Simplified modeling of the time behavior of a system whose failure is caused by a hardware failure ($Z_i \rightarrow Z_i''$) or by the occurrence of a software defect ($Z_i \rightarrow Z_i'$)

The first two models are based on the concept of *failure rate* $\lambda(t)$, as defined by Eq. (A6.27), and the two other ones assume that the *flow of occurrence* of software defects constitutes particular *point processes* (not necessarily *renewal processes*, since the structure of the process can changes at the occurrence of each defect). Although the models described above are correct from a mathematical point of view, they often suffer from the lack of reliable *field data*, especially concerning the *number* and *criticality* (effect at system level) of the *defects still present* in the software under consideration. *Defect prevention* should thus remain the main focus of software quality assurance efforts.

For systems with hardware and software, one can often assume that defects in the software will be detected and eliminated one after the other. Only hardware failures should then remain. Figure 5.4 shows a possibility to take this into account. However, the interdependence between hardware and software can be greater than that assumed in Fig. 5.4. Also is the number (n) of defects in the software at the time $t = 0$ generally not known and furthermore, by eliminating a software defect new defects can be introduced. Investigations in this field are in progress.

6 Reliability and Availability of Repairable Systems

Reliability and availability analysis of *repairable systems* is generally performed using stochastic processes, including *Markov*, *semi-Markov*, and *semi-regenerative processes*. The mathematical foundation of these processes is given in Appendix A7. Equations used for the investigation of Markov and semi-Markov models are summarized in Table 6.2. This chapter investigates systematically most of the reliability models encountered in practical applications (the index S stays for system and designates the highest integration level considered). After a short introduction (Section 6.1), Section 6.2 investigates in detail the one-item structure (under very general assumptions). Sections 6.3 to 6.6 deal then extensively with basic series/parallel structures. To unify models and simplify calculations, it is assumed that the system has *only one repair crew* and that *no further failures can occur at system down*. Starting from constant failure and repair rates between successive states (Markov processes), generalization is performed step by step (beginning with the repair rates) up to the case in which the process involved is *regenerative* with a minimum number of regeneration states. Section 6.7 deals in depth with *approximate expressions* for large series/parallel structures, when *independent elements* or *macro structures* are used. Sections 6.8 to 6.10 consider systems with complex structures as well as the influence of imperfect switching and of preventive maintenance. For the basic structures investigated in Sections 6.2 to 6.7, important results are summarized in tables and graphs (Tables 6.3 to 6.10, Figs. 6.17 and 6.18). The use of dedicated software packages is discussed in Section 6.8.2.

6.1 Introduction and General Assumptions

Investigation of the *time behavior of repairable systems* spans a very large class of *stochastic processes*, from simple Poisson processes through Markov and semi-Markov processes up to sophisticated regenerative process with only one, or just some few *regeneration states*. Nonregenerative processes are seldom considered because of mathematical difficulties. Important for the choice of the class of processes to be used are the distribution functions of the failure-free operating time and of the repair time of each element in the system. When the failure and repair rates of all elements in the system are *constant* (*during the sojourn time in each state*, but not necessarily at a state change, e.g. because of load sharing as in Figs. 2.12, 6.8, 6.13), the process involved is a *time homogeneous Markov process* with finitely many states (exponentially distributed *sojourn time in each state*). Similarly if Erlang distributions occur, supplementary states can be introduced (Section 6.3). The possibility to transform a given stochastic process into a Markov process by introducing supplementary variables and/or states is theoretically possible, but often difficult to apply in practical situations (Appendix A7.8). A generalization of the distribution functions of repair times leads to *semi-regenerative processes*, i.e. to processes with an *embedded semi-Markov process*. This is always true if the system has *only one repair crew*, then each termination of a repair is a renewal point (because of the assumed constant failure rates). Arbitrary distributions of the repair and failure-free operating times lead in general to *nonregenerative stochastic processes*.

Table 6.1 gives an overview of the processes used in reliability investigations of repairable systems, with their possibilities and limits. Appendix A7 introduces these process with particular emphasis on reliability applications. All equations necessary for the reliability and availability computation of systems described by Markov and semi-Markov processes are summarized in Table 6.2.

Basic for reliability and availability calculations are the *reliability block diagram*, the *distribution functions* of the *failure-free operating time* and of the *repair time* of each element, as well as information about *maintenance strategy*, *logistical support*, *type of redundancy*, and *dependence between elements*. Reliability and availability are given as functions of time by solving appropriate systems of *differential or integral equations*, or expressed by the mean time to failure or the steady-state point availability at system level ($MTTF_{Si}$ or PA_S) by solving appropriate systems of *algebraic equations*. If the system has no redundancy, the reliability function is the same as in the nonrepairable case. In the presence of redundancy, it is generally assumed that redundant elements will be repaired *without interrupting the operation at system level*. *Reliability* investigations thus aim to find the occurrence of the *first system down*, whereas the *point availability* is the probability to find the system in an up state at the time t considered, independent of whether down states have occurred before t.

In order to unify the models and simplify calculations, the following assumptions are made for the analyses in Sections 6.2 to 6.6:

1. *Continuous operation*: Each element of the system is operating when not under repair or waiting for repair because of a busy repair crew. (6.1)
2. *No further failures at system down*: At system down the system is restored according to a given maintenance strategy to an up state at system level from which operation is continued, failures during a repair at system down are neglected. (6.2)

3. *Only one repair crew* is available at system level: Repair is performed according to a stated strategy, e.g. first-in/first-out, given priority, etc. (6.3)
4. *Redundancy*: Redundant elements are repaired without interruption of operation at system level. (6.4)
5. *As-good-as-new*: The repaired element is after repair as-good-as-new. (6.5)
6. *Independence*: Failure-free operating times and repair times of each element are statistically independent, positive, and continuous random variables with finite mean (*MTTF* =mean time to failure, *MTTR* = mean time to repair) and variance. (6.6)
7. *Support*: Preventive maintenance is neglected, switching is ideal, logistical support is unlimited (apart assumption 6.3). (6.7)

The above assumptions apply in many practical situations (Section 6.8). However, assumption (6.5) should be *verified critically*, in particular when the repaired element does not consist of just one part which has been replaced by a new one, but contains parts which have not been replaced during the repair. This assumption is valid if the non-replaced parts have *constant (time independent) failure rates* and applies also for considerations at system level.

Table 6.1 Stochastic processes used in reliability and availability analysis of repairable systems

Stochastic processes	Can be used in modeling	Background	Degree of difficulty
Renewal processes	Spare parts reservation in the case of arbitrary failure rates and negligible replacement or repair times	Renewal theory	Medium
Alternating renewal processes	One-item renewable (repairable) structures with arbitrary failure and repair rates	Renewal theory	Medium
Markov processes (finite state space, time-homogeneous)	Systems of arbitrary structure whose elements have constant failure rate and constant repair rates (*during the sojourn time in each state*, but not necessarily at a state change, e.g. because of load sharing)	Differential equations	Low
Semi-Markov processes	Some systems whose elements have constant or erlangian failure rates (Erlang distributed failure-free times) and arbitry repair rates	Integral equations	Medium
Semi-regenerative processes (processes with only few) regeneration states	Systems of arbitrary structure whose elements have constant failure rates and arbitrary repair rates; some redundant structures whose elements have not only constant failure rates and arbitrary repair rates	Integral equations	High
Nonregenerative proc. (e.g. superimposed renewal or alternating renewal processes)	Systems of arbitrary structure whose elements have arbitrary failure and repair rates	Sophisticated methods, partial diff. equations	High to very high

One-item repairable structures are analyzed in Section 6.2 under very general assumptions allowing a careful investigation of the *asymptotic and stationary behavior*. For all basic reliability structures (series, parallel, and series/parallel), investigations *begin* in Sections 6.3 to 6.6 by assuming *constant* failure and repair rates for each element in the reliability block diagram. Distributions of the repair times, and as far as possible of the failure-free operating times, are then generalized step by step up to the case in which the process involved remain regenerative with just a minimum number of *regeneration states*. This is also in order to show the capability and limits of the models considered. For large series/parallel structures, *approximate equations* are developed in Section 6.7 comparing different methods (independent elements, macro-structures, omission of states with low probability). More complex structures as well as the influence of preventive maintenance and of imperfect switching are discussed in Sections 6.8 to 6.10.

From the systematic investigations given in Sections 6.2 to 6.10, the following conclusions can be drawn:

1. *Analytical methods* work well for reasonably simple structures. For large state spaces, one of the following possibilities has to be used:
   - Computation of the mean time to failure and of the steady-state availability at system level only.
   - Combination of elementary structures (series, parallel, and series/parallel) of the reliability block diagram into *macro-structures* for which the expressions for the reliability and availability (exact solutions) are known (Table 6.10).
   - Combination of states in the transition probabilities diagram in $(t, t+\delta t]$.
   - Solution with the help of *approximate expressions*
     - use of *macro-structures* (see above),
     - omission of states with *more than k failures* (e.g. $k>2$), i.e. of states with low probabilities,
     - assumption that each element in the reliability block diagram works and is repaired *independently* of all other elements (the Boolean function for the reliability in the nonrepairable case (Section 2.3.4) applies for the point availability in the repairable case too, Table 6.9),
   - Use of *Monte-Carlo simulations.*
2. Assumption (6.2) allows a reduction of the state space and thus simplifies the calculation of the availability and interval reliability; it has *no influence on the reliability function* and delivers in many practical applications a good *approximate expression* for the availability and interval reliability.
3. As long as for each element in the reliability block diagram the condition *MTTR* << *MTTF* holds, the *form of the distribution functions* of the repair times has a small influence on the mean time to failure and on the steady-state availability at system level (Examples 6.7 and 6.8).

Table 6.2 Relationships for the reliability, point availability, and interval reliability of systems described by Markov and semi-Markov processes

	Reliability Function	Point Availability	Interval Reliability
Semi - Markov Processes (SMP)	$\mathrm{R}_{Si}(t)=1-\mathrm{Q}_i(t)+\sum_{Z_j\in U,\ j\ne i}\int_0^t \mathrm{q}_{ij}(x)\mathrm{R}_{Sj}(t-x)\,dx, \quad Z_i\in U$ $MTTF_{Si}=T_i+\sum_{Z_j\in U,\ j\ne i} p_{ij}\,MTTF_{Sj}, \quad Z_i\in U$ *with* $\mathrm{Q}_{ij}(x)=\Pr\{\tau_{ij}\le x\cap\tau_{ik}>\tau_{ij},\ k\ne j\}=p_{ij}\,\mathrm{F}_{ij}(x)$ $\mathrm{q}_{ij}(x)=\frac{d\mathrm{Q}_{ij}(x)}{dx}=p_{ij}\frac{d\mathrm{F}_{ij}(x)}{dx}, \quad \mathrm{Q}_{ii}(x)\equiv 0$ $p_{ij}=\Pr\{\tau_{ik}>\tau_{ij},\ k\ne j\}=\mathrm{Q}_{ij}(\infty), \quad p_{ii}=0$ $\mathrm{F}_{ij}(x)=\Pr\{\tau_{ij}\le x \mid \tau_{ik}>\tau_{ij},\ k\ne j\}$	$\mathrm{PA}_{Si}(t)=\sum_{Z_j\in U}\mathrm{P}_{ij}(t), \quad i=0,\dots,m$ $PA_S=\sum_{Z_j\in U}P_j=\sum_{Z_j\in U}\frac{T_j}{T_{jj}}$, see $\mathrm{IR}_S(\theta)$ for P_j, T_{jj} *with* $\mathrm{P}_{ij}(t)=\delta_{ij}(1-\mathrm{Q}_i(t))+\sum_{k=0,\ k\ne i}^{m}\int_0^t \mathrm{q}_{ik}(x)\mathrm{P}_{kj}(t-x)\,dx$ $i,j=0,\dots,m, \quad \delta_{ii}=1, \quad \delta_{ij}=0$ for $j\ne i$ $\mathrm{Q}_i(x)=\sum_{j=0,\ j\ne i}^{m}\mathrm{Q}_{ij}(x), \quad T_i=\int_0^\infty(1-\mathrm{Q}_i(x))\,dx$	Problem oriented computation; for *constant failure rates* it follows in steady-state that $\mathrm{IR}_S(\theta)=\sum_{Z_j\in U}P_j\,\mathrm{R}_{Sj}(\theta)=\sum_{Z_j\in U}\frac{T_j}{T_{jj}}\mathrm{R}_{Sj}(\theta)$ *with* $P_j=\lim_{t\to\infty}\mathrm{P}_j(t)=\frac{T_j}{T_{jj}}, \quad T_{jj}=\frac{1}{p_j}\sum_{k=0}^{m}p_k T_k$, and p_j from $p_j=\sum_{i=0}^{m}p_i p_{ij}$, for the embedded Markov chain with $p_{ii}\equiv 0$, $p_j>0$, and $p_0+\dots+p_m=1$
Time Homogeneous Markov Processes (method of integral equations)	$\mathrm{R}_{Si}(t)=e^{-\rho_i t}+\sum_{Z_j\in U,\ j\ne i}\int_0^t\rho_{ij}e^{-\rho_i x}\mathrm{R}_{Sj}(t-x)\,dx, \quad Z_i\in U$ $MTTF_{Si}=\frac{1}{\rho_i}+\sum_{Z_j\in U,\ j\ne i}\frac{\rho_{ij}}{\rho_i}MTTF_{Sj}, \quad Z_i\in U$ *with* ρ_{ij} = transition rates (see below) and $\rho_i=\sum_{j=0,\ j\ne i}^{m}\rho_{ij}$	$\mathrm{PA}_{Si}(t)=\sum_{Z_j\in U}\mathrm{P}_{ij}(t), \quad i=0,\dots,m$ $PA_S=\sum_{Z_j\in U}P_j$, see $\mathrm{IR}_S(\theta)$ for P_j *with* $\mathrm{P}_{ij}(t)=\delta_{ij}e^{-\rho_i t}+\sum_{k=0,\ k\ne i}^{m}\int_0^t\rho_{ik}e^{-\rho_i x}\mathrm{P}_{kj}(t-x)\,dx$ $i,j=0,\dots,m, \quad \delta_{ii}=1, \quad \delta_{ij}=0$ for $j\ne i$	$\mathrm{IR}_{Si}(t,t+\theta)=\sum_{Z_j\in U}\mathrm{P}_{ij}(t)\mathrm{R}_{Sj}(\theta), \quad i=0,\dots,m$ $\mathrm{IR}_S(\theta)=\sum_{Z_j\in U}P_j\,\mathrm{R}_{Sj}(\theta)$ see below for the computation of P_j

Table 6.2 (cont.)

Time Homogeneous Markov Processes (method of differential equations)	$\mathrm{R}_{Si}(t)=\sum_{Z_j\in U}\mathrm{P}'_{ij}(t), \quad Z_i\in U$ $MTTF_{Si}=\frac{1}{\rho_i}+\sum_{Z_j\in U,\ j\ne i}\frac{\rho_{ij}}{\rho_i}MTTF_{Sj}, \quad Z_i\in U$ *with* $\mathrm{P}'_{ij}(t)\equiv\mathrm{P}'_j(t)$ and $\mathrm{P}'_j(t)$ obtained from $\dot{\mathrm{P}}'_j(t)=-\rho'_j\,\mathrm{P}'_j(t)+\sum_{i=0,\ i\ne j}^{m}\mathrm{P}'_i(t)\rho'_{ij}, \quad j=0,\dots,m$ $\rho'_{ij}=\rho_{ij}$ for $Z_i\in U$, $\rho'_{ij}=0$ for $Z_i\in\overline{U}$, $\rho'_j=\sum_{i=0,\ i\ne j}^{m}\rho'_{ji}$ $\mathrm{P}'_i(0)=1$, $\mathrm{P}'_j(0)=0$ for $j\ne i$	$\mathrm{PA}_{Si}(t)=\sum_{Z_j\in U}\mathrm{P}_{ij}(t), \quad i=0,\dots,m$ $PA_S=\sum_{Z_j\in U}P_j$, see $\mathrm{IR}_S(\theta)$ for P_j *with* $\mathrm{P}_{ij}(t)\equiv\mathrm{P}_j(t)$ and $\mathrm{P}_j(t)$ obtained from $\dot{\mathrm{P}}_j(t)=-\rho_j\,\mathrm{P}_j(t)+\sum_{i=0,\ i\ne j}^{m}\mathrm{P}_i(t)\rho_{ij}, \quad j=0,\dots,m$ $\rho_j=\sum_{i=0,\ i\ne j}^{m}\rho_{ji}$, $\mathrm{P}_i(0)=1$, $\mathrm{P}_j(0)=0$ for $j\ne i$	$\mathrm{IR}_{Si}(t,t+\theta)=\sum_{Z_j\in U}\mathrm{P}_{ij}(t)\mathrm{R}_{Sj}(\theta), \quad i=0,\dots,m$ $\mathrm{IR}_S(\theta)=\sum_{Z_j\in U}P_j\,\mathrm{R}_{Sj}(\theta)$ *with* P_j from $\rho_j P_j=\sum_{i=0,\ i\ne j}^{m}P_i\,\rho_{ij}$, $j=0,\dots,m$ $P_j>0$, $P_0+\dots+P_m=1$; one equation for P_j (arbitrarily chosen because of the assumed ergodicity) must be dropped and replaced by $P_0+\dots+P_m=1$

$\mathrm{R}_{Si}(t)$ = Pr{ system *up* in $(0,t] \mid Z_i$ is entered at $t=0$ }, $Z_i\in U$ $^{*)}$; S stays for system

$MTTF_{Si}$ = E[system failure free time $\mid Z_i$ is entered at $t=0$] $=\int_0^\infty\mathrm{R}_{Si}(t)\,dt=\tilde{\mathrm{R}}_{Si}(0)$, $Z_i\in U$ $^{*)}$; $\tilde{\mathrm{R}}_{Si}(s)=\int_0^\infty\mathrm{R}_{Si}(t)e^{-st}dt$ = Laplace transform of $\mathrm{R}_{Si}(t)$

$\mathrm{PA}_{Si}(t)$ = Pr{ system *up* at $t \mid Z_i$ is entered at $t=0$ }, $i=0,\dots,m$ $^{*)}$

PA_S = Pr{ system *up* at t, in steady-state or for $t\to\infty$ } $=\lim_{s\downarrow 0}s\,\tilde{\mathrm{PA}}_{Si}(s)$, $i=0,\dots,m$; $AA_S=\lim_{t\to\infty}\frac{1}{t}\mathrm{E}$[system *up* in $(0,t]$] $=PA_S$

$\mathrm{IR}_{Si}(t,t+\theta)$ = Pr{ system *up* in $[t,t+\theta] \mid Z_i$ is entered at $t=0$ }, $i=0,\dots,m$ $^{*)}$

$\mathrm{IR}_S(\theta)$ = Pr{ system *up* in $[t,t+\theta]$, in steady-state or for $t\to\infty$ }

U = set of the *up* states, $\overline{U}$ = set of the *down* states, $U\cup\overline{U}=\{Z_0,\dots,Z_m\}$, $U\cap\overline{U}=\emptyset$

$\mathrm{P}_{ij}(t)$ = Pr{ system in state Z_j at $t \mid Z_i$ is entered at $t=0$ } $^{*)}$; $\mathrm{P}_j(t)$ = Pr{ system in state Z_j at t }, $P_j=\lim_{t\to\infty}\mathrm{P}_j(t)=\mathrm{P}_j(t)$ in steady - state

ρ_{ij} $=\lim_{\delta t\downarrow 0}\frac{1}{\delta t}$Pr{ transition from Z_i to Z_j in $(t,t+\delta t] \mid$ system in Z_i at t }, defined only for Markov processes and for $j\ne i$ ($p_{ii}=0$ in a SMP), t arbitrary

$^{*)}$ For Markov processes, the condition "Z_i is entered at $t=0$" can be replaced by "system in Z_i at $t=0$"

4. For large systems, it is preferable to start investigations by assuming ***Markov models***, i.e. constant failure and repair rates for all elements. If *MTTR* << *MTTF* holds, experience shows that the precision obtained is in general good for practical applications (Example 6.8). In a second step, more appropriate distribution functions for repair times or even for failure-free operating times can be considered. With this approach in mind, Table 6.2 gives the equations for the computation of the reliability, availability, and interval reliability of Markov and semi-Markov models. However, for large systems even the use of Markov or semi-Markov models can become time consuming, because of the large number of states involved (often close to $e \cdot n!$ for n elements in the reliability block diagram). In such cases, one of the approximation method given in Section 6.7 can be used (see also Section 6.8 for more general considerations).

6.2 One-Item Structure

A ***one-item structure*** is an entity of arbitrary complexity, which for investigation purposes is considered as a *unit*. Its reliability block diagram consists of a single element, see Fig. 6.1.

Under the assumptions (6.1) to (6.3) and (6.5) to (6.7), the repairable one-item structure is completely characterized by the distribution function of the *failure-free operating times* $\tau_0, \tau_1, \ldots$

$$F_A(t) = \Pr\{\tau_0 \le t\} \quad \text{and} \quad F(t) = \Pr\{\tau_i \le t\}, \quad i = 1, 2, \ldots, \tag{6.8}$$

with densities

$$f_A(t) = \frac{dF_A(t)}{dt} \quad \text{and} \quad f(t) = \frac{dF(t)}{dt}, \tag{6.9}$$

of the distribution function of the *repair times* $\tau'_0, \tau'_1, \ldots$

$$G_A(t) = \Pr\{\tau'_0 \le t\} \quad \text{and} \quad G(t) = \Pr\{\tau'_i \le t\}, \quad i = 1, 2, \ldots, \tag{6.10}$$

Figure 6.1 Reliability block diagram of a one-item structure

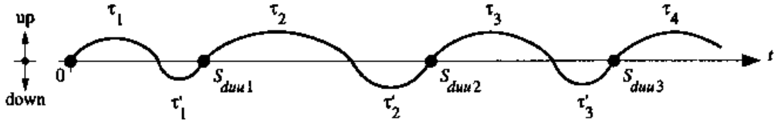


Figure 6.2 Time behavior of a repairable one-item structure new at $t = 0$ (repair times are exaggerated, • = renewal point, *duu* = transition from down state to up state given that the item is up at $t = 0$)

with densities

$$g_A(t) = \frac{dG_A(t)}{dt} \quad \text{and} \quad g(t) = \frac{dG(t)}{dt}, \tag{6.11}$$

and the probability p that the one-item structure is *up at* $t = 0$

$$p = \Pr\{up \text{ at } t = 0\} \tag{6.12}$$

or

$$1 - p = \Pr\{down \text{ (i.e. under repair) at } t = 0\},$$

respectively ($F_A(0) = F(0) = G_A(0) = G(0) = 0$). The time behavior of a one-item structure can be investigated in this case with help of the *alternating renewal process* introduced in Appendix A7.3.

Section 6.2.1 considers the one-item structure *new* at $t = 0$, i.e. the case $p = 1$ and $F_A(t) = F(t)$, with arbitrary $F(t)$ and $G(t)$. Generalization of the initial conditions at $t = 0$ will allow in Sections 6.2.4 and 6.2.5 an investigation in depth of the *asymptotic* and *stationary behavior*.

6.2.1 One-Item Structure New at Time $t = 0$

Figure 6.2 shows the time behavior of a one-item structure new at $t = 0$. $\tau_0, \tau_1, \ldots$ are the failure-free operating times. They are statistically independent and distributed according to $F(t)$ as in Eq. (6.8). Similarly, $\tau'_0, \tau'_1, \ldots$, are the repair times, distributed according to $G(t)$ as in Eq. (6.10). Considering assumption (6.3) the time points $0, S_{duu1}, \ldots$ are *renewal points* and constitute an ordinary *renewal process* embedded in the original alternating renewal process. Investigations of this Section are based on this property. To emphasize that the one-item structure is new at $t = 0$, all related quantities will have the index 0 (state Z_0 is entered at $t = 0$).

6.2.1.1 Reliability Function

The *reliability function* gives the probability that the item operates failure free in $(0, t]$. Thus (considering that the item is new at $t = 0$)

$$\mathrm{R}_0(t) = \Pr\{\text{up in } (0,t] \mid \text{new at } t = 0\}, \tag{6.13}$$

and from Eq. (6.8) it follows that

$$\mathrm{R}_0(t) = \Pr\{\tau_1 > t\} = 1 - \mathrm{F}(t). \tag{6.14}$$

The *mean time to failure* follows from Eqs. (6.14) and (A6.38)

$$MTTF_0 = \int_0^\infty \mathrm{R}_0(t)\,dt, \tag{6.15}$$

with the upper limit of the integral being T_L should the useful life of the item be limited to T_L (in this case $\mathrm{F}(t)$ jumps to 1 and $\mathrm{R}_0(t)$ to 0 at $t = T_L$). In the following it will be tacitly assumed that $T_L = \infty$.

6.2.1.2 Point Availability

The *point availability* gives the probability of finding the item operating at time t. Thus (considering that the item is new at $t = 0$)

$$\mathrm{PA}_0(t) = \Pr\{up \text{ at } t \mid \text{new at } t = 0\}, \tag{6.16}$$

and the following holds

$$\mathrm{PA}_0(t) = 1 - \mathrm{F}(t) + \int_0^t \mathrm{h}_{duu}(x)(1 - \mathrm{F}(t - x))\,dx. \tag{6.17}$$

Equation (6.17) has been derived in Appendix A7.3 using the theorem of total probability, see Eq. (A7.56). $1 - \mathrm{F}(t)$ is the probability of no failure in $(0, t]$, $\mathrm{h}_{duu}(x)\,dx$ gives the probability that any one of the renewal points $S_{duu1}, S_{duu2}, \ldots$ lies in $(x, x + dx]$, and $1 - \mathrm{F}(t - x)$ is the probability that no further failure occurs in $(x, t]$. Using the *Laplace transform* (Appendix A9.7) and considering Eq. (A7.50) with $\mathrm{F}_A(t) = F(t)$, Eq. (6.17) yields

$$\tilde{\mathrm{P}}\mathrm{A}_0(s) = \frac{1 - \tilde{\mathrm{f}}(s)}{s(1 - \tilde{\mathrm{f}}(s)\tilde{\mathrm{g}}(s))}. \tag{6.18}$$

$\tilde{\mathrm{f}}(s)$ and $\tilde{\mathrm{g}}(s)$ are the Laplace transforms of the failure-free operating times and repair times densities, respectively (defined by Eqs. (6.9) and (6.11)).

Example 6.1

a) Give the Laplace transform of the point availability for the case of a *constant failure rate* λ.

Solution

With $\mathrm{F}(t) = 1 - e^{-\lambda t}$ or $\mathrm{f}(t) = \lambda e^{-\lambda t}$, Eq. (6.18) yields

$$\tilde{\mathrm{P}}\mathrm{A}_0(s) = \frac{1}{s + \lambda(1 - \tilde{\mathrm{g}}(s))}. \tag{6.19}$$

A Gamma distribution, i.e. $\mathrm{g}(t) = \alpha(\alpha t)^{\beta-1} e^{-\alpha t} / \Gamma(\beta)$ as Eq. (A6.98) with $\lambda = \alpha$, would lead to

$$\tilde{\mathrm{P}}\mathrm{A}_0(s) = \frac{(s+\alpha)^\beta}{(s+\lambda)(s+\alpha)^\beta - \lambda\alpha^\beta}.$$

b) Give the Laplace transform and the corresponding time function of the point availability for the case of *constant failure and repair rates* λ and μ.

Solution

With $\mathrm{f}(t) = \lambda e^{-\lambda t}$ and $\mathrm{g}(t) = \mu e^{-\mu t}$, Eq. (6.18) yields

$$\tilde{\mathrm{P}}\mathrm{A}_0(s) = \frac{s + \mu}{s(s + \lambda + \mu)},$$

and thus (Table A9.7b)

$$\mathrm{PA}_0(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}. \tag{6.20}$$

$\mathrm{PA}_0(t)$ converges rapidly, exponentially with a time constant $1/(\lambda + \mu) \approx 1/\mu = MTTR$, to the asymptotic value $\mu/(\lambda + \mu)$, see Section 6.2.4 for a more extensive discussion.

$\mathrm{PA}_0(t)$ can also be derived using *regenerative process* arguments (Appendix A7.6). After the first repair the item is *as-good-as-new*, S_{duu1} is a *renewal point* and from this time point the process *restarts anew* as at $t = 0$. Therefore,

$$\Pr\{up \text{ at } t \mid S_{duu1} = x\} = \mathrm{PA}_0(t - x). \tag{6.21}$$

Since the event

up at t

occurs with exactly one of the following two mutually exclusive events

no failure in $(0, t]$

or

$S_{duu1} \le t \cap up$ at t

it follows that

$$\mathrm{PA}_0(t) = 1 - \mathrm{F}(t) + \int_0^t (\mathrm{f}(x) * \mathrm{g}(x))\,\mathrm{PA}_0(t - x)\,dx, \tag{6.22}$$

where $f(x) * g(x)$ is the density of the sum $\tau_1 + \tau_1'$ (see Fig 6.2 and Eq. (A6.75)). The Laplace transform of $PA_0(t)$ as per Eq. (6.22) is still given by Eq. (6.18).

6.2.1.3 Average Availability

The *average availability* is defined as the expected proportion of time in which the item is operating in $(0, t]$. Thus (considering that the item is new at $t = 0$)

$$AA_0(t) = \frac{1}{t} E[\text{total } up \text{ time in } (0, t] \mid \text{new at } t = 0], \tag{6.23}$$

and

$$AA_0(t) = \frac{1}{t} \int_0^t PA_0(x) dx \tag{6.24}$$

holds with $PA_0(x)$ from Eq. (6.17) or Eq. (6.22). To prove Eq. (6.24), consider that the time behavior of the item can be described by a binary random function taking values 1 with probability $PA_0(t)$ for up state and 0 for down state, the mean (expected value) of this random function is equal to $AA_0(t)$, see also Eq. (A6.39).

6.2.1.4 Interval Reliability

The *interval reliability* gives the probability that the item operates failure free during an interval $[t, t + \theta]$. Thus (considering that the item is new at $t = 0$)

$$IR_0(t, t + \theta) = \Pr\{up \text{ in } [t, t + \theta] \mid \text{new at } t = 0\} \tag{6.25}$$

and the same method used to find Eq. (6.17) leads to

$$IR_0(t, t + \theta) = 1 - F(t + \theta) + \int_0^t h_{duu}(x)(1 - F(t + \theta - x)) dx. \tag{6.26}$$

Example 6.2

Give the interval reliability for the case of a *constant failure rate* λ.

Solution

With $F(t) = 1 - e^{-\lambda t}$ it follows that

$$IR_0(t, t + \theta) = e^{-\lambda(t+\theta)} + \int_0^t h_{duu}(x) e^{-\lambda(t+\theta-x)} dx = [e^{-\lambda t} + \int_0^t h_{duu}(x) e^{-\lambda(t-x)} dx]\, e^{-\lambda\theta}.$$

Comparison with Eq. (6.17) for $F(t) = 1 - e^{\lambda t}$ yields

$$IR_0(t, t + \theta) = PA_0(t) e^{-\lambda\theta}. \tag{6.27}$$

It must be pointed out that the *product rule* of Eq. (6.27), expressing that $\Pr\{up \text{ in } [t, t + \theta]\} = \Pr\{up \text{ at } t\} \cdot \Pr\{\text{no failure in } (t, t + \theta]\}$, is valid *only* because of the *constant failure rate* λ (*memoryless property*); in the general case, the second term would have the form $\Pr\{\text{no failure in } (t, t + \theta] \mid up \text{ at } t\}$ which differs from $\Pr\{\text{no failure in } (t, t + \theta]\}$.

6.2.1.5 Special Kinds of Availability

In addition to the point and average availabilities given in Sections 6.2.1.2 and 6.2.1.3 there are several other kinds of availability frequently used in applications:

1. *Mission Availability*: The mission availability $MA_0(T_o, t_f)$ gives the probability that in a mission of total operating time T_o each failure can be repaired within a time span t_f. Hence (considering that the item is new at $t = 0$)

   $$MA_0(T_o, t_f) = \Pr\{\text{each individual failure occuring in a mission with total operating time } T_o \text{ can be repaired in a time} \le t_f \mid \text{new at } t = 0\}. \tag{6.28}$$

   Mission availability is important in applications where *interruptions of duration* $\le t_f$ can be accepted. Its calculation takes care of all cases in which there are (exactly) 0, 1, 2, ... failures, by considering that at the end to the mission the item is operating (to reach T_o). Thus,

   $$MA_0(T_o, t_f) = 1 - F(T_o) + \sum_{n=1}^{\infty} (F_n(T_o) - F_{n+1}(T_o))(G(t_f))^n. \tag{6.29}$$

   $F_n(T_o) - F_{n+1}(T_o)$ is the probability of n failures during the total operating time T_o, as given by Eq. (A7.14) with $F_1(t) = F(t)$ for $t \le T_o$, and $(G(t_f))^n$ is the probability that each repair will be shorter than t_f. The case of *constant failure rate* λ yields

   $$F_n(T_o) - F_{n+1}(T_o) = \frac{(\lambda T_o)^n}{n!} e^{-\lambda T_o}$$

   and thus

   $$MA_0(T_o, t_f) = e^{-\lambda T_o (1 - G(t_f))}. \tag{6.30}$$

2. *Work-Mission Availability*: The work-mission availability $WMA_0(T_o, t_d)$ gives the probability that the *sum* of the repair times for all failures occurring in a mission with *total operating time* T_o is smaller than t_d. Hence (considering that the item is new at $t = 0$)

$$\text{WMA}_0(T_o,t_d)=\Pr\{\text{sum of the repair times for all failures occurring in a mission of total operating time } T_o \text{ is} \le t_d \mid \text{new at } t=0\}. \qquad (6.31)$$

Similarly as for Eq. (6.29) it follows that

$$\text{WMA}_0(T_o, t_d) = 1-\text{F}(T_o)+\sum_{n=1}^{\infty}(\text{F}_n(T_o)-\text{F}_{n+1}(T_o))\text{G}_n(t_d), \qquad (6.32)$$

where $\text{G}_n(t_d)$ is the distribution function of the sum of n repair times distributed according to $\text{G}(t)$. It is not difficult to recogize that $\text{WMA}_0(t-x, x)$ is the distribution function of the *total down time* (sojourn time in the repair state) in $(0, t]$ for a process described by Fig. 6.2, i.e. for any $0 < x \le t$ it holds that

$$\Pr\{\text{total repair time in } (0, t] \le x \mid \text{new at } t = 0\}$$
$$= \Pr\{\text{total up time in } (0, t] > t-x \mid \text{new at } t=0\} = \text{WMA}_0(t-x, x),$$

with $\Pr\{\text{total repair time in}(0, t]=0 \mid \text{new at } t=0\}=1-\text{F}(t)=\text{WMA}_0(t,0)$. From this interpretaion of $\text{WMA}_0(T_o,t_d)$ it follows that

$$\text{E}[\text{total } up \text{ time in } (0, t] \mid \text{new at } t = 0] = \int_0^t \text{WMA}_0(t-x, x)dx. \qquad (6.33)$$

Equation (6.33) is a further expression for $t\cdot\text{AA}_0(t)$ given in Eq. (6.24).

3. *Joint Availability*: The joint availability $\text{JA}_0(t, t+\theta)$ gives the probability of finding the item operating at the time points t and $t+\theta$, hence (considering that the item is new at $t=0$)

$$\text{JA}_0(t, t+\theta) = \Pr\{(\text{up at } t \cap \text{up at } t+\theta) \mid \text{new at } t = 0\}. \qquad (6.34)$$

For the case of a *constant failure rate*, the two events *up at t* and *up at t + θ* of Eq. (6.34) are independent, thus

$$\text{JA}_0(t, t+\theta) = \text{PA}_0(t)\text{PA}_0(\theta). \qquad (6.35)$$

For an arbitrary failure rate, it is necessary to consider that the event

up at $t \cap$ *up* at $t+\theta$

occurs with one of the following two mutually exclusive events (Appendix A7.3)

up in $[t, t+\theta]$

or

up at $t \cap$ next failure occurs before $t+\theta \cap$ *up* at $t+\theta$.

The probability for the first event is the interval reliability $\text{IR}_0(t,t+\theta)$ given by Eq. (6.26). For the second event, it is necessary to consider the distribution function of the *forward recurrence time in the up state* $\tau_{Ru}(t)$. As shown in Fig. 6.3, $\tau_{Ru}(t)$ can only be defined if the item is in the up state at time t, hence

$$\Pr\{\tau_{Ru}(t) > x \mid \text{new at } t = 0\} = \Pr\{up \text{ in } (t, t+x] \mid (up \text{ at } t \cap \text{new at } t = 0)\}$$

and thus, as in Example A7.2 and considering Eqs. (6.16) and (6.25),

$$\Pr\{\tau_{Ru}(t)>x \mid \text{new at } t=0\}=\frac{\Pr\{up \text{ in } [t,t+x] \mid \text{new at } t = 0\}}{\Pr\{up \text{ at } t \mid \text{new at } t = 0\}}=\frac{\text{IR}_0(t,t+x)}{\text{PA}_0(t)}$$
$$=1-\text{F}_{\tau_{Ru}}(x) \qquad (6.36)$$

From Eq. (6.36) and the above considerations, it follows that

$$\text{JA}_0(t,t+\theta) = \text{IR}_0(t,t+\theta)+\text{PA}_0(t)\int_0^{\theta} \text{f}_{\tau_{Ru}}(x)\text{PA}_1(\theta-x)dx$$
$$= \text{IR}_0(t,t+\theta)-\int_0^{\theta}\frac{\partial\text{IR}_0(t,t+x)}{\partial x}\text{PA}_1(\theta-x)dx, \qquad (6.37)$$

where $\text{PA}_1(t) = \Pr\{up \text{ at } t \mid \text{a repair begins at } t = 0\}$ is given by

$$\text{PA}_1(t) = \int_0^t h_{dud}(x)(1-\text{F}(t-x))dx, \qquad (6.38)$$

with $\text{h}_{dud}(t)=\text{g}(t)+\text{g}(t)*\text{f}(t)*\text{g}(t)+\text{g}(t)*\text{f}(t)*\text{g}(t)*\text{f}(t)*\text{g}(t)+\ldots$, see Eq. (A7.50). $\text{JA}_0(t,t+\theta)$ can also be obtained in a similar way to $\text{PA}_0(t)$ in Eq. (6.17), i.e. by considering the alternating renewal process starting up at the time t with $\tau_{Ru}(t)$ distributed according to $\text{F}_{\tau_{Ru}}(x)$ as per Eq. (6.36), this leads to

$$\text{JA}_0(t,t+\theta) = \text{IR}_0(t,t+\theta)+\int_0^{\theta}\text{h}'_{duu}(x)(1-\text{F}(\theta-x))dx, \qquad (6.39)$$

with $\text{h}'_{duu}(x)=\text{f}'_{\tau_{Ru}}(x)*g(x)+\text{f}'_{\tau_{Ru}}(x)*\text{g}(x)*\text{f}(x)*\text{g}(x)+\ldots$, see Eq. (A7.50), and $\text{f}'_{\tau_{Ru}}(x)=\text{PA}_0(t)\text{f}_{\tau_{Ru}}(x)=\text{PA}_0(t)\,d\text{F}_{\tau_{Ru}}(x)/dx=-\partial\text{IR}_0(t,t+x)/\partial x$, see Eqs. (6.36) and (6.37). Similarly as for $\tau_{Ru}(t)$, the distribution function for the *forward recurrence time in the down state* $\tau_{Rd}(t)$, see Fig. 6.3, is given by

$$\Pr\{\tau_{Rd}(t) \le x \mid \text{new at } t = 0\} = 1-\frac{\int_0^t \text{h}_{udu}(y)(1-\text{G}(t+x-y))dy}{1-\text{PA}_0(t)} \qquad (6.40)$$

with $\text{h}_{udu}(t)=\text{f}(t)+\text{f}(t)*\text{g}(t)*\text{f}(t)+\ldots$, see Eq. (A7.50).

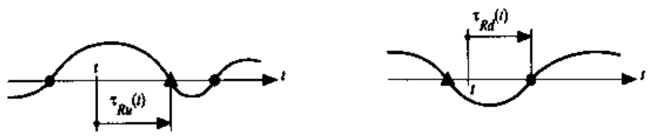


Figure 6.3 Forward recurrence times $\tau_{Ru}(t)$ and $\tau_{Rd}(t)$ in an alternating renewal process

6.2.2 One-Item Structure New at Time $t = 0$ and with Constant Failure Rate λ

In many practical applications, a *constant failure rate* λ can be assumed. In this case, the expressions of Section 6.2.1 can be simplified making use of the *memoryless property* given by the constant failure rate. Table 6.3 summarizes the results for the cases of constant failure rate λ and constant or arbitrary repair rate μ or $\mu(t) = g(t)/(1-G(t))$.

6.2.3 One-Item Structure with Arbitrary Initial Conditions at Time $t = 0$

Generalization of the initial conditions at time $t = 0$, i.e. the introduction of p, $\mathrm{F}_A(t)$ and $\mathrm{G}_A(t)$as defined by Eqs. (6.12), (6.8), and (6.10), leads to a time behavior of the one-item repairable structure described by Fig. A7.3 and to the following results:

1. *Reliability function* $\mathrm{R}(t)$

$$\mathrm{R}(t) = \Pr\{up \text{ in } (0, t] \mid up \text{ at } t=0\} = 1 - \mathrm{F}_A(t). \tag{6.41}$$

Equation (6.41) follows from the more general expression $\Pr\{up \text{ at } t=0 \cap up \text{ in } (0,t]\} = \Pr\{up \text{ at } t=0\} \cdot \Pr\{up \text{ in } (0,t] \mid up \text{ at } t=0\} = p \cdot (1-\mathrm{F}_A(t) = p \cdot R(t)$.

2. *Point availability* $\mathrm{PA}(t)$

$$\begin{aligned}\mathrm{PA}(t) &= \Pr\{up \text{ at } t\}\\ &= p[1-\mathrm{F}_A(t) + \int_0^t \mathrm{h}_{duu}(x)(1-\mathrm{F}(t-x))dx] + (1-p)\int_0^t \mathrm{h}_{dud}(x)(1-\mathrm{F}(t-x))dx,\end{aligned} \tag{6.42}$$

with $\mathrm{h}_{duu}(t) = \mathrm{f}_A(t) * \mathrm{g}(t) + \mathrm{f}_A(t) * (t) * \mathrm{f}(t) * \mathrm{g}(t) + \ldots$ and $\mathrm{h}_{dud}(t) = \mathrm{g}_A(t) + \mathrm{g}_A(t) * \mathrm{f}(t) * \mathrm{g}(t) + \mathrm{g}_A(t) * \mathrm{f}(t) * \mathrm{g}(t) * \mathrm{f}(t) * \mathrm{g}(t) + \ldots$.

3. *Average availability* $\mathrm{AA}(t)$

$$\mathrm{AA}(t) = \frac{1}{t}\mathrm{E}[\text{total } up \text{ time in } (0,t]] = \frac{1}{t}\int_0^t \mathrm{PA}(x)dx. \tag{6.43}$$

4. *Interval reliability* $\mathrm{IR}(t, t+\theta)$

$$\begin{aligned}\mathrm{IR}(t, t+\theta) &= \Pr\{\text{up in } [t, t+\theta]\}\\ &= p[1-\mathrm{F}_A(t+\theta) + \int_0^t \mathrm{h}_{duu}(x)(1-\mathrm{F}(t+\theta-x))dx]\\ &\quad + (1-p)\int_0^t \mathrm{h}_{dud}(x)(1-\mathrm{F}(t+\theta-x))dx.\end{aligned} \tag{6.44}$$

5. *Joint availability* $\mathrm{JA}(t, t+\theta)$

$$\begin{aligned}\mathrm{JA}(t, t+\theta) &= \Pr\{up \text{ at } t \cap up \text{ at } t+\theta\}\\ &= \mathrm{IR}(t, t+\theta) - \int_0^\theta \frac{\partial \mathrm{IR}(t, t+x)}{\partial x}\mathrm{PA}_1(\theta - x)dx,\end{aligned} \tag{6.45}$$

with $\mathrm{IR}(t, t+\theta)$ from Eq. (6.44) and $\mathrm{PA}_1(t)$ from Eq. (6.38).

Table 6.3 Results for the repairable one-item structure new at $t = 0$ with *constant failure rate* λ

	Repair rate		Remarks, Assumptions
	arbitrary	constant (μ)	
1. Reliability function $\mathrm{R}_0(t)$	$e^{-\lambda t}$	$e^{-\lambda t}$	$\mathrm{R}_0(t) = \Pr\{up \text{ in } (0, t] \mid \text{new at } t = 0\}$
2. Point avail ability $\mathrm{PA}_0(t)$	$e^{-\lambda t} + \int_0^t \mathrm{h}_{duu}(x)e^{-\lambda(t-x)}dx$	$\frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}$	$\mathrm{PA}_0(t) = \Pr\{up \text{ at } t \mid \text{new at } t = 0\}$, $h_{duu} = f*g + f*g*f*g + \ldots$
3. Average availability $\mathrm{AA}_0(t)$	$\frac{1}{t}\int_0^t \mathrm{PA}_0(x)dx$	$\frac{\mu}{\lambda+\mu} + \frac{\lambda(1-e^{-(\lambda+\mu)t})}{t(\lambda+\mu)^2}$	$\mathrm{AA}_0(t) = \mathrm{E}[\text{total } up \text{ time in } (0, t] \mid \text{new at } t = 0\} / t$
4. Interval reliability $\mathrm{IR}_0(t, t+\theta)$	$\mathrm{PA}_0(t)e^{-\lambda\theta}$	$\frac{\mu e^{-\lambda\theta}}{\lambda+\mu} + \frac{\lambda e^{-(\lambda+\mu)t-\lambda\theta}}{\lambda+\mu}$	$\mathrm{IR}_0(t, t+\theta) = \Pr\{up \text{ in } [t, t+\theta] \mid \text{new at } t = 0)$
5. Joint availability $\mathrm{JA}_0(t, t+\theta)$	$\mathrm{PA}_0(t)\mathrm{PA}_0(\theta)$	$\mathrm{PA}_0(t)\mathrm{PA}_0(\theta)$	$\mathrm{JA}_0(t, t+\theta) = \Pr\{up \text{ at } t \cap up \text{ at } t+\theta \mid \text{new at } t = 0\}$, $\mathrm{PA}_0(x)$ as in point 2
6. Mission availability $\mathrm{MA}_0(T_o, t_f)$	$e^{-\lambda T_o (1-\mathrm{G}(t_f))}$	$e^{-\lambda T_o e^{-\mu t_f}}$	$\mathrm{MA}_0(T_o, t_f) =$ Pr{each failure in a mission with total operating time T_o can be repaired in a time $\le t_f \mid$ new at $t = 0$}

λ = failure rate, $\Pr\{\tau_{Ru}(t) \le x\} = 1 - e^{-\lambda x}$ (Fig. 6.3), *up* means in the operating state

6. *Forward recurrence time* ($\tau_{Ru}(t)$ and $\tau_{Rd}(t)$ as in Fig. 6.3)

$$\Pr\{\tau_{Ru}(t) \le x\} = 1 - \frac{\mathrm{IR}(t, t+x)}{\mathrm{PA}(t)}, \tag{6.46}$$

with $\mathrm{IR}(t, t+x)$ according to Eq. (6.44) and $\mathrm{PA}(t)$ from Eq. (6.42), and

$$\Pr\{\tau_{Rd}(t) \le x\} = 1 - \frac{\Pr\{down \text{ in } [t, t+x]\}}{1 - \mathrm{PA}(t)}, \tag{6.47}$$

where

$$\Pr\{down \text{ in } [t, t+x]\} = p \int_0^t \mathrm{h}_{udu}(y)(1 - \mathrm{G}(t+x-y))\,dy$$
$$+(1-p)[1 - \mathrm{G}_A(t+x) + \int_0^t \mathrm{h}_{udd}(y)(1 - \mathrm{G}(t+x-y))\,dy],$$

with $\mathrm{h}_{udu}(t) = \mathrm{f}_A(t) + \mathrm{f}_A(t) * \mathrm{g}(t) * \mathrm{f}(t) + \mathrm{f}_A(t) * \mathrm{g}(t) * \mathrm{f}(t) * \mathrm{g}(t) * \mathrm{f}(t) + \ldots$ and $\mathrm{h}_{udd}(t) = \mathrm{g}_A(t) * \mathrm{f}(t) + \mathrm{g}_A(t) * \mathrm{f}(t) * \mathrm{g}(t) * \mathrm{f}(t) + \ldots$.

Expressions for mission availability and work-mission availability are generally only used with items new at time $t = 0$ (Eqs. (6.29) and (6.32)), see [6.4 (1973)] for a generalization.

6.2.4 Asymptotic Behavior

As $t \to \infty$ expressions for the point availability, average availability, interval reliability, joint availability, and for the distribution of the forward recurrence times (Eqs. (6.42) – (6.47)) converge to quantities which are independent of t and the initial conditions at time $t = 0$. Using the *key renewal theorem* (Eq. (A7.29)) it follows that

$$\lim_{t\to\infty} \mathrm{PA}(t) = PA = \frac{MTTF}{MTTF + MTTR}, \tag{6.48}$$

$$\lim_{t\to\infty} \mathrm{AA}(t) = AA = \frac{MTTF}{MTTF + MTTR} = PA, \tag{6.49}$$

$$\lim_{t\to\infty} \mathrm{IR}(t, t+\theta) = \mathrm{IR}(\theta) = \frac{1}{MTTF + MTTR} \int_\theta^\infty (1 - \mathrm{F}(y))\,dy, \tag{6.50}$$

$$\lim_{t\to\infty} \mathrm{JA}(t, t+\theta) = \mathrm{JA}(\theta) = \frac{MTTF}{MTTF + MTTR} \mathrm{PA}_{0e}(\theta), \tag{6.51}$$

$$\lim_{t\to\infty} \Pr\{\tau_{Ru}(t) \le x\} = \frac{1}{MTTF} \int_0^x (1 - \mathrm{F}(y))\,dy, \tag{6.52}$$

$$\lim_{t\to\infty} \Pr\{\tau_{Rd}(t) \le x\} = \frac{1}{MTTR} \int_0^x (1 - \mathrm{G}(y))\,dy, \tag{6.53}$$

where $MTTF = \mathrm{E}[\tau_i]$, $MTTR = \mathrm{E}[\tau'_i]$, $i = 1, 2, \ldots$, and $\mathrm{PA}_{0e}(\theta)$ is the point availability according to Eq. (6.42) with $p = 1$ and $\mathrm{F}_A(t)$ from Eq. (6.57) or Eq. (6.52).

Example 6.3

Show that for a repairable one-item structure in continuous operation, the limit

$$\lim_{t\to\infty} \mathrm{PA}(t) = PA = \frac{MTTF}{MTTF + MTTR}$$

is valid for any distribution function $\mathrm{F}(t)$ of the failure-free operating times and $\mathrm{G}(t)$ of the repair times, if $MTTF < \infty$, $MTTR < \infty$, and the densities $\mathrm{f}(t)$ and $\mathrm{g}(t)$ go to 0 as $t \to \infty$.

Solution

Using the *renewal density theorem* (A7.30) it follows that

$$\lim_{t\to\infty} \mathrm{h}_{duu}(t) = \lim_{t\to\infty} \mathrm{h}_{dud}(t) = \frac{1}{MTTF + MTTR}.$$

Furthermore, applying the *key renewal theorem* (A 7.29) to $\mathrm{PA}(t)$ given by Eq. (6.42) yields

$$\lim_{t\to\infty} \mathrm{PA}(t) = p(1 - 1 + \frac{\int_0^\infty (1 - \mathrm{F}(x))\,dx}{MTTF + MTTR}) + (1-p)\frac{\int_0^\infty (1 - \mathrm{F}(x))\,dx}{MTTF + MTTR}$$
$$= p\frac{MTTF}{MTTF + MTTR} + (1-p)\frac{MTTF}{MTTF + MTTR} = \frac{MTTF}{MTTF + MTTR}.$$

The limit $MTTF \,/\, (MTTF + MTTR)$ can also be obtained from the final value theorem of the Laplace transform (Table A9.7), considering for $s \to 0$

$$\tilde{\mathrm{f}}(s) \approx 1 - s\,MTTF \quad \text{and} \quad \tilde{\mathrm{g}}(s) \approx 1 - s\,MTTR. \tag{6.54}$$

In the case of a *constant* failure rate λ and a *constant* repair rate μ, Eq. (6.42) yields

$$\mathrm{PA}(t) = \frac{\mu}{\lambda + \mu} + (p - \frac{\mu}{\lambda + \mu})e^{-(\lambda+\mu)t}. \tag{6.55}$$

Thus, for this important case, the convergence of $\mathrm{PA}(t)$ toward $PA = \mu / (\lambda + \mu)$ is *exponential* with a time constant $1 / (\lambda + \mu) < 1 / \mu = MTTR$. Considering $0 \le p \le 1$ and $\lambda < \mu$ ($\lambda << \mu$ holds in all practical applications) it follows that

$$\left|\mathrm{PA}(t) - PA\right| \le \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)t} \le e^{-\mu t} = e^{-t / MTTR}.$$

In particular, for $p = 1$, i.e. for $\mathrm{PA}(0) = 1$ and $\mathrm{PA}(t) \equiv \mathrm{PA}_0(t)$, one obtains

$$\left| \mathrm{PA}_0(t) - PA \right| = \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \le \frac{\lambda}{\mu} e^{-\mu t} = \lambda\, MTTR\, e^{-t/MTTR}. \tag{6.56}$$

Inequality (6.56) can be generalized to a wide class of distribution functions $\mathrm{G}(x)$ with density $\mathrm{g}(x)$ for the repair time [6.12]. Let $\mathrm{F}(x) = 1 - e^{-\lambda x}$ and the Laplace transform $\tilde{\mathrm{g}}(s) = \int_0^\infty e^{-sx}\,\mathrm{g}(x)\,dx$ of the repair time density $\mathrm{g}(x)$ be finite for some value $s = -z$, with $z > 0$. If λ is so small that $\lambda(\tilde{\mathrm{g}}(-z) - 1) < z$, then the inequality

$$PA - \lambda^2 \frac{(\tilde{\mathrm{g}}(-z)-1)^2}{z^2 - \lambda^2(\tilde{\mathrm{g}}(-z)-1)^2} e^{-zt} \le \; \mathrm{PA}_0(t) \; \le PA + \lambda z \frac{\tilde{\mathrm{g}}(-z)-1}{z^2 - \lambda^2(\tilde{\mathrm{g}}(-z)-1)^2} e^{-zt}$$

holds for every $t \ge 0$ [6.12], with $PA = 1/(1+\lambda\, MTTR)$. If the repair rate $\mu(t) = \mathrm{g}(t)/(1-\mathrm{G}(t))$ is increasing, then $\tilde{\mathrm{g}}(s)$ is finite for $s = -1/MTTR$ and for $\lambda\, MTTR\,(\tilde{\mathrm{g}}(-1/MTTR) - 1) < 1$ it follows [6.12] that

$$\left| \mathrm{PA}_0(t) - PA \right| \le \lambda\, MTTR \frac{\tilde{\mathrm{g}}(-1/MTTR) - 1}{1 - (\lambda\, MTTR(\tilde{\mathrm{g}}(-1/MTTR) - 1))^2} e^{-t/MTTR}.$$

The important case of a *gamma distribution* with mean β/α and shape parameter $\beta \ge 3$, i.e. for $\mathrm{g}(x) = \alpha^\beta x^{\beta-1} e^{-\alpha x}/\Gamma(\beta)$, leads to $\left| \mathrm{PA}_0(t) - PA \right| \le \lambda\, MTTR\, e^{-t/MTTR}$ at least for $t \ge 3\,MTTR = 3\beta/\alpha$ [6.12].

6.2.5 Steady-State Behavior

For

$$p = \frac{MTTF}{MTTF + MTTR}, \quad \mathrm{F}_A(x) = \frac{1}{MTTF}\int_0^x (1-\mathrm{F}(y))\,dy, \quad \mathrm{G}_A(x) = \frac{1}{MTTR}\int_0^x (1-\mathrm{G}(y))\,dy \tag{6.57}$$

the *alternating renewal process* describing the time behavior of a *one-item repairable structure* is *stationary* (in *steady-state*), see Appendix A7.3. With p, $\mathrm{F}_A(t)$, and $\mathrm{G}_A(t)$ as per Eq. (6.57), the expressions for the point availability (6.42), average availability (6.43), interval reliability (6.44), joint availability (6.45), and the distribution functions of the forward recurrence times (6.46) and (6.47) take the values given by Eqs. (6.48) – (6.53) for all $t \ge 0$, see Example 6.4 for the point availability PA. This relationship between asymptotic and steady-state (stationary) behavior is important in practical applications because it allows the following interpretation:

> *A one-item repairable structure is in a steady-state (stationary behavior) if it began operating at the time $t = -\infty$ and will be considered only for $t \ge 0$, the time $t = 0$ being an arbitrary time point.*

Table 6.4 Results for the repairable one-item structure in steady-state (stationary behavior)

	Failure and repair rates		Remarks, assumptions
	Arbitrary	Constant	
1. Pr{*up* at $t = 0$} (p)	$\frac{MTTF}{MTTF + MTTR}$	$\frac{\mu}{\lambda+\mu}$	$MTTF = \mathrm{E}[\tau_i], \quad i \ge 1$ $MTTR = \mathrm{E}[\tau'_i], \quad i \ge 1$
2. Distributions of τ_0 ($\mathrm{F}_A(t) = \Pr\{\tau_0 \le t\}$)	$\frac{1}{MTTF}\int_0^t (1-\mathrm{F}(x))\,dx$	$1 - e^{-\lambda t}$	$\mathrm{F}_A(x)$ is also the distribution function of $\tau_{Ru}(t)$ as in Fig. 6.3 ($\mathrm{F}_A(x) = \Pr\{\tau_{Ru}(t) \le x\}$)
3. Distributions of τ'_0 ($\mathrm{G}_A(t) = \Pr\{\tau'_0 \le t\}$)	$\frac{1}{MTTR}\int_0^t (1-\mathrm{G}(x))\,dx$	$1 - e^{-\mu t}$	$\mathrm{G}_A(x)$ is also the distribution function of $\tau_{Rd}(t)$ as in Fig. 6.3 ($\mathrm{G}_A(x) = \Pr\{\tau_{Rd}(t) \le x\}$)
4. Renewal densities $\mathrm{h}_{du}(t)$ and $\mathrm{h}_{ud}(t)$	$\frac{1}{MTTF + MTTR}$	$\frac{\lambda\mu}{\lambda+\mu}$	$\mathrm{h}_{du}(t) = p\,\mathrm{h}_{duu}(t) + (1-p)\,\mathrm{h}_{dud}(t)$, $\mathrm{h}_{ud}(t) = p\,\mathrm{h}_{udu}(t) + (1-p)\,\mathrm{h}_{udd}(t)$, p as in point 1
5. Point availability (PA)	$\frac{MTTF}{MTTF + MTTR}$	$\frac{\mu}{\lambda+\mu}$	$PA = \Pr\{up \text{ at } t\}, \quad t \ge 0$
6. Average availability (AA)	$\frac{MTTF}{MTTF + MTTR}$	$\frac{\mu}{\lambda+\mu}$	$AA = \frac{1}{t}\mathrm{E}[\text{total up time in } (0,t]]$, $t > 0$
7. Interval reliability ($\mathrm{IR}(\theta)$)	$\frac{1}{MTTF + MTTR} \cdot \int_\theta^\infty (1-\mathrm{F}(x))\,dx$	$\frac{\mu}{\lambda+\mu} e^{-\lambda\theta}$	$\mathrm{IR}(\theta) = \Pr\{up \text{ in } [t, t+\theta]\}, \quad t \ge 0$
8. Joint availability ($\mathrm{JA}(\theta)$)	$\frac{MTTF \cdot \mathrm{PA}_{0e}(\theta)}{MTTF + MTTR}$	$(\frac{\mu}{\lambda+\mu})^2 + \frac{e^{-(\lambda+\mu)\theta}}{(\lambda+\mu)^2/\lambda\mu}$	$\mathrm{JA}(\theta) = \Pr\{up \text{ at } t \cap up \text{ at } t+\theta\}$, $\mathrm{PA}_{0e}(\theta) = \mathrm{PA}(\theta)$ as per Eq. (6.42) with $p = 1$ and $\mathrm{F}_A(t)$ as in point 2

λ = failure rate, μ = repair rate, *up* means in operating state

For a *constant failure rate* λ and a *constant repair rate* μ, the convergence to steady-state is *exponential with time constant* $\approx 1/\mu = MTTR$ (Eq. (6.55)). From the results of Section 6.2.4, it seems reasonable to assume that for most practical applications the time to reach steady-state is of the *order of magnitude of 10 MTTR*, according to the precision desired. Important results for the steady-state behavior of a repairable one-item structure are given in Table 6.4.

Example 6.4

Show that for a repairable one-item structure in steady-state, i.e. with p, $F_A(t)$, and $G_A(t)$ as given in Eq. (6.57), the point availability is $PA(t) = PA = MTTF/(MTTF + MTTR)$ for all $t \geq 0$.

Solution

Applying the Laplace transform to Eq. (6.42) and using Eqs. (A7.50) and (6.57) yields

$$\tilde{PA}(s) = \frac{MTTF}{MTTF + MTTR}\left(\frac{1}{s} - \frac{1 - \tilde{f}(s)}{s^2\, MTTF} + \frac{\frac{1 - \tilde{f}(s)}{s\, MTTF}\tilde{g}(s)}{1 - \tilde{f}(s)\tilde{g}(s)} \cdot \frac{1 - \tilde{f}(s)}{s}\right)$$

$$+ \frac{MTTR}{MTTF + MTTR} \cdot \frac{\frac{1 - \tilde{g}(s)}{s\, MTTR}}{1 - \tilde{f}(s)\tilde{g}(s)} \cdot \frac{1 - \tilde{f}(s)}{s},$$

and finally

$$\tilde{PA}(s) = \frac{MTTF}{MTTF + MTTR}\left(\frac{1}{s} - \frac{1 - \tilde{f}(s)}{s^2\, MTTF}\right) + \frac{[1 - \tilde{f}(s)][\tilde{g}(s) - \tilde{f}(s)\tilde{g}(s) + 1 - \tilde{g}(s)]}{s^2\,(MTTF + MTTR)[1 - \tilde{f}(s)\tilde{g}(s)]},$$

from which

$$\tilde{PA}(s) = \frac{MTTF}{MTTF + MTTR} \cdot \frac{1}{s},$$

and thus $PA(t) = PA$ for all $t \geq 0$.

6.3 Systems without Redundancy

The reliability block diagram of a *system without redundancy* consists of the series connection of all its elements E_1 to E_n, see Fig. 6.4. Each element E_i in Fig. 6.4 is characterized by the distribution functions $F_i(t)$ for the failure-free operating times and $G_i(t)$ for the repair times.

6.3.1 Series Structure with Constant Failure and Repair Rates for Each Element

In this section, *constant* failure and repair rates are assumed, i.e.

$$F_i(t) = 1 - e^{-\lambda_{0i}\, t} \tag{6.58}$$

and

$$G_i(t) = 1 - e^{-\mu_{i0}\, t} \tag{6.59}$$

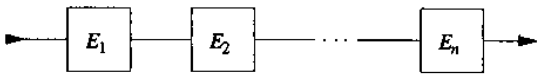


Figure 6.4 Reliability block diagram for a system without redundancy (series structure)

holds for $i = 1, \ldots, n$. Because of Eqs. (6.58) and (6.59), the stochastic behavior of the system is described by a (time) *homogeneous Markov process*. Let Z_0 be the system up state and Z_i the state in which element E_i is down. Taking assumption (6.2) into account, i.e. neglecting further failures during a repair at system level (in short: *no further failures at system down*), the corresponding *diagram of transition probabilities* in $(t, t + \delta t]$ is given in Fig. 6.5. Equations of Table 6.2 can be used to obtain the expressions for the reliability function, point availability and interval reliability. With $U = \{Z_0\}$, $\overline{U} = \{Z_1, \ldots, Z_n\}$ and the transition rates according to Fig. 6.5, the *reliability function* (see Table 6.2 for definitions) follows from

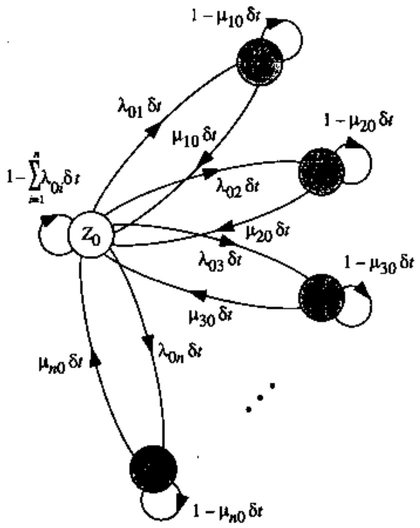


Figure 6.5 Diagram of the transition probabilities in $(t, t + \delta t]$ for a repairable series structure (constant failure and repair rates λ_{0i} and μ_{i0}, only one repair crew, no further failures at system down, arbitrary t, $\delta t \downarrow 0$, Markov process)

$$R_{S0}(t) = e^{-\lambda_{S0}\, t}, \qquad \text{with} \qquad \lambda_{S0} = \sum_{i=1}^{n} \lambda_{0i}, \tag{6.60}$$

and thus, for the *mean time to failure*,

$$MTTF_{S0} = \frac{1}{\lambda_{S0}}. \tag{6.61}$$

The *point availability* is given by

$$PA_{S0}(t) = P_{00}(t), \tag{6.62}$$

with $P_{00}(t)$ from (see Table 6.2)

$$P_{00}(t) = e^{-\lambda_{S0}\, t} + \sum_{i=1}^{n} \int_0^t \lambda_{0i}\, e^{-\lambda_{S0}\, x}\, P_{i0}(t-x)\, dx$$

$$P_{i0}(t) = \int_0^t \mu_{i0}\, e^{-\mu_{i0}\, x}\, P_{00}(t-x)\, dx, \qquad i = 1, \dots, n. \tag{6.63}$$

The solution Eq. (6.63) leads to the following Laplace transform (Table A9.7) for $PA_{S0}(t)$

$$\tilde{PA}_{S0}(s) = \frac{1}{s\left(1 + \sum\limits_{i=1}^{n} \dfrac{\lambda_{0i}}{s + \mu_{i0}}\right)}. \tag{6.64}$$

From Eq. (6.64) there follows the *asymptotic* and *steady-state* value of the *point* and *average availability*

$$PA_S = AA_S = \frac{1}{1 + \sum\limits_{i=1}^{n} \dfrac{\lambda_{0i}}{\mu_{i0}}}. \tag{6.65}$$

Because of the *constant* failure rate of all elements, the *interval reliability* can be directly obtained from Eq. (6.27) by

$$IR_{S0}(t, t+\theta) = PA_{S0}(t)\, e^{-\lambda_{S0}\, \theta}, \tag{6.66}$$

with the asymptotic and steady-state value

$$IR_S(\theta) = PA_S\, e^{-\lambda_{S0}\, \theta}. \tag{6.67}$$

6.3.2 Series Structure with Constant Failure Rate and Arbitrary Repair Rate for Each Element

Generalization of the repair time distribution functions $G_i(t)$, with densities $g_i(t)$ and $G_i(0) = 0$, leads to a *semi-Markov process* with state space $Z_0, \dots, Z_n$, as for the Markov process of Fig. 6.5. The *reliability function* and the *mean time to failure* are still given by Eqs. (6.60) and (6.61). For the *point availability* let us first compute the *semi-Markov transition probabilities* $Q_{ij}(x)$ using Table 6.2

$$\begin{aligned} Q_{0i}(x) &= \Pr\{\tau_{0i} \le x \cap \tau_{0k} > \tau_{0i}, \quad k \ne i\} \\ &= \int_0^x \lambda_{0i}\, e^{-\lambda_{0i}\, y} \prod_{k \ne i} e^{-\lambda_{0k}\, y}\, dy = \frac{\lambda_{0i}}{\lambda_{S0}}(1 - e^{-\lambda_{S0}\, x}) \\ Q_{i0}(x) &= G_i(x), \qquad i = 1, \dots, n. \end{aligned} \tag{6.68}$$

The system of integral Equations for the *transition probabilities* (conditional state probabilities) $P_{ij}(t)$ follows then from Table 6.2

$$P_{00}(t) = e^{-\lambda_{S0}\, t} + \sum_{i=1}^{n} \int_0^x \lambda_{0i}\, e^{-\lambda_{S0}\, x}\, P_{i0}(t-x)\, dx,$$

$$P_{i0}(t) = \int_0^t g_i(x)\, P_{00}(t-x)\, dx, \qquad i = 1, \dots, n. \tag{6.69}$$

For the Laplace transform of the *point availability* $PA_{S0}(t) = P_{00}(t)$ one obtains finally from Eq. (6.69)

$$\tilde{PA}_{S0}(s) = \frac{1}{s + \lambda_{S0} - \sum\limits_{i=1}^{n} \lambda_{0i}\, \tilde{g}_i(s)} = \frac{1}{s + \sum\limits_{i=1}^{n} \lambda_{0i}(1 - \tilde{g}_i(s))}, \tag{6.70}$$

from which follows the *asymptotic* and *steady-state value* of the point and *average availability*

$$PA_S = AA_S = \frac{1}{1 + \sum\limits_{i=1}^{n} \lambda_{0i}\, MTTR_i}, \tag{6.71}$$

with $\lim\limits_{s \to 0} (1 - \tilde{g}(s)) \approx s\, MTTR$, as per Eq. (6.54), and

$$MTTR_i = \int_0^\infty (1 - G_i(t))\, dt. \tag{6.72}$$

The *interval reliability* can be computed either from Eq. (6.66) with $PA_{S0}(t)$ from Eq. (6.70) or from Eq. (6.67) with PA_S from Eq. (6.71).

Example 6.5

A system consists of elements E_1 to E_4 which are necessary for the fulfillment of the required function (series structure). Let the failure rates $\lambda_1 = 10^{-3}\text{h}^{-1}$, $\lambda_2 = 0.5 \cdot 10^{-3}\text{h}^{-1}$, $\lambda_3 = 10^{-4}\text{h}^{-1}$, $\lambda_4 = 2 \cdot 10^{-3}\text{h}^{-1}$ be constant and assume that the repair time of all elements is lognormally distributed with parameters $\lambda = 0.5\text{h}^{-1}$ and $\sigma = 0.6$. The system has only one repair crew and no further failure can occur at system down (failures during repair are neglected). Determine the reliability function for a mission of duration $t = 168\text{h}$, the mean time to failure, the asymptotic and stationary values of the point and average availability, and he asymptotic and stationary values of the interval reliability for $\theta = 12\text{h}$.

Solution

The system failure rate is $\lambda_{S0} = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 36 \cdot 10^{-4}\text{h}^{-1}$ according to Eq. (6.60). The reliability function follows as $R_{S0}(t) = e^{-0.0036t}$, from which $R_{S0}(168\text{h}) \approx 0.55$. The mean time to failure is $MTTF_{S0} = 1/\lambda_{S0} \approx 278\text{h}$. The mean time to repair is obtained from Table A6.2 as $E[\tau'] = (e^{\sigma^2/2})/\lambda = MTTR \approx 2.4\text{h}$. For the asymptotic and steady-state values of the point and average availability as well as for the interval reliability for $\theta = 12\text{h}$ it follows from Eqs. (6.71) and (6.67) that $PA_S = AA_S = 1/(1 + 36 \cdot 10^{-4} \cdot 2.4) \approx 0.991$ and $IR_S(12) \approx 0.991 \cdot e^{-0.0036 \cdot 12} \approx 0.95$.

6.3.3 Series Structure with Arbitrary Failure and Repair Rates for Each Element

Generalization of repair and failure-free operating time distribution functions leads to a *nonregenerative stochastic process*. This model can be investigated using supplementary variables, or by approximating the distribution functions of the failure-free operating times in such a way that the involved stochastic process can be reduced to a regenerative process. Using for the approximation an *Erlang distribution function*, the process is semi-Markovian. As an example, let us consider the case of a two-element series structure (E_1, E_2) and assume that the repair times are arbitrary, with densities $g_{10}(t)$ and $g_{20}(t)$, and the failure-free operating times have densities

$$f_{01}(t) = \lambda_{01}^2 \, t \, e^{-\lambda_{01} t} \tag{6.73}$$

and

$$f_{02}(t) = \lambda_{02} \, e^{-\lambda_{02} t}. \tag{6.74}$$

Equation (6.73) is the density of the sum of two exponentially distributed random time intervals with density $\lambda_{01} e^{-\lambda_{01} t}$. Under these assumptions, the two-element series structure corresponds to a *1-out-of-2 standby redundancy* with constant failure λ_{01}, in series with an element with constant failure rate λ_{02}. Figure 6.6 gives the equivalent reliability block diagram and the corresponding *state transition diagram*. This diagram only *clarifies* the possible transitions and *can not be considered* as a the diagram of the transition probabilities in $(t, t+\delta t]$. Z_0 is the system up state, $Z_{1'}$ and $Z_{2'}$ are *supplementary states* necessary for computation only. For the *semi-Markov transition probabilities* $Q_{ij}(x)$one obtains (see Table 6.2)

$$Q_{01'}(x) = Q_{1'1}(x) = \frac{\lambda_{01}}{\lambda_{01} + \lambda_{02}} (1 - e^{-(\lambda_{01} + \lambda_{02})x})$$

$$Q_{02}(x) = Q_{1'2'}(x) = \frac{\lambda_{02}}{\lambda_{01} + \lambda_{02}} (1 - e^{-(\lambda_{01} + \lambda_{02})x})$$

$$Q_{20}(x) = Q_{2'1'}(x) = \int_0^x g_{20}(y) \, dy$$

$$Q_{10}(x) = \int_0^x g_{10}(y) \, dy. \tag{6.75}$$

From Eq. (6.75) it follows that (Table 6.2 and Eq. (6.54))

$$R_{S0}(t) = (1 + \lambda_{01} t) e^{-(\lambda_{01} + \lambda_{02})t}, \tag{6.76}$$

$$MTTF_{S0} = \frac{2\lambda_{01} + \lambda_{02}}{(\lambda_{01} + \lambda_{02})^2}, \tag{6.77}$$

$$\tilde{PA}_{S0}(s) = \tilde{P}_{00}(s) + \tilde{P}_{01'}(s) = \frac{(s + \lambda_{01} + \lambda_{02}(1 - \tilde{g}_{20}(s))) + \lambda_{01}}{[s + \lambda_{01} + \lambda_{02}(1 - \tilde{g}_{20}(s))]^2 - \lambda_{01}^2 \tilde{g}_{10}(s)}, \tag{6.78}$$

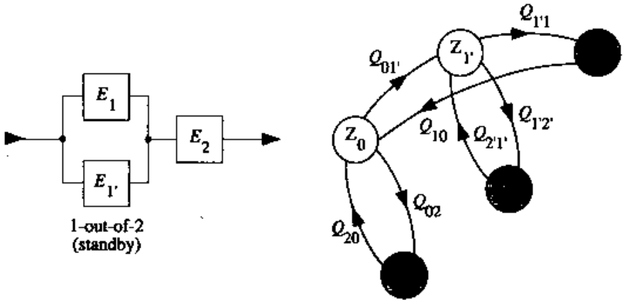


Figure 6.6 Equivalent reliability block diagram and *state transition diagram* of a two series element system (E_1 and E_2) with arbitrarily distributed repair times, constant failure rate for E_2, and Erlangian ($n = 2$) distributed failure-free operating time for E_1 (5-state semi-Markov process)

$$PA_S = AA_S = \frac{2}{2 + 2\lambda_{02}\, MTTR_2 + \lambda_{01}\, MTTR_1}, \tag{6.79}$$

$$\mathrm{IR}_S(\theta) = \frac{(2 + \lambda_{01}\theta)\, e^{-(\lambda_{01}+\lambda_{02})\theta}}{2 + 2\lambda_{02}\, MTTR_2 + \lambda_{01}\, MTTR_1}. \tag{6.80}$$

The *interval reliability* $\mathrm{IR}_{S0}(t, t+\theta)$ can be obtained from

$$\mathrm{IR}_{S0}(t, t+\theta) = \mathrm{P}_{00}(t)\mathrm{R}_{S0}(\theta) + \mathrm{P}_{01'}(t)\mathrm{R}_{S1'}(\theta)$$

with $\mathrm{R}_{S1}(\theta) = e^{-(\lambda_{01}+\lambda_{02})\theta}$, because of the constant failure rates λ_{01} and λ_{02}.

Important results for repairable series structures are summarized in Table 6.5.

Table 6.5 Results for a repairable system without redundancy (elements $E_1, \ldots, E_n$ in series), one repair crew, no further failures at system down (n repair crews would give $\mathrm{PA}_{S0}(t) = \Pi_i \mathrm{PA}_i(t)$

Quantity	Expression	Remarks, assumptions
1. Reliability function ($\mathrm{R}_{S0}(t)$)	$\prod_{i=1}^{n} \mathrm{R}_i(t)$	Independent elements (at least up to system failure)
2. Mean time to system failure ($MTTF_{S0}$)	$\int_0^{\infty} \mathrm{R}_{S0}(t)\,dt$	$\mathrm{R}_i(t) = e^{-\lambda_{0i}t} \rightarrow R_{S0}(t) = e^{-\lambda_{S0}t}$ and $MTTF_{S0} = 1/\lambda_{S0}$ with $\lambda_{S0} = \lambda_{01} + \ldots + \lambda_{0n}$
3. System failure rate ($\lambda_{S0}(t)$)	$\sum_{i=1}^{n} \lambda_{0i}(t)$	Independent elements (at least up to system failure)
4. Asymptotic and steady-state value of the point availability and average availability ($PA_S = AA_S$)	a) $\frac{1}{1 + \sum_{i=1}^{n} \frac{\lambda_{0i}}{\mu_{i0}}} \approx 1 - \sum_{i=1}^{n} \frac{\lambda_{0i}}{\mu_{i0}}$ *) b) $\frac{1}{1 + \sum_{i=1}^{n} \lambda_{0i}\, MTTR_i} \approx 1 - \sum_{i=1}^{n} \lambda_{0i}\, MTTR_i$ c) $\frac{2}{2 + 2\lambda_{02} MTTR_2 + \lambda_{01} MTTR_1}$	At system down, no further failures can occur: a) Constant failure rate λ_{0i} and constant repair rate μ_{i0} for each element ($i = 1, \ldots, n$) b) Constant failure rate λ_{0i} and arbitrary repair rate $\mu_{i0}(t)$ with $MTTR_i$ = mean time to repair for each element ($i = 1, \ldots, n$) c) 2-element series structure with failure rates $\lambda_{01}^2 t/(1 + \lambda_{01} t)$ for E_1 and λ_{02} for E_2
5. Asymptotic and steady-state value of the interval reliability ($\mathrm{IR}_S(\theta)$)	$PA_S\, e^{-\lambda_{S0}\theta}$	Each element has constant failure rate λ_{0i}, $\lambda_{S0} = \lambda_{01} + \ldots + \lambda_{0n}$

*) If *n repair crews* were available, then $PA_S = \Pi_i (1/(1 + \lambda_{0i}/\mu_{i0})) \approx 1 - \Sigma_i \lambda_{0i}/\mu_{i0}$ (for $\lambda_{0i}/\mu_{i0} << 1$)

6.4 1-out-of-2 Redundancy

The *1-out-of-2 redundancy*, also known as *1-out-of-2: G*, is the simplest redundant structure arising in practical applications. It consists of two elements E_1 and E_2, one of which is in the operating state and the other in reserve. When a failure occurs, one element is repaired while the other continues the operation. The system is down when an element fails while the other one is being repaired. Assuming ideal *switching* and *failure detection*, the reliability block diagram is a parallel connection of elements E_1 and E_2, see Fig. 6.7.

Investigations are based on the assumptions (6.1) to (6.7). This implies in particular, that the repair of a redundant element begins immediately on failure occurrence and is performed *without interruption of the system level operation*. The distribution functions of the repair times, and of the failure-free operating times are generalized step by step, beginning with the exponential distribution (memoryless), up to the case in which the process involved has only *one regeneration state* (Section 6.4.3). Influence of switching is considered in Sections 6.6 and 6.9.

6.4.1 1-out-of-2 Redundancy with Constant Failure and Repair Rates for Each Element

Because of the constant failure and repair rates, the time behavior of the 1-out-of-2 redundancy can be described by a (time) *homogeneous Markov process*. The number of states is 3 if elements E_1 and E_2 are identical (Fig. 6.8) and 5 if they are different (Fig. 6.9), the corresponding diagrams of transition probabilities in $(t, t + \delta t]$ are also given in Fig. A7.4.

Let us consider first the case of identical elements E_1 and E_2 (see Example 6.6 for different elements) and assume as distribution function of the failure-free operating times

$$\mathrm{F}(t) = 1 - e^{-\lambda t} \tag{6.81}$$

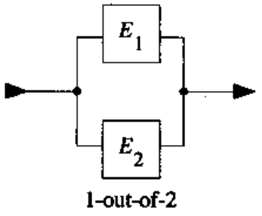


Figure 6.7 Reliability block diagram for a 1-out-of-2 redundancy (ideal failure detection and switching)

in the *operating state* and

$$F_r(t) = 1 - e^{-\lambda_r t} \tag{6.82}$$

in the *reserve state*. This includes *active* (parallel) redundancy for $\lambda_r = \lambda$, *warm* redundancy for $\lambda_r < \lambda$, and *standby* redundancy for $\lambda_r \equiv 0$. Repair times are assumed to be distributed (independently of λ_r) according to

$$G(t) = 1 - e^{-\mu t}. \tag{6.83}$$

For the investigation of more general situations (arbitrary load sharing, more than one repair crew, or other cases in which failure and/or repair rates change at a state transition) one can use the *birth and death process* introduced in Appendix A7.5.4. For all these cases, it is usual to perform investigations using the *method of differential equations* (Table 6.2 and Appendix A7.5.3.1). Figure 6.8 gives the *diagrams of transition probabilities* in $(t, t+\delta t]$ for the calculation of the point availability (Fig. 6.8a) and of the reliability function (Fig. 6.8b).

Considering the system behavior at times t and $t+\delta t$, the following *difference equations* can be established for the *state probabilities* $P_0(t)$, $P_1(t)$, and $P_2(t)$ according to Fig. 6.8a, where $P_i(t) = \Pr\{\text{process in } Z_i \text{ at } t\}$, $i = 0, 1, 2$.

$$\begin{aligned}
P_0(t+\delta t) &= P_0(t)(1-(\lambda+\lambda_r)\delta t) + P_1(t)\mu\,\delta t\\
P_1(t+\delta t) &= P_1(t)(1-(\lambda+\mu)\delta t) + P_0(t)(\lambda+\lambda_r)\delta t + P_2(t)\mu\,\delta t\\
P_2(t+\delta t) &= P_2(t)(1-\mu\,\delta t) + P_1(t)\lambda\,\delta t.
\end{aligned} \tag{6.84}$$

For $\delta t \downarrow 0$, it follows that

$$\begin{aligned}
\dot{P}_0(t) &= -(\lambda+\lambda_r)P_0(t) + P_1(t)\mu\\
\dot{P}_1(t) &= -(\lambda+\mu)P_1(t) + P_0(t)(\lambda+\lambda_r) + P_2(t)\mu\\
\dot{P}_2(t) &= -\mu P_2(t) + P_1(t)\lambda.
\end{aligned} \tag{6.85}$$

The system of differential equations (6.85) can also be obtained directly from Table 6.2 and Fig. 6.8a. Its solution leads to the state probabilities $P_i(t)$, $i = 0, 1, 2$. Assuming as *initial conditions* at $t = 0$, $P_0(0) = 1$ and $P_1(0) = P_2(0) = 0$, the above state probabilities are identical to the *transition probabilities* $P_{0i}(t)$, $i = 0, 1, 2$, i.e. $P_{00}(t) \equiv P_0(t)$, $P_{01}(t) \equiv P_1(t)$, and $P_{02}(t) \equiv P_2(t)$. The *point availability* $PA_{S0}(t)$ is then given by (see also Table 6.2)

$$PA_{S0}(t) = P_{00}(t) + P_{01}(t). \tag{6.86}$$

$PA_{S1}(t)$ or $PA_{S2}(t)$ could have been determined for other initial conditions. From Eq. (6.86) follows for the Laplace transform of $PA_{S0}(t)$

$$\tilde{PA}_{S0}(s) = \tilde{P}_{00}(s) + \tilde{P}_{01}(s) = \frac{((s+\mu)^2 + s\lambda)) + (s+\mu)(\lambda+\lambda_r)}{s[(s+\lambda+\lambda_r)(s+\lambda+\mu) + \mu(s+\mu)]}, \tag{6.87}$$

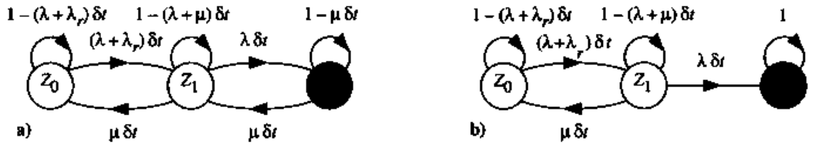


Figure 6.8 Diagram of the transition probabilities in $(t, t+\delta t]$ for a repairable 1-out-of-2 warm redundancy (two identical elements, constant failure (λ, λ_r) and repair (μ) rates, one repair crew, arbitrary t, $\delta t \downarrow 0$, Markov processes): **a)** For the point availability; **b)** For the reliability function

and thus for $t \to \infty$

$$\lim_{t\to\infty} PA_{S0}(t) = PA_S = \frac{\mu(\lambda+\lambda_r+\mu)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2} \approx 1 - \frac{\lambda(\lambda+\lambda_r)}{\mu(\lambda+\lambda_r+\mu)}. \tag{6.88}$$

If $PA_{S0}(t) = PA_S$, then PA_S is also the value of the point and *average availability* in the *steady-state* for all $t \ge 0$.

To compute the *reliability function* (by the method of differential equations) it is necessary to consider that the 1-out-of-2 redundancy will operate failure free in $(0, t]$ *only* if in this time interval the *down state at system level* (state Z_2 in Fig. 6.8) will *not be visited*. To recognize if the state Z_2 has been entered before t it is sufficient to make Z_2 *absorbing* (Fig. 6.8b). In this case, if Z_2 is entered the process *remains there indefinitely*, thus the probability of being in Z_2 at t is the probability of having entered Z_2 *before* the time t, i.e. the unreliability $1 - R_S(t)$. To avoid ambiguities, the *state probabilities* in Fig. 6.8b will be marked by an *apostrophe* (prime). The procedure is similar to that for Eq. (6.85) and leads to the following system of differential equations

$$\begin{aligned}
\dot{P}'_0(t) &= -(\lambda+\lambda_r)P'_0(t) + P'_1(t)\mu\\
\dot{P}'_1(t) &= -(\lambda+\mu)P'_1(t) + P'_0(t)(\lambda+\lambda_r)\\
\dot{P}'_2(t) &= P'_1(t)\lambda,
\end{aligned} \tag{6.89}$$

and to the corresponding state probabilities $P'_0(t)$, $P'_1(t)$, and $P'_2(t)$. With the *initial conditions* at $t = 0$, $P'_0(0) = 1$ and $P'_1(0) = P'_2(0) = 0$, the state probabilities $P'_0(t)$, $P'_1(t)$, and $P'_2(t)$ are identical to the *transition probabilities* $P'_{00}(t) \equiv P'_0(t)$, $P'_{01}(t) \equiv P'_1(t)$, and $P'_{02}(t) \equiv P'_2(t)$. The *reliability function* is then given by (Table 6.2)

$$R_{S0}(t) = P'_{00}(t) + P'_{01}(t). \tag{6.90}$$

With the initial condition $P'_1(0) = 1$, $R_{S1}(t)$ would have been obtained. Eq. (6.90) yields the following Laplace transform for $R_{S0}(t)$

$$\tilde{\mathrm{R}}_{S0}(s)=\frac{s+2\lambda+\lambda_r+\mu}{(s+\lambda+\lambda_r)(s+\lambda)+s\mu}, \tag{6.91}$$

from which the *mean time to failure* ($MTTF_{S0}=\tilde{\mathrm{R}}_{S0}(0)$) follows as

$$MTTF_{S0}=\frac{2\lambda+\lambda_r+\mu}{\lambda(\lambda+\lambda_r)}. \tag{6.92}$$

Important for practical applications is the situation for $\lambda,\lambda_r \ll \mu$. To investigate this case let us consider an *active redundancy* $(\lambda_r=\lambda)$. From Eq. (6.91) it follows that

$$\tilde{\mathrm{R}}_{S0}(s)=\frac{s+3\lambda+\mu}{(s+2\lambda)(s+\lambda)+s\mu}=\frac{s+3\lambda+\mu}{(s-r_1)(s-r_2)},$$

with

$$r_{1,2}=\frac{-(3\lambda+\mu)\pm\sqrt{(3\lambda+\mu)^2-8\lambda^2}}{2},$$

and thus, using Table A9.7b,

$$\mathrm{R}_{S0}(t)=\frac{r_2 e^{r_1 t}-r_1 e^{r_2 t}}{r_2-r_1}.$$

For $\lambda \ll \mu$, $r_1 \approx 0$, and $r_2 \approx -\mu$ yielding

$$\mathrm{R}_{S0}(t)\approx e^{r_1 t}.$$

Using $\sqrt{1-\varepsilon}\approx 1-\varepsilon/2$ for $2r_1=-(3\lambda+\mu)(1-\sqrt{1-8\lambda^2/(3\lambda+\mu)^2})$ leads to $r_1\approx -2\lambda^2/(3\lambda+\mu)$. $\mathrm{R}_{S0}(t)$ can thus be *approximated by a decreasing exponential function* with time constant $MTTF_{S0}\approx(3\lambda+\mu)/2\lambda^2$. This result is *important*. It shows that:

> *For $\lambda \ll \mu$, a repairable 1-out-of-2 active redundancy with constant failure rate λ and constant repair rate μ behaves approximately like a one-item structure with constant failure rate $\lambda_S \approx 2\lambda^2/(3\lambda+\mu)$; an equivalent repair rate μ_S for the one-item structure can be obtained by comparing the equations for the steady-state point availability and leads to $\mu_S\approx\lambda_S/(1-PA_S)\approx\mu$, see Tables 6.4 and 6.10.*

Extension of the above result to warm redundancy ($\lambda_r<\lambda$) leads to

$$\mathrm{R}_{S0}(t)\approx e^{-\lambda_{S0}t} \qquad \text{with} \qquad \lambda_{S0}=\frac{1}{MTTF_{S0}}\approx\frac{\lambda(\lambda+\lambda_r)}{2\lambda+\lambda_r+\mu}. \tag{6.93}$$

As in all these considerations, $\lambda_r=\lambda$ yields active and $\lambda_r\equiv 0$ standby redundancy.

Because of the *memoryless property* of the (time) homogeneous Markov process, the *interval reliability* can be obtained directly from the *transition probabilities* $\mathrm{P}_{ij}(t)$ and the reliability functions $\mathrm{R}_{Si}(t)$, see Table 6.2. Assuming the initial condition $\mathrm{P}_0(0)=1$ yields

$$\mathrm{IR}_{S0}(t,t+\theta)=\mathrm{P}_{00}(t)\mathrm{R}_{S0}(\theta)+\mathrm{P}_{01}(t)\mathrm{R}_{S1}(\theta). \tag{6.94}$$

The Laplace transform of $\mathrm{IR}_{S0}(t,t+\theta)$ is then given by

$$\tilde{\mathrm{IR}}_{S0}(s,\theta)=\frac{[(s+\mu)^2+s\lambda]\mathrm{R}_{S0}(\theta)+(s+\mu)(\lambda+\lambda_r)\mathrm{R}_{S1}(\theta)}{s[(s+\lambda+\lambda_r)(s+\lambda+\mu)+\mu(s+\mu)]},$$

which leads to the following asymptotic and steady-state value (Table 6.2)

$$\mathrm{IR}_S(\theta)=P_0\mathrm{R}_{S0}(\theta)+P_1\mathrm{R}_{S1}(\theta)=\frac{\mu^2\mathrm{R}_{S0}(\theta)+\mu(\lambda+\lambda_r)\mathrm{R}_{S1}(\theta)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2}\approx\mathrm{R}_{S0}(\theta). \tag{6.95}$$

To compare the effectiveness of computational methods, let us now express the reliability function, point availability, and interval reliability using the *method of integral equations* (Appendix A7.5.3.2). The $\mathrm{Q}_{ij}(x)$ are given according to Eq. (A7.102) and Fig. 6.8a by

$$\mathrm{Q}_{01}(x)=\Pr\{\tau_{01}\le x\}=1-\Pr\{\tau_{01}>x\}=1-e^{-\lambda x}e^{-\lambda_r x}=1-e^{-(\lambda+\lambda_r)x}$$

$$\mathrm{Q}_{10}(x)=\Pr\{\tau_{10}\le x\cap\tau_{12}>\tau_{10}\}=\int_0^x\mu e^{-\mu y}e^{-\lambda y}dy=\frac{\mu}{\lambda+\mu}(1-e^{-(\lambda+\mu)x})$$

$$\mathrm{Q}_{12}(x)=\Pr\{\tau_{12}\le x\cap\tau_{10}>\tau_{12}\}=\int_0^x\lambda e^{-\lambda y}e^{-\mu y}dy=\frac{\lambda}{\lambda+\mu}(1-e^{-(\lambda+\mu)x})$$

$$\mathrm{Q}_{21}(x)=\Pr\{\tau_{21}\le x\}=1-e^{-\mu x}.$$

From Table 6.2 it follows then that

$$\mathrm{R}_{S0}(t)=e^{-(\lambda+\lambda_r)t}+\int_0^t(\lambda+\lambda_r)e^{-(\lambda+\lambda_r)x}\mathrm{R}_{S1}(t-x)dx$$

$$\mathrm{R}_{S1}(t)=e^{-(\lambda+\mu)t}+\int_0^t\mu e^{-(\lambda+\mu)x}\mathrm{R}_{S0}(t-x)dx, \tag{6.96}$$

for the *reliability functions* $\mathrm{R}_{S0}(t)$ and $\mathrm{R}_{S1}(t)$, as well as

$$\mathrm{P}_{00}(t)=e^{-(\lambda+\lambda_r)t}+\int_0^t(\lambda+\lambda_r)e^{-(\lambda+\lambda_r)x}\mathrm{P}_{10}(t-x)\,dx$$

$$\mathrm{P}_{10}(t)=\int_0^t\mu e^{-(\lambda+\mu)x}\mathrm{P}_{00}(t-x)dx+\int_0^t\lambda e^{-(\lambda+\mu)x}\mathrm{P}_{20}(t-x)dx$$

$$\mathrm{P}_{20}(t)=\int_0^t\mu e^{-\mu x}\mathrm{P}_{10}(t-x)dx$$

and

$$P_{01}(t) = \int_0^t (\lambda + \lambda_r) e^{-(\lambda+\lambda_r)x}\, P_{11}(t-x)\, dx$$

$$P_{11}(t) = e^{-(\lambda+\mu)t} + \int_0^t \mu\, e^{-(\lambda+\mu)x}\, P_{01}(t-x)\, dx + \int_0^t \lambda\, e^{-(\lambda+\mu)x}\, P_{21}(t-x)\, dx$$

$$P_{21}(t) = \int_0^t \mu\, e^{-\mu x}\, P_{11}(t-x)\, dx, \tag{6.97}$$

for the transition probabilities. The solution of Eqs. (6.96) and (6.97) yields Eqs. (6.87), (6.91), and (6.94). Equations (6.96) and (6.97) show how the use of integral Equations can lead to quicker solutions than differential equations.

Table 6.6 summarizes the main results of Section 6.4.1. It gives *approximate equations* valid for $\lambda \ll \mu$ and distinguishes between the cases of active redundancy ($\lambda_r = \lambda$), warm redundancy ($\lambda_r < \lambda$), and standby redundancy ($\lambda_r \equiv 0$). From Table 6.6, the improvement in $MTTF_{S0}$) through repair, without interruption of operation at system level, is given as lower and upper bounds by

$$\lambda\, MTTF_{S0} = \quad \underset{active}{\frac{\mu}{2\lambda} = \frac{MTBF}{2MTTR}} \qquad \underset{standby}{\frac{\mu}{\lambda} = \frac{MTBF}{MTTR}}$$

Investigation of the *unavailability* in steady-state $1 - PA_S$ leads to

$$1 - PA_S = 1 - AA_S = \quad \underset{active}{2(\frac{\lambda}{\mu})^2 = 2(\frac{MTTR}{MTBF})^2} \qquad \underset{standby}{(\frac{\lambda}{\mu})^2 = (\frac{MTTR}{MTBF})^2}$$

The above results can easily be extended to cover situations in which failure or repair rates are modified at *state changes*, for example because of *load sharing*, *differences within the element*, *repair priority*, etc. These cases, simply modify the transition rates on the diagram of transition probabilities in $(t, t+\delta t]$, see for example Figs. 2.12 and A7.4 to A7.6.

Example 6.6

Find the mean time to failure $MTTF_{S0}$ and the asymptotic and steady-state value of the point availability PA_S for a 1-out-of-2 active redundancy with two *different elements* E_1 and E_2, constant failure rates λ_1, λ_2, and constant repair rates μ_1, μ_2 (one repair crew).

Table 6.6 Reliability function $R_{S0}(t)$ for $\lambda, \lambda_r \ll \mu$, mean time to failure $MTTF_{S0}$, steady-state availability $PA_S = AA_S$, and interval reliability $IR_S(\theta)$ of a repairable 1-out-of-2 redundancy with identical elements E_1 and E_2 (constant failure rates λ, λ_r and repair rate μ, one repair crew)

	Standby ($\lambda_r \equiv 0$)	Warm ($\lambda_r < \lambda$)	Active ($\lambda_r = \lambda$)
$R_{S0}(t)^*$	$\approx e^{-\frac{\lambda^2 t}{2\lambda+\mu}}$	$\approx e^{-\frac{\lambda(\lambda+\lambda_r)t}{2\lambda+\lambda_r+\mu}}$	$\approx e^{-\frac{2\lambda^2 t}{3\lambda+\mu}}$
$MTTF_{S0}{}^*$	$\frac{2\lambda+\mu}{\lambda^2} \approx \frac{\mu}{\lambda^2}$	$\frac{2\lambda+\lambda_r+\mu}{\lambda(\lambda+\lambda_r)} \approx \frac{\mu}{\lambda(\lambda+\lambda_r)}$	$\frac{3\lambda+\mu}{2\lambda^2} \approx \frac{\mu}{2\lambda^2}$
$PA_S = AA_S{}^{**}$	$\frac{\mu(\lambda+\mu)}{\lambda(\lambda+\mu)+\mu^2}$ $\approx 1-(\lambda/\mu)^2$	$\frac{\mu(\lambda+\lambda_r+\mu)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2}$ $\approx 1-\lambda(\lambda+\lambda_r)/\mu^2$	$\frac{\mu(2\lambda+\mu)}{2\lambda(\lambda+\mu)+\mu^2}$ $\approx 1-2(\lambda/\mu)^2$
$IR_S(\theta)^{**}$	$\approx R_{S0}(\theta)$	$\approx R_{S0}(\theta)$	$\approx R_{S0}(\theta)$

λ, λ_r = failure rate (r = reserve), μ = repair rate, * new at $t = 0$, ** asymptotic and steady-state value

Solution

Figure 6.9 gives the reliability block diagram and the diagram of transition probabilities in $(t, t+\delta t]$. $MTTF_{S0}$ and PA_S can be calculated from appropriate systems of algebraic equations. According to Table 6.2 and considering Fig. 6.9 it follows for the *mean time to failure* that

$$MTTF_{S0} = \frac{1}{\lambda_1+\lambda_2}(1 + \lambda_1\, MTTF_{S1} + \lambda_2\, MTTF_{S2})$$

$$MTTF_{S1} = \frac{1}{\lambda_2+\mu_1}(1 + \mu_1\, MTTF_{S0})$$

$$MTTF_{S2} = \frac{1}{\lambda_1+\mu_2}(1 + \mu_2\, MTTF_{S0}),$$

which leads to

$$MTTF_{S0} = \frac{(\lambda_1+\mu_2)(\lambda_2+\mu_1) + \lambda_1(\lambda_1+\mu_2) + \lambda_2(\lambda_2+\mu_1)}{\lambda_1\lambda_2(\lambda_1+\lambda_2+\mu_1+\mu_2)}, \tag{6.98}$$

and in particular for $\lambda_1 \ll \mu_1$ and $\lambda_2 \ll \mu_2$,

$$MTTF_{S0} \approx \frac{\mu_1\mu_2}{\lambda_1\lambda_2(\mu_1+\mu_2)}. \tag{6.99}$$

As for Eq. (6.93), the *reliability function* can be expressed by

$$R_{S0}(t) \approx e^{-\lambda_{S0}t} \quad \text{with} \quad \lambda_{S0} = \frac{1}{MTTF_{S0}} \approx \frac{\lambda_1\lambda_2(\mu_1+\mu_2)}{\mu_1\mu_2} = \lambda_1\lambda_2(\frac{1}{\mu_1} + \frac{1}{\mu_2}). \tag{6.100}$$

For the asymptotic and steady-state value of the point availability and average availability $PA_S = P_0 + P_1 + P_2$ holds with P_0, P_1, and P_2 as solution of (Table 6.2)

$$
\begin{aligned}
(\lambda_1+\lambda_2)P_0 &= \mu_1 P_1 + \mu_2 P_2\\
(\lambda_2+\mu_1)P_1 &= \lambda_1 P_0 + \mu_2 P_4\\
(\lambda_1+\mu_2)P_2 &= \lambda_2 P_0 + \mu_1 P_3\\
\mu_1 P_3 &= \lambda_2 P_1\\
\mu_2 P_4 &= \lambda_1 P_2\\
P_0 + P_1 + P_2 + P_3 + P_4 &= 1.
\end{aligned}
$$

One (arbitrarily chosen, because of the implicitly assumed ergodicity) of the first five equations must be dropped and replaced by the sixth one (linear dependence). The solution yields P_0 through P_4, from which

$$PA_S = \frac{\mu_1\mu_2[\mu_1\mu_2 + (\lambda_1+\lambda_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)]}{\mu_1^2\mu_2^2 + \mu_1\mu_2(\lambda_1+\lambda_2)(\lambda_1+\lambda_2+\mu_1+\mu_2) + \lambda_1\lambda_2\{\mu_1^2+\mu_2^2+(\lambda_1+\lambda_2)(\mu_1+\mu_2)\}} \tag{6.101}$$

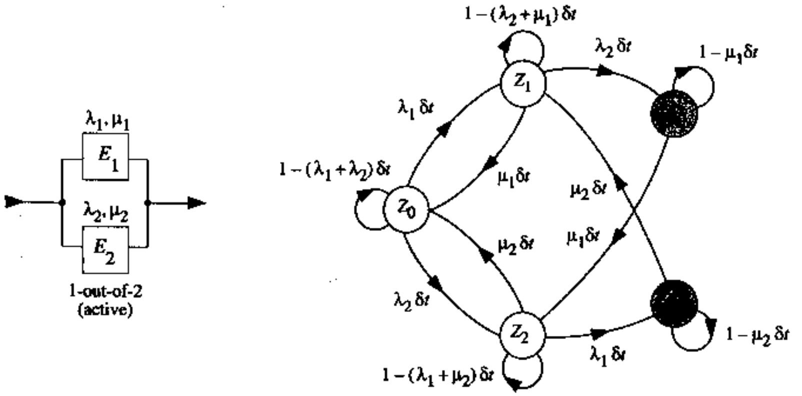
Equation (6.101) can also be written in the form

$$PA_S = \frac{1}{1+\dfrac{\lambda_1\lambda_2[\mu_1^2+\mu_2^2+(\lambda_1+\lambda_2)(\mu_1+\mu_2)]}{\mu_1\mu_2[\mu_1\mu_2+(\lambda_1+\lambda_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)]}}, \tag{6.102}$$

yielding, for $\lambda_1 \ll \mu_1$ and $\lambda_2 \ll \mu_2$,

$$PA_S \approx 1 - \frac{\lambda_1\lambda_2}{\mu_1^2\mu_2^2}(\mu_1^2+\mu_2^2) = 1 - \frac{\lambda_1}{\mu_1}\frac{\lambda_2}{\mu_2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right). \tag{6.103}$$

With $\lambda_1 = \lambda_2 = \lambda$ and $\mu_1 = \mu_2 = \mu$, Eqs. (6.98) and (6.101) become Eqs. (6.92) and (6.88), respectively (with $\lambda_r = \lambda$).



$\rho_{01} = \rho_{24} = \lambda_1$; $\rho_{02} = \rho_{13} = \lambda_2$; $\rho_{10} = \rho_{32} = \mu_1$; $\rho_{20} = \rho_{41} = \mu_2$
(for $R_S(t)$ set $\rho_{32} = \rho_{41} = 0$)

Figure 6.9 Reliability block diagram and diagram of transition probabilities in $(t, t+\delta t]$ for a repairable 1-out-of-2 active redundancy with different elements, ideal failure detection and switch, const. failure rates λ_1, λ_2 and repair rates μ_1, μ_2 (one repair crew, arbitr. t, $\delta t \downarrow 0$, Markov process)

6.4.2 1-out-of-2 Redundancy with Constant Failure Rate and Arbitrary Repair Rate for Each Element

Consider now a 1-out-of-2 *warm redundancy* with 2 identical elements E_1 and E_2, failure-free times distributed according to Eqs. (6.81) and (6.82), and repair times with mean *MTTR*, distributed according to an arbitrary distribution function $G(t)$ with $G(0) = 0$ and density $g(t)$. The time behavior of this system can be described by a process with states Z_0, Z_1, and Z_2. Because of the arbitrary repair rate, only states Z_0 and Z_1 are *regeneration states.* These states constitute a semi-Markov process *embedded* in the original *semi-regenerative process* (Fig. A.7.10). The *semi-Markov transition probabilities* $Q_{ij}(x)$ are given by Eq. (A7.167). Setting these quantities in the equations of Table 6.2 (SMP) by considering $Q_0(x) = Q_{01}(x)$ and $Q_1(x) = Q_{10}(x) + Q'_{12}(x)$, see Example A7.12, it follows that

$$
\begin{aligned}
R_{S0}(t) &= e^{-(\lambda+\lambda_r)t} + \int_0^t (\lambda+\lambda_r)e^{-(\lambda+\lambda_r)x} R_{S1}(t-x)\,dx\\
R_{S1}(t) &= e^{-\lambda t}(1-G(t)) + \int_0^t g(x)e^{-\lambda x} R_{S0}(t-x)\,dx,
\end{aligned} \tag{6.104}
$$

for the *reliability functions* $R_{S0}(t)$ and $R_{S1}(t)$, with solution

$$\tilde{R}_{S0}(s) = \frac{s+\lambda+(\lambda+\lambda_r)(1-\tilde{g}(s+\lambda))}{(s+\lambda)[(s+(\lambda+\lambda_r)(1-\tilde{g}(s+\lambda))]} \tag{6.105}$$

and in particular

$$MTTF_{S0} = \frac{1}{\lambda} + \frac{1}{(\lambda+\lambda_r)(1-\tilde{g}(\lambda))}, \tag{6.106}$$

and to

$$
\begin{aligned}
P_{00}(t) &= e^{-(\lambda+\lambda_r)t} + \int_0^t (\lambda+\lambda_r)e^{-(\lambda+\lambda_r)x} P_{10}(t-x)\,dx\\
P_{10}(t) &= \int_0^t g(x)e^{-\lambda x} P_{00}(t-x)\,dx + \int_0^t g(x)(1-e^{-\lambda x})P_{10}(t-x)\,dx\\
P_{01}(t) &= \int_0^t (\lambda+\lambda_r)e^{-(\lambda+\lambda_r)x} P_{11}(t-x)\,dx\\
P_{11}(t) &= (1-G(t))e^{-\lambda t}\\
&\quad + \int_0^t g(x)e^{-\lambda x} P_{01}(t-x)\,dx + \int_0^t g(x)(1-e^{-\lambda x})P_{11}(t-x)\,dx
\end{aligned} \tag{6.107}
$$

for the *transition probabilities* $P_{ij}(t)$ of the embedded semi-Markov process with states Z_0 and Z_1. The Laplace transform of the *point availability* $PA_{S0}(t) = P_{00}(t) + P_{01}(t)$ follows then as a solution of Eq. (6.107)

$$P\tilde{A}_{S0}(s) = \frac{(s+\lambda)(1-\tilde{g}(s)) + \lambda_r(1-\tilde{g}(s+\lambda)) + \lambda + s\,\tilde{g}(s+\lambda)}{(s+\lambda)[(s+\lambda+\lambda_r)(1-\tilde{g}(s)) + s\,\tilde{g}(s+\lambda)]}, \tag{6.108}$$

and leads to the asymptotic and steady-state value of the *point availability* PA_S and *average availability* AA_S (considering $\lim_{s\to 0}(1-\tilde{g}(s)) \approx s \cdot MTTR$, as per Eq. (6.54))

$$PA_S = AA_S = \frac{\lambda + \lambda_r(1-\tilde{g}(\lambda))}{\lambda(\lambda+\lambda_r)\,MTTR + \lambda\,\tilde{g}(\lambda)}, \tag{6.109}$$

where

$$MTTR = \int_0^\infty (1-G(x))\,dx \tag{6.110}$$

and $\tilde{g}(\lambda)$ is the Laplace transform of the density $g(t)$ for $s = \lambda$, see Eq. (6.88) for $g(t) = \mu e^{-\mu t}$, i.e. $\tilde{g}(\lambda) = \mu/(\lambda+\mu)$, and Examples 6.7 and 6.8 for the approximation of $\tilde{g}(\lambda)$. Calculation of the *interval reliability* is difficult because state Z_1 is regenerative only at its occurrence point (Fig. A7.10). However, if $\tilde{g}(\lambda) \to 1$ or $\lambda\,MTTR << 1$, the asymptotic value of the state probability for Z_1 ($P_1 = \lim_{t\to\infty} P_{01}(t)$) becomes very small with respect to the state probability for Z_0 ($P_0 = \lim_{t\to\infty} P_{00}(t)$). For the asymptotic and steady-state value of the *interval reliability* it holds then that

$$IR_S(\theta) \approx P_0\,R_{S0}(\theta) = \frac{\lambda\,\tilde{g}(\lambda)}{\lambda(\lambda+\lambda_r)\,MTTR + \lambda\,\tilde{g}(\lambda)} R_{S0}(\theta). \tag{6.111}$$

In many practical applications, $\lambda\,MTTR < 0.01$. In such cases, Eq. (6.111) can be further simplified to

$$IR_S(\theta) \approx R_{S0}(\theta). \tag{6.112}$$

Example 6.7

Let the density $g(t)$ of the repair time τ' of a system with constant failure rate $\lambda>0$ be continuous and assume furthermore that $\lambda\,E[\tau'] = \lambda\,MTTR << 1$ and $\lambda\sqrt{Var[\tau']} << 1$. Investigate the quantity $\tilde{g}(\lambda)$ for $\lambda \to 0$.

Solution

For $\lambda \to 0$, $\lambda\,MTTR << 1$, and $\lambda\sqrt{Var[\tau']} << 1$, the three first terms of the series expansion of $e^{-\lambda t}$ lead to

$$\tilde{g}(\lambda) = \int_0^\infty g(t) e^{-\lambda t} dt \approx \int_0^\infty g(t)\left(1 - \lambda t + \frac{(\lambda t)^2}{2}\right) dt = 1 - \lambda\,E[\tau'] + E[\tau'^2]\,\lambda^2/2.$$

From this, follows the *approximate expression*

$$\tilde{g}(\lambda) \approx 1 - \lambda\,MTTR + \frac{\lambda^2\,(MTTR^2 + Var[\tau'])}{2}. \tag{6.113}$$

In many practical applications,

$$\tilde{g}(\lambda) \approx 1 - \lambda\,MTTR \tag{6.114}$$

is a sufficiently good approximation, however not in computing steady-state availabilities (Eq. (6.114) would give for Eq. (6.109) $PA_S = 1$, thus Eq. (6.113) has to be used).

Example 6.8

In a 1-out-of-2 warm redundancy with identical elements E_1 and E_2 let the failure rates λ in the operating state and λ_r in the reserve state be constant. For the repair time let us assume that it is distributed according to $G(t) = 1 - e^{-\mu'(t-\psi)}$ for $t \geq \psi$ and $G(t) = 0$ for $t < \psi$, with $MTTR = 1/\mu > \psi$. Assuming $\lambda\psi << 1$, investigate the influence of ψ on the mean time to failure $MTTR_{S0}$ and on the asymptotic and steady-state value of the point availability PA_S.

Solution

With

$$\tilde{g}(\lambda) = \int_\psi^\infty \mu' e^{-\mu'(t-\psi)-\lambda t} dt = \frac{\mu'}{\lambda+\mu'} e^{-\lambda\psi} \approx \frac{\mu'}{\lambda+\mu'}(1-\lambda\psi)$$

and considering that

$$MTTR = \int_0^\infty t\,g(t)\,dt = \int_\psi^\infty t\,\mu' e^{-\mu'(t-\psi)} dt = \psi + \frac{1}{\mu'} \equiv \frac{1}{\mu},$$

i.e., $\mu' = \frac{\mu}{1-\mu\psi}$ obviously with $0 \leq \psi < 1/\mu$. Eqs. (6.106) and (6.109) lead to the *approximate expressions*

$$MTTF_{S0,\psi>0} \approx \frac{2\lambda + \lambda_r + \mu(1-\lambda\psi)}{\lambda(\lambda+\lambda_r)}$$

and

$$PA_{S,\psi>0} \approx \frac{\mu(\lambda + \lambda_r + \mu(1-\lambda\psi))}{(\lambda+\lambda_r)(\lambda + \mu(1-\lambda\psi)) + \mu^2(1-\lambda\psi)}.$$

On the other hand, $\psi = 0$ leads to $1 - \tilde{g}(\lambda) = \lambda/(\lambda+\mu)$ and thus (Eqs. (6.92) and (6.88))

$$MTTF_{S0,\psi=0} = \frac{2\lambda + \lambda_r + \mu}{\lambda(\lambda+\lambda_r)} \quad \text{and} \quad PA_{S,\psi=0} = \frac{\mu(\lambda+\lambda_r+\mu)}{(\lambda+\mu)(\lambda+\lambda_r)+\mu^2}.$$

Assuming $\mu >> \lambda, \lambda_r$ yields (with $0 \leq \lambda\psi < \lambda/\mu$)

$$\frac{MTTF_{S0,\psi>0}}{MTTF_{S0,\psi=0}} \approx 1 - \lambda\psi \quad \text{and} \quad \frac{PA_{S,\psi>0}}{PA_{S,\psi=0}} \approx 1 + \lambda\psi\frac{\lambda+\lambda_r}{\mu} - (\lambda\psi)^2 \approx 1. \tag{6.115}$$

Equation (6.115) allows the conclusion to be made that for $\lambda MTTR << 1$, the *shape* of the distribution function of the repair time has (as long as *MTTR* is unchanged) a *small influence* on the mean time to failure $MTTF_{S0}$, and on the asymptotic and steady-state value of the point (and average) availability PA_S of a 1-out-of-2 redundancy. This important result can be extended to complex structures.

Example 6.9

A 1-out-of-2 warm redundancy with identical elements E_1 and E_2 has a failure rate $\lambda = 10^{-2}\,h^{-1}$ in the operating state and $\lambda_r = 5 \cdot 10^{-3}\,h^{-1}$ in the reserve state. Repair times are lognormally distributed with mean $E[\tau'] = MTTR = 2.4\,h$ and variance $Var[\tau'] = 0.6\,h^2$. Compute the mean time to failure $MTTF_{S0}$ and the asymptotic and steady-state value of the point (and average) availability PA_S with *approximate expressions*: (i) $\tilde{g}(\lambda)$ from Eq. (6.114); (ii) $\tilde{g}(\lambda)$ from Eq. (6.113); (iii) $g(t) = \mu' e^{-\mu'(t-\psi)}$, $t \ge \psi$, $\psi = 1.4\,h$ and $1/\mu' = 1\,h$; (iv) $g(t) = \mu e^{-\mu t}$ and $1/\mu = 2.4\,h$.

Solution

(i) With $\tilde{g}(\lambda) = 0.976$ it follows (Eq. (6.106)) that $MTTF_{S0} \approx 2878\,h$ and (Eq. (6.109)) $PA_S = 1$. (ii) With $\tilde{g}(\lambda) \approx 0.9763$ it follows (Eq. (6.106)) that $MTTF_{S0} \approx 2915\,h$ and (Eq. (6.109)) $PA_S \approx 0.9996$. (iii) Example 6.8 yields $MTTF_{S0,\psi=1.4\,h} \approx 2906\,h$ and $PA_{S,\psi=1.4\,h} \approx 0.9996$. (iv) From Eqs. (6.92) and (6.88) it follows that $MTTF_{S0} \approx 2944\,h$ and $PA_S \approx 0.9992$. These results confirm the conclusions of Example 6.8 that for $\lambda MTTR << 1$, the *shape* of the distribution function of the repair time has a *small influence* on $MTTF_{S0}$ and PA_S.

6.4.3 1-out-of-2 Redundancy with Constant Failure Rate only in the Reserve State

Generalization of the repair and failure rates of a 1-out-of-2 redundancy leads to a *nonregenerative stochastic process*. However, in many practical applications it can be assumed that the failure rate in the *reserve state* is constant. If this holds, and the 1-out-of-2 redundancy has only one repair crew, then the process involved has exactly *one regeneration state* [6.4 (1975)].

To see this, consider a 1-out-of-2 warm redundancy, satisfying assumptions (6.1) to (6.7), with failure-free operating times distributed according to $F(t)$ in the operating state and $V(t) = 1 - e^{-\lambda_r t}$ in the reserve state, and repair times distributed according to $G(t)$ for repair of failures in the operating state and $W(t)$ for repair of failures in the reserve state ($F(0)=V(0)= G(0)=W(0)=0$, densities $f(t)$, $g(t)$, $w(t)$). Figure 6.10a shows a time schedule of such a system and Fig. 6.10b gives the *state transition diagram* (to visualize possible state transitions only) of the corresponding stochastic process. States Z_0, Z_1, and Z_2 are up states. State Z_1 is the only *regeneration state* present here (Fig. 6.10a). The occurrence of Z_1 brings the process to a situation of total independence from the previous time development. It is therefore sufficient to investigate the time behavior from $t = 0$ up to the first regeneration point and between *two consecutive regeneration points* (Appendix A7.7).

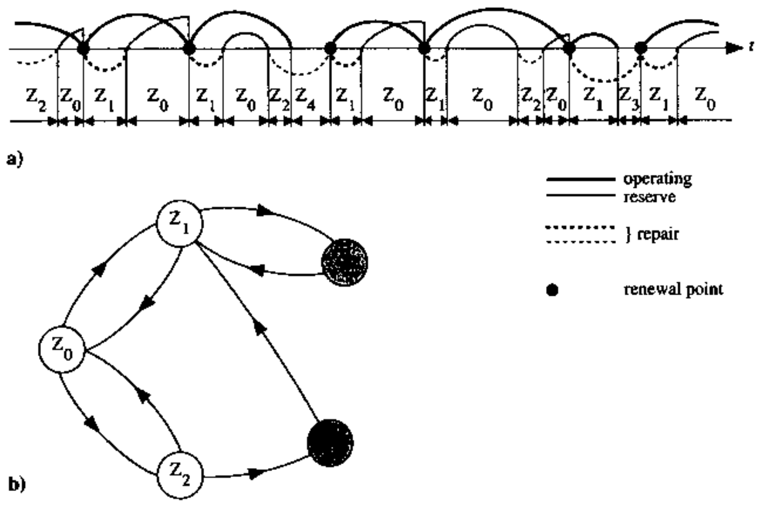


Figure 6.10 Repairable 1-out-of-2 warm redundancy with constant failure rate λ_r only in the reserve state and one repair crew; **a)** Time schedule (repair times are shown greatly exaggerated); **b)** state transition diagram (to visualize possible state transitions, only Z_1 is a regeneration state)

Let us consider first the case in which the regenerative state Z_1 is *entered* at $t = 0$ (S_{EP0}) and let S_{EP1} be the first renewal point after $t = 0$. The *reliability function* $R_{S1}(t)$ is given by (see Table 6.2 for definitions)

$$R_{S1}(t) = 1 - F(t) + \int_0^t u_1(x) R_{S1}(t-x)\,dx, \tag{6.116}$$

with

$1 - F(t) = \Pr\{$failure-free operating time of the element operating at $t = 0$ is $> t \mid Z_1$ entered at $t = 0\}$

and

$$\int_0^t u_1(x) R_{S1}(t-x)\,dx = \Pr\{(S_{EP1} \le t \cap up \text{ in } (S_{EP1}, t]) \mid Z_1 \text{ entered at } t = 0\}$$

The *first renewal point* S_{EP1} occurs at the time x (i.e. within the interval $(x, x+dx]$) only if at this time *the operating element fails and the reserve element is ready to enter the operating state*. The quantity $u_1(x)$, defined as

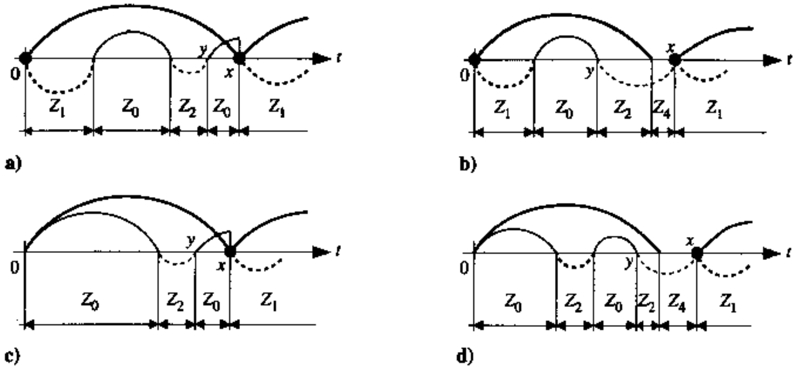


Figure 6.11 Time schedules at $t = 0$ for the 1-out-of-2 redundancy according to Fig. 6.10

$$\mathrm{u}_1(x) = \lim_{\delta x \downarrow 0} \frac{1}{\delta x} \Pr\{(x < S_{EP1} \le x + \delta x \mid Z_1 \text{ entered at } t = 0\}$$

can be obtained as (Fig. 6.11a)

$$\mathrm{u}_1(x) = \mathrm{f}(x)\mathrm{PA}_d(x), \tag{6.117}$$

with

$$\begin{aligned}\mathrm{PA}_d(x) &= \Pr\{\text{reserve element } up \text{ at time } x \mid Z_1 \text{ entered at } t = 0\} \\ &= \int_0^x \mathrm{h}'_{dud}(y) e^{-\lambda_r (x-y)} dy\end{aligned} \tag{6.118}$$

and

$$\mathrm{h}'_{dud}(y) = \mathrm{g}(y) + \mathrm{g}(y) * \mathrm{v}(y) * \mathrm{w}(y) + \mathrm{g}(y) * \mathrm{v}(y) * \mathrm{w}(y) * \mathrm{v}(y) * \mathrm{w}(y) + \dots \tag{6.119}$$

The *point availability* is given by

$$\mathrm{PA}_{S1}(t) = 1 - \mathrm{F}(t) + \int_0^t \mathrm{u}_1(x)\mathrm{PA}_{S1}(t-x)dx + \int_0^t \mathrm{u}_2(x)\mathrm{PA}_{S1}(t-x)dx, \tag{6.120}$$

with

$$\begin{aligned}1 - \mathrm{F}(t) = \Pr\{&\text{failure-free operating time of the element operating} \\ &\text{at } t = 0 \text{ is } > t \mid Z_1 \text{ entered at } t = 0\},\end{aligned}$$

$$\int_0^t \mathrm{u}_1(x)\mathrm{PA}_{S1}(t-x)dx = \Pr\{(S_{EP1} \le t \cap up \text{ at } t) \mid Z_1 \text{ entered at } t = 0\},$$

and

$$\begin{aligned}\int_0^t \mathrm{u}_2(x)\mathrm{PA}_{S1}(t-x)dx = \Pr\{(S_{EP1} \le t \cap \text{system failed in } (0, S_{EP1}] \\ \cap\, up \text{ at } t) \mid Z_1 \text{ entered at } t = 0\}.\end{aligned}$$

The quantity $\mathrm{u}_2(x)$, defined as

$$\begin{aligned}\mathrm{u}_2(x) = \lim_{\delta x \downarrow 0} \frac{1}{\delta x} \Pr\{(x < S_{EP1} \le x + \delta x \cap \text{system failed in } (0, x]) \\ \mid Z_1 \text{ entered at } t = 0\},\end{aligned}$$

can be obtained as (Fig. 6.11b)

$$\mathrm{u}_2(x) = \mathrm{g}(x)\mathrm{F}(x) + \int_0^x \mathrm{h}'_{udd}(y)\mathrm{w}(x-y)(\mathrm{F}(x) - \mathrm{F}(y))dy \tag{6.121}$$

with

$$\mathrm{h}'_{udd}(y) = \mathrm{g}(y) * \mathrm{v}(y) + \mathrm{g}(y) * \mathrm{v}(y) * \mathrm{w}(y) * \mathrm{v}(y) + \dots. \tag{6.122}$$

$\mathrm{u}_1(x) + \mathrm{u}_2(x)$ is the density of the *embedded renewal process* with *renewal points* $0, S_{EP1}, S_{EP2}, \dots$.

Consider now the case in which at $t = 0$ the state Z_0 is *entered*. The *reliability function* $\mathrm{R}_{S0}(t)$ is given by

$$\mathrm{R}_{S0}(t) = 1 - \mathrm{F}(t) + \int_0^t \mathrm{u}_3(x)\mathrm{R}_{S1}(t-x)dx, \tag{6.123}$$

with (Fig. 6.11c)

$$\begin{aligned}\mathrm{u}_3(x) &= \lim_{\delta x \downarrow 0} \frac{1}{\delta x} \Pr\{x < S_{EP1} \le x + \delta x \\ &\qquad \mid Z_0 \text{ entered at } t = 0\} = \mathrm{f}(x)\mathrm{PA}_0(x),\end{aligned} \tag{6.124}$$

where

$$\begin{aligned}\mathrm{PA}_0(x) &= \Pr\{\text{reserve element } up \text{ at time } x \mid Z_0 \text{ entered at } t = 0\} \\ &= e^{-\lambda_r x} + \int_0^x \mathrm{h}'_{duu}(y) e^{-\lambda_r (x-y)} dy,\end{aligned} \tag{6.125}$$

with

$$\mathrm{h}'_{duu}(y) = \mathrm{v}(y) * \mathrm{w}(y) + \mathrm{v}(y) * \mathrm{w}(y) * \mathrm{v}(y) * \mathrm{w}(y) + \ldots . \tag{6.126}$$

The *point availability* $\mathrm{PA}_{S0}(t)$ is given by

$$\mathrm{PA}_{S0}(t) = 1 - \mathrm{F}(t) + \int_0^t \mathrm{u}_3(x)\mathrm{PA}_{S1}(t-x)\,dx + \int_0^t \mathrm{u}_4(x)\mathrm{PA}_{S1}(t-x)\,dx, \tag{6.127}$$

with (Fig. 6.11d)

$$\begin{aligned}\mathrm{u}_4(x) &= \lim_{\delta x \downarrow 0} \frac{1}{\delta x} \Pr\{(x < S_{EP1} \le x + \delta x \cap \text{system failed in } (0, x]) \\ &\quad \mid Z_0 \text{ entered at } t = 0\} = \int_0^x \mathrm{h}'_{udu}(y)\,\mathrm{w}(x-y)(\mathrm{F}(x) - \mathrm{F}(y))\,dy\end{aligned} \tag{6.128}$$

and

$$\mathrm{h}'_{udu}(y) = \mathrm{v}(y) + \mathrm{v}(y) * \mathrm{w}(y) * \mathrm{v}(y) + \mathrm{v}(y) * \mathrm{w}(y) * \mathrm{v}(y) * \mathrm{w}(y) * \mathrm{v}(y) + \ldots . \tag{6.129}$$

Equations (6.116), (6.120), (6.123), and (6.127) can be solved using Laplace transforms. However, analytical difficulties can arise when calculating Laplace transforms for $\mathrm{F}(t)$, $\mathrm{G}(t)$, $\mathrm{W}(t)$, $\mathrm{u}_1(t)$, $\mathrm{u}_2(t)$, $\mathrm{u}_3(t)$ and $\mathrm{u}_4(t)$ as well as at the inversion of the final equations. Easier is the computation of the the *mean time to failure* $MTTF_{S0}$ and of the asymptotic and steady-state values of the *point and average availability* $PA_S = AA_S$, for which the following expressions can be found (Eqs. (6.123) and (6.116) for $MTTF_{S0}$ and Eq. (6.120) for PA_S)

$$MTTF_{S0} = MTTF\,(1 + \frac{\int_0^\infty \mathrm{u}_3(t)\,dt}{1 - \int_0^\infty \mathrm{u}_1(t)\,dt}) \tag{6.130}$$

and

$$\lim_{t\to\infty} \mathrm{PA}_{S1}(t) = PA_S = AA_S = \frac{MTTF}{\int_0^\infty t(\mathrm{u}_1(t) + \mathrm{u}_2(t))\,dt}, \tag{6.131}$$

with

$$MTTF = \int_0^\infty (1 - \mathrm{F}(t))\,dt. \tag{6.132}$$

The model investigated in this section has as special cases that of Section 6.4.2, with $\mathrm{F}(t) = 1 - e^{-\lambda t}$ and $\mathrm{W}(t) = \mathrm{G}(t)$, as well as the 1-out-of-2 *standby redundancy* with arbitrarily distributed failure-free operating times and repair times, see Example 6.10.

Example 6.10

Using the results of Section 6.4.3, develop the expressions for the reliability function $R_{S0}(t)$ and the point availability $\mathrm{PA}_{S0}(t)$ for a 1-out-of-2 *standby* redundancy with a failure-free operating time distributed according to $\mathrm{F}(t)$, with density $\mathrm{f}(t)$, and repair time distributed according to $\mathrm{G}(t)$ with density $\mathrm{g}(t)$.

Solution

For a standby redundancy, $\mathrm{u}_1(x) = \mathrm{f}(x)\mathrm{G}(x)$, $\mathrm{u}_2(x) = \mathrm{g}(x)\mathrm{F}(x)$, $\mathrm{u}_3(x) = \mathrm{f}(x)$, and $\mathrm{u}_4(x) \equiv 0$ (Eqs. (6.117), (6.121), (6.124), and (6.128)). From this, the expressions for $\mathrm{R}_{S0}(t)$, $\mathrm{R}_{S1}(t)$, $\mathrm{PA}_{S0}(t)$, and $\mathrm{PA}_{S1}(t)$ can be derived. The Laplace transforms of $\mathrm{R}_{S0}(t)$ and $\mathrm{PA}_{S0}(t)$ are

$$\tilde{\mathrm{R}}_{S0}(s) = \frac{1 - \tilde{\mathrm{f}}(s)}{s} + \frac{\tilde{\mathrm{f}}(s)(1 - \tilde{\mathrm{f}}(s))}{s(1 - \tilde{\mathrm{u}}_1(s))}, \tag{6.133}$$

$$\mathrm{P\tilde{A}}_{S0}(s) = \frac{1 - \tilde{\mathrm{f}}(s)}{s} + \frac{\tilde{\mathrm{f}}(s)(1 - \tilde{\mathrm{f}}(s))}{s(1 - \tilde{\mathrm{u}}_1(s) - \tilde{\mathrm{u}}_2(s))}, \tag{6.134}$$

with

$$\tilde{\mathrm{u}}_1(s) = \int_0^\infty \mathrm{f}(t)\mathrm{G}(t)e^{-st}\,dt$$

and

$$\tilde{\mathrm{u}}_2(s) = \int_0^\infty \mathrm{g}(t)\mathrm{F}(t)e^{-st}\,dt.$$

The mean time to failure $MTTF_{S0}$ follows then from Eq. (6.133)

$$MTTF_{S0} = \int_0^\infty (1 - \mathrm{F}(t))\,dt + \frac{\int_0^\infty (1 - \mathrm{F}(t))\,dt}{1 - \int_0^\infty \mathrm{f}(t)\mathrm{G}(t)\,dt}. \tag{6.135}$$

For the asymptotic and steady-state value of the point and average availability $PA_S = AA_S$, Eq. (6.134) yields

$$PA_S = AA_S = \frac{\int_0^\infty (1 - \mathrm{F}(t))\,dt}{\int_0^\infty t\,d(\mathrm{F}(t)\mathrm{G}(t))}. \tag{6.136}$$

Important results for a 1-out-of-2 redundancy with arbitrary repair rates, and failure rates as general as possible within regenerative processes, are given in Table 6.7.

Table 6.7 Mean time to failure $MTTF_{S0}$, asymptotic and steady-state values of the point and average availability $PA_S = AA_S$ as well as of the interval reliability $IR_S(\theta)$ for a repairable 1-out-of-2 redundancy with one repair crew, arbitrary repair rates, and failure rates as general as possible

			Standby ($\lambda_r \equiv 0$)	Warm ($\lambda_r < \lambda$)		Active ($\lambda_r = \lambda$)
Element E_1 and E_2	Distribution of the failure-free operating times	OS	$F(t)$	$1-e^{-\lambda t}$	$F(t)$	$1-e^{-\lambda t}$
		RS	-	$1-e^{-\lambda_r t}$	$1-e^{-\lambda_r t}$	$1-e^{-\lambda t}$
	Distribution of the repair times	OS	$G(t)$	$G(t)$	$G(t)$	$G(t)$
		RS	–	$G(t)$	$W(t)$	$G(t)$
	Mean of the failure-free operating times		$MTTF = \int_0^\infty (1-F(t))\,dt$	$\frac{1}{\lambda}$ or $\frac{1}{\lambda_r}$	$MTTF$ or $\frac{1}{\lambda_r}$	$\frac{1}{\lambda}$
	Mean of the repair times		$MTTR = \int_0^\infty (1-G(t))\,dt$	$MTTR$	$MTTR$ or $MTTR_W$	$MTTR$
1-out-of-2 redundancy	Mean time to failure ($MTTF_{S0}$)		$MTTF + \frac{MTTF}{1-\int_0^\infty f(t)G(t)\,dt}$	$\frac{1}{\lambda}+\frac{1}{(\lambda+\lambda_r)(1-\tilde{g}(\lambda))}$ $\approx \frac{1}{\lambda}(1+\frac{1}{(\lambda+\lambda_r)MTTR})$	$MTTF + \frac{MTTF\int_0^\infty u_3(t)\,dt}{1-\int_0^\infty u_1(t)\,dt}$	$\frac{1}{\lambda}+\frac{1}{2\lambda(1-\tilde{g}(\lambda))}$ $\approx \frac{1}{\lambda}(1+\frac{1}{2\lambda MTTR})$
	Point and average availability ($PA_S = AA_S$)*		$\frac{MTTF}{\int_0^\infty t\,d(F(t)G(t))}$	$\frac{\lambda+\lambda_r(1-\tilde{g}(\lambda))}{\lambda(\lambda+\lambda_r)MTTR+\lambda\tilde{g}(\lambda)}$	$\frac{MTTF}{\int_0^\infty t(u_1(t)+u_2(t))\,dt}$	$\frac{2-\tilde{g}(\lambda)}{2\lambda MTTR+\tilde{g}(\lambda)}$
	Interval reliability ($IR_S(\theta)$)*		$\approx R_{S0}(\theta)$	$\approx R_{S0}(\theta)$	$\approx R_{S0}(\theta)$	$\approx R_{S0}(\theta)$

OS = operating state, RS = reserve state, * asymptotic and steady-state value

6.5 *k*-out-of-*n* Redundancy

A *k-out-of-n redundancy,* also known as *k-out-of-n: G,* often consists of *n* identical elements, of which *k* are necessary for the required function and *n - k* stay in the reserve state. Assuming ideal failure detection and switching, the reliability block diagram is as given in Fig. 6.12. The investigation in this Section will assume identical elements $E_1, \ldots, E_n$, only one repair crew, and *no further failures at system down* (failures during a repair at system level are neglected, as per assumption (6.2)). Section 6.5.1 will consider the case of *warm redundancy* with constant failure rates λ in the operation state and $\lambda_r < \lambda$ in the *reserve state,* and constant repair rate μ (only one repair crew as per assumption (6.3)). This case includes *active redundancy* ($\lambda_r = \lambda$) and *standby redundancy* ($\lambda_r \equiv 0$). An extension to cover other situations in which the failure rate is modified at state changes, e.g. for a particular *load sharing,* is easily possible using the general equations for the *birth and death process* developed in Appendix A7.5.4. Section 6.5.2 investigates then a 2-out-of-3 active redundancy with constant failure rate and arbitrary repair rate. The influence of *switching elements* is considered in Sections 6.6 to 6.9.

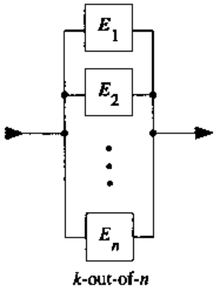


Figure 6.12 Reliability block diagram for a *k*-out-of-*n* redundancy (ideal failure detection and switching)

6.5.1 *k*-out-of-*n* Warm Redundancy with Identical Elements and Constant Failure and Repair Rates

Assuming constant failure and repair rates, the time behavior of the *k*-out-of-*n* redundancy can be investigated by a *birth and death process* (Appendix A7.5.4). Figure 6.13 gives the corresponding diagram of the transition probabilities in $(t, t+\delta t]$. Using Fig. 6.13 and Table 6.2, the following system of differential equations can be established for the state probabilities $P_j(t) = \Pr\{\text{in state } Z_j \text{ at } t\}$ of a *k*-out-of-*n* warm redundancy with one repair crew and no further failures at system down

$$\dot{P}_0(t) = -\nu_0 P_0(t) + \mu P_1(t)$$

$$\dot{P}_j(t) = \nu_{j-1} P_{j-1}(t) - (\nu_j + \mu) P_j(t) + \mu P_{j+1}(t), \qquad j = 1, \ldots, n-k,$$

$$\dot{P}_{n-k+1}(t) = \nu_{n-k} P_{n-k}(t) - \mu P_{n-k+1}(t), \tag{6.137}$$

with

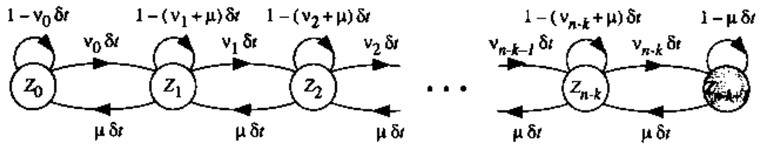


Figure 6.13 Diagram of the transition probabilities in $(t, t+\delta t]$ for a repairable k-out-of-n *warm* redundancy (n identical elements, constant failure and repair rates, no further failures at system down, one repair crew, arbitrary t, $\delta t \downarrow 0$, birth-and-death process, Z_0 to Z_{n-k} are up states)

$$\nu_j = k\lambda + (n-k-j)\lambda_r, \qquad j = 0, \ldots, n-k. \tag{6.138}$$

For the investigation of more general situations (arbitrary load sharing, more than one repair crew, or other cases in which failure and/or repair rates change at a state transition) one can use the *birth and death process* introduced in Appendix A7.5.4. The solution of the system (6.137) with the *initial conditions* at $t=0$, $\mathrm{P}_i(0)=1$ and $\mathrm{P}_j(0)=0$ for $j \neq i$, yields the *point availability* (see Table 6.2 for definitions)

$$\mathrm{PA}_{Si}(t) = \sum_{j=0}^{n-k} \mathrm{P}_{ij}(t), \tag{6.139}$$

with $\mathrm{P}_{ij}(t) \equiv \mathrm{P}_j(t)$ from Eq. (6.137) with $\mathrm{P}_i(0)=1$. In many practical applications, only the *asymptotic* and *steady-state* value of the point availability PA_S is required. This can be obtained by setting $\dot{\mathrm{P}}_j(t) = 0$ and $\mathrm{P}_j(t) = P_j$ in Eq. (6.137), see Appendix A7.5.4. The solution is

$$PA_S = \sum_{j=0}^{n-k} P_j = 1 - P_{n-k+1}, \quad \text{with } P_j = \frac{\pi_j}{\sum\limits_{i=0}^{n-k+1} \pi_i}, \quad \pi_i = \frac{\nu_0 \cdots \nu_{i-1}}{\mu^i}, \quad \pi_0 = 1. \tag{6.140}$$

PA_S is also the asymptotic and steady-state value of the *average availability* AA_S. As shown in Example A 7.11 (Eq. (A7.143)), for $2\nu_j < \mu$ it holds that

$$P_j \geq \sum_{i=j+1}^{n-k+1} P_i, \qquad j = 0, \ldots, n-k.$$

From this, the following *bounds for* PA_S can be used in many practical applications (assuming $2\nu_j < \mu$, $j = 0, \ldots, n$-k) to obtain an *approximate expression* for PA_S.

$$\sum_{j=0}^{i} P_j \leq PA_S < P_i + \sum_{j=0}^{i} P_j, \qquad i = 0, \ldots, n-k. \tag{6.141}$$

The *reliability function* follows from Table 6.2 and Fig. 6.13

$$\mathrm{R}_{S0}(t) = e^{-\nu_0 t} + \int_0^t \nu_0 e^{-\nu_0 x} \mathrm{R}_{S1}(t-x)\,dx$$

$$\mathrm{R}_{Sj}(t) = e^{-(\nu_j+\mu)t} + \int_0^t (\nu_j \mathrm{R}_{Sj+1}(t-x) + \mu \mathrm{R}_{Sj-1}(t-x)) e^{-(\nu_j+\mu)x}\,dx, \qquad j = 1, \ldots, n-k-1,$$

$$\mathrm{R}_{Sn-k}(t) = e^{-(\nu_{n-k}+\mu)t} + \int_0^t \mu \mathrm{R}_{Sn-k-1}(t-x) e^{-(\nu_{n-k}+\mu)x}\,dx, \tag{6.142}$$

with ν_i as in Eq. (6.138). Similar results hold for the *mean time to failure*.

$$MTTF_{S0} = \frac{1}{\nu_0} + MTTF_{S1}$$

$$MTTF_{Sj} = \frac{1}{\nu_j + \mu}(1 + \nu_j\, MTTF_{Sj+1} + \mu\, MTTF_{Sj-1}), \qquad j = 1, \ldots, n-k-1,$$

$$MTTF_{Sn-k} = \frac{1}{\nu_{n-k} + \mu}(1 + \mu\, MTTF_{Sn-k-1}). \tag{6.143}$$

The solution of Eqs. (6.142) and (6.143), shows that both $\mathrm{R}_{Si}(t)$ and $MTTF_{Si}$ *depend* on $n-k$. This leads for $n-k=1$ to

$$\tilde{\mathrm{R}}_{S0_1}(s) = \frac{s + \nu_0 + \nu_1 + \mu}{(s+\nu_0)(s+\nu_1) + s\mu}$$

$$MTTF_{S0_1} = \frac{\nu_0 + \nu_1 + \mu}{\nu_0\,\nu_1} \approx \frac{\mu}{\nu_0\,\nu_1}, \tag{6.144}$$

and for $n-k=2$ to

$$\tilde{\mathrm{R}}_{S0_2}(s) = \frac{(s+\nu_0+\nu_1+\mu)(s+\nu_2+\mu) + \nu_1(\nu_0-\mu)}{s(s+\nu_0+\nu_1+\mu)(s+\nu_2+\mu) + \nu_0\,\nu_1\,\nu_2 + s\,\nu_1(\nu_0-\mu)}$$

$$MTTF_{S0_2} = \frac{\nu_2(\nu_0+\nu_1+\mu) + \mu(\nu_0+\mu) + \nu_0\,\nu_1}{\nu_0\,\nu_1\,\nu_2} \approx \frac{\mu^2}{\nu_0\,\nu_1\,\nu_2}. \tag{6.145}$$

This holds similarly for the *point availability* PA_S (Table 6.8).

Because of the constant failure rate, the *interval reliability* follows directly from

$$\mathrm{IR}_{Si}(t, t+\theta) = \sum_{j=0}^{n-k} \mathrm{P}_{ij}(t) \mathrm{R}_{Sj}(\theta), \qquad i = 0, \ldots, n-k \tag{6.146}$$

with $\mathrm{P}_{ij}(t)$ as in Eq. (6.139) and $\mathrm{R}_{Si}(\theta)$ from Eq. (6.142) with $t = \theta$. The asymptotic and steady-state value is then given by

$$\mathrm{IR}_S(\theta) = \sum_{j=0}^{n-k} P_j \, \mathrm{R}_{Sj}(\theta), \tag{6.147}$$

with P_j from Eq. (6.141). Table 6.8 summarizes the main results for a k-out-of-n warm redundancy with constant failure and repair rates.

Assuming (for comparative investigations in Section 6.7) *n repair crews*, i. e. independent elements, following *approximate expressions* can be given for *active redundancy* with arbitrary n and k [6.13, 6.19]

$$MTTF_{S0} \approx \frac{1}{k\lambda\binom{n}{k}}(\mu/\lambda)^{n-k}, \qquad n \text{ repair crews}, \quad \lambda/\mu \ll 1 \tag{6.148}$$

$$PA_S \approx 1 - \frac{k}{n-k+1}\binom{n}{k}(\lambda/\mu)^{n-k+1}, \qquad n \text{ repair crews}, \quad \lambda/\mu \ll 1. \tag{6.149}$$

6.5.2 *k*-out-of-*n* Active Redundancy with Identical Elements, Constant Failure Rate, and Arbitrary Repair Rate

Generalization of the repair rate leads to stochastic process with *only two regeneration states* (Z_0 and Z_1 in Fig. 6.13). The investigation is similar to that of the 1-out-of-2 redundancy of Section 6.4.2. As an example let us consider a 2-out-of-3 active redundancy with 3 identical elements, with failure rate λ and repair time distributed according to $\mathrm{G}(t)$ with $\mathrm{G}(0) = 0$ and density $g(t)$. Using Fig. 6.14a, the following integral equation can be established for the *reliability function* $\mathrm{R}_{S0}(t)$, see Table 6.2 for definitions,

$$\mathrm{R}_{S0}(t) = e^{-3\lambda t} + \int_0^t 3\lambda e^{-3\lambda x} e^{-2\lambda(t-x)}(1 - \mathrm{G}(t-x))\,dx$$
$$+ \int_0^t\int_0^y 3\lambda e^{-3\lambda x} g(y-x) e^{-2\lambda(y-x)} \mathrm{R}_{S0}(t-y))\,dx\,dy. \tag{6.150}$$

The Laplace transform of $\mathrm{R}_{S0}(t)$ follows then as

$$\tilde{\mathrm{R}}_{S0}(s) = \frac{s + 5\lambda - 3\lambda\tilde{g}(s+2\lambda)}{(s+2\lambda)(s+3\lambda) - 3\lambda(s+2\lambda)\tilde{g}(s+2\lambda)}, \tag{6.151}$$

and the *mean time to failure* as

$$MTTF_{S0} = \frac{5 - 3\tilde{g}(2\lambda)}{6\lambda(1 - \tilde{g}(2\lambda))}. \tag{6.152}$$

Table 6.8 Mean time to failure $MTTF_{S0}$ and asymptotic as well as steady-state values of the point and average availability $PA_S = AA_S$, and of the interval reliability for a repairable k-out-of-n *warm* redundancy with n identical elements, one repair crew, constant failure rate (λ in operating state and $\lambda_r < \lambda$ in reserve state), constant repair rate μ, and no further failures at system down

		Mean time to failure ($MTTF_{S0}$)	Asymptotic and steady-state point and average availability ($PA_S = AA_S$)	Interval reliability ($\mathrm{IR}_S(\theta)$)*
$n-k=1$	gen. case	$\frac{\nu_0+\nu_1+\mu}{\nu_0\nu_1} \approx \frac{\mu}{\nu_0\nu_1}$	$\frac{\nu_0\mu+\mu^2}{\nu_0\nu_1+\nu_0\mu+\mu^2} \approx 1-\frac{\nu_0\nu_1}{\mu^2}$	$\approx \mathrm{R}_{S0}(\theta)$
	$n=2$ $k=1$	$\frac{2\lambda+\lambda_r+\mu}{\lambda(\lambda+\lambda_r)}$	$\frac{\mu(\lambda+\lambda_r+\mu)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2} \approx 1-\frac{\lambda(\lambda+\lambda_r)}{\mu^2}$	$\approx \mathrm{R}_{S0}(\theta)$
	$n=3$ $k=2$	$\frac{4\lambda+\lambda_r+\mu}{2\lambda(2\lambda+\lambda_r)}$	$\frac{\mu(2\lambda+\lambda_r+\mu)}{(2\lambda+\lambda_r)(2\lambda+\mu)+\mu^2} \approx 1-\frac{2\lambda(2\lambda+\lambda_r)}{\mu^2}$	$\approx \mathrm{R}_{S0}(\theta)$
$n-k=2$	gen. case	$\frac{\nu_2(\nu_0+\nu_1+\mu)}{\nu_0\nu_1\nu_2} + \frac{\mu(\nu_0+\mu)+\nu_0\nu_1}{\nu_0\nu_1\nu_2} \approx \frac{\mu^2}{\nu_0\nu_1\nu_2}$	$\frac{\nu_0\nu_1\mu+\nu_0\mu^2+\mu^3}{\nu_0\nu_1\nu_2+\nu_0\nu_1\mu+\nu_0\mu^2+\mu^3} \approx 1-\frac{\nu_0\nu_1\nu_2}{\mu^3}$	$\approx \mathrm{R}_{S0}(\theta)$
	$n=3$ $k=1$	$\frac{\mu^2}{\lambda(\lambda+\lambda_r)(\lambda+2\lambda_r)}$	$\approx 1-\frac{\lambda(\lambda+\lambda_r)(\lambda+2\lambda_r)}{\mu^3}$	$\approx \mathrm{R}_{S0}(\theta)$
	$n=5$ $k=3$	$\approx \frac{\mu^2}{3\lambda(3\lambda+\lambda_r)(3\lambda+2\lambda_r)}$	$\approx 1-\frac{3\lambda(3\lambda+\lambda_r)(3\lambda+2\lambda_r)}{\mu^3}$	$\approx \mathrm{R}_{S0}(\theta)$
$n-k$ arbitrary		$\approx \frac{\mu^{n-k}}{\nu_0\cdots\nu_{n-k}}$	$\approx 1-\frac{\nu_0\cdots\nu_{n-k}}{\mu^{n-k+1}}$	$\approx \mathrm{R}_{S0}(\theta)$

$\nu_i = k\lambda + (n-k-i)\lambda_r$, $i = 0, \ldots, n-k$; λ, λ_r = failure rate ($\lambda_r = \lambda \rightarrow$ *active red.* $\rightarrow \nu_0 \cdots \nu_{n-k} = \lambda^{n-k+1} n!/(k-1)!$, $\lambda_r \equiv 0 \rightarrow$ *standby red.* $\rightarrow \nu_0 \cdots \nu_{n-k} = (k\lambda)^{n-k+1}$); μ = repair rate; $\mathrm{R}_{Sj}(t)$ from Eq. (6.142)

* see [6.4 (1985)] for exact solutions

For the *point availability*, Fig. 6.14b yields

$$\mathrm{PA}_{S0}(t) = e^{-3\lambda t} + \int_0^t 3\lambda e^{-3\lambda x}\,\mathrm{PA}_{S1}(t-x)\,dx$$

$$\mathrm{PA}_{S1}(t) = e^{-2\lambda t}(1-\mathrm{G}(t)) + \int_0^t \mathrm{g}(x) e^{-2\lambda x}\, \mathrm{PA}_{S0}(t-x)dx$$
$$+\int_0^t \mathrm{g}(x)(1-e^{-2\lambda x})\mathrm{PA}_{S1}(t-x)dx \tag{6.153}$$

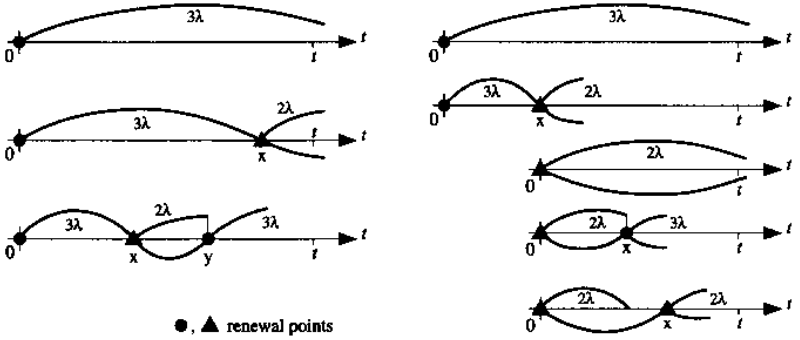
from which,

$$\mathrm{P\tilde{A}}_{S0}(s) = \frac{(s+2\lambda)[1+\tilde{\mathrm{g}}(s+2\lambda)-\tilde{\mathrm{g}}(s)]+3\lambda(1-\tilde{\mathrm{g}}(s+2\lambda))}{s(s+2\lambda)[1+\tilde{\mathrm{g}}(s+2\lambda)-\tilde{\mathrm{g}}(s)]+3\lambda(s+2\lambda)(1-\tilde{\mathrm{g}}(s))}. \tag{6.154}$$

The *asymptotic* and *steady-state* value of the point and average availability is then given by (see also Eq. (6.54))

$$PA_S = AA_S = \frac{3-\tilde{\mathrm{g}}(2\lambda)}{2\,\tilde{\mathrm{g}}(2\lambda)+6\lambda\, MTTR}, \tag{6.155}$$

with *MTTR* as per Eq. (6.110). For the asymptotic and steady-state value of the *interval reliability*, an approximate expression similar to that of Eq. (6.112) can be used in most applications. Generalization of the failure and repair rates leads to *nonregenerative stochastic processes.*



a) Calculation of $\mathrm{R}_{S0}(t)$ b) Calculation of $\mathrm{PA}_{S0}(t)$

Figure 6.14 Time schedule of a repairable 2-out-of-3 active redundancy (constant failure rate, arbitrary repair rate, one repair crew, no further failures at system down, repair times are shown greatly exaggerated)

6.6 Simple Series/Parallel Structures

A series/parallel structure is an arbitrary combination of series and parallel models, see Table 2.1 for some examples. Such a structure is generally investigated on a case-by-case basis using the methods of Sections 6.3 – 6.5. If the time behavior can be described by a Markov or semi-Markov process, Table 6.2 can be used to establish equations for the reliability function, point availability, and interval reliability.

As a first example, let us consider a repairable *1-out-of-2 active redundancy* with elements $E_1 = E_2 = E$ in series with a *switching element* E_v. The failure rates λ and λ_v as well as the repair rates μ and μ_v are constant. The system has only one repair crew, *repair priority* is for E_v (a repair on E_1 or E_2 is stopped as soon as a failure of E_v occurs, see Example 6.11 for the case of no priority), and *no further failures at system down* (i.e. failures occurring during a repair at system level are neglected, as per assumption (6.2)). Figure 6.15 gives the reliability block diagram and the diagram of transition probabilities in $(t, t+\delta t]$. The *reliability function* can be computed using Table 6.2, or directly by considering that for a series structure the reliability at system level is still the product of the reliabilities of the elements

$$\mathrm{R}_{S0}(t) = \mathrm{R}_{S0_{\text{1-out-of-2}}}(t)e^{-\lambda_v t}. \tag{6.156}$$

Because of the term $e^{-\lambda_v t}$, the Laplace transform of $\mathrm{R}_{S0}(t)$ follows directly from the Laplace transform of the reliability function for the 1-out-of-2 parallel redundancy $\mathrm{R}_{S0_{\text{1-out-of-2}}}$, by replacing s with $s+\lambda_v$ (Table A9.7)

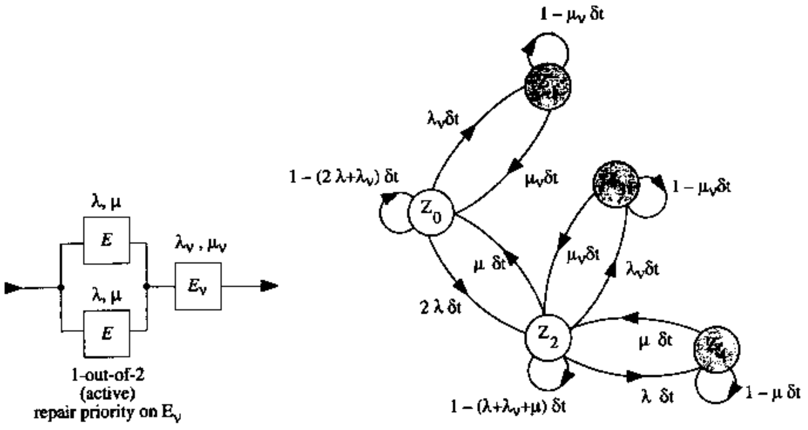
$$\tilde{\mathrm{R}}_{S0}(s) = \frac{s+3\lambda+\lambda_v+\mu}{(s+2\lambda+\lambda_v)(s+\lambda+\lambda_v)+(s+\lambda_v)\mu}. \tag{6.157}$$

The *mean time to failure* $MTTF_{S0}$ follows then from $MTTF_{S0} = \tilde{\mathrm{R}}_{S0}(0)$

$$MTTF_{S0} = \frac{3\lambda+\lambda_v+\mu}{(2\lambda+\lambda_v)(\lambda+\lambda_v)+\mu\lambda_v} = \frac{1}{\lambda_v+2\lambda^2/(3\lambda+\lambda_v+\mu)} \le \frac{1}{\lambda_v}. \tag{6.158}$$

The last part of Eq. (6.158) clearly shows the effect of the series element E_v. The *asymptotic* and *steady-state* value of the *point availability* and *average availability* $PA_S = AA_S$ is given by the solution of following system of algebraic equations obtained from Fig. 6.15 and Table 6.2

$$P_0 = \frac{(\mu_v P_1 + \mu P_2)}{2\lambda+\lambda_v}, \quad P_1 = \frac{\lambda_v}{\mu_v} P_0, \quad P_2 = \frac{1}{\lambda+\lambda_v+\mu}(\mu_v P_3 + \mu P_4 + 2\lambda P_0),$$
$$P_3 = \frac{\lambda_v}{\mu_v} P_2, \qquad P_4 = \frac{\lambda}{\mu} P_2, \qquad P_0 + P_1 + P_2 + P_3 + P_4 = 1. \tag{6.159}$$



$\rho_{01} = \rho_{23} = \lambda_v$; $\rho_{02} = 2\lambda$; $\rho_{10} = \rho_{32} = \mu_v$; $\rho_{20} = \rho_{42} = \mu$; $\rho_{24} =$
(for $R_S(t)$ set $\rho_{10} = \rho_{32} = \rho_{42} = 0$)

Figure 6.15 Reliability block diagram and diagram of transition probabilities in $(t, t+\delta t]$ for a repairable 1-out-of-2 active redundancy with a switching element (constant failure rates (λ, λ_v) and constant repair rates (μ, μ_v), one repair crew, repair priority for E_v, no further failures at system down, arbitrary t, $\delta t \downarrow 0$, Markov process, Z_0 and Z_2 are up states)

One (arbitrarily chosen, because of the implicitly assumed ergodicity) of the first five equations (6.159) must be dropped and replaced by the sixth one, $\sum P_i = 1$, because of linear dependence. The solution yields P_0 through P_4, from which

$$PA_S = AA_S = P_0 + P_2 = \frac{\mu^2\mu_v + 2\lambda\mu\mu_v}{\mu^2\mu_v + 2\lambda\mu\mu_v + 2\lambda(\lambda\mu_v + \lambda_v\mu) + \mu^2\lambda_v}$$

$$= \frac{1}{1 + \frac{\lambda_v}{\mu_v} + \frac{2(\lambda/\mu)^2}{1+2\lambda/\mu}} \approx 1 - \frac{\lambda_v}{\mu_v} - \frac{2(\lambda/\mu)^2}{1+2\lambda/\mu} \le 1 - \frac{\lambda_v}{\mu_v}. \tag{6.160}$$

As for the mean time to failure (Eq. (6.158)), the last part of Eq. (6.160) shows the effect of the series element E_v. For the asymptotic and steady-state value of the *interval reliability* one obtains (Table 6.2)

$$IR_S(\theta) = P_0 R_{S0}(\theta) + P_2 R_{S2}(\theta) \approx P_0 R_{S0}(\theta) \approx R_{S0}(\theta). \tag{6.161}$$

Example 6.11

Give the reliability function and the asymptotic and steady-state value of the point and average availability for a 1-out-of-2 active redundancy in series with a switching element, as in Fig. 6.15, but *without repair priority* on the switching element.

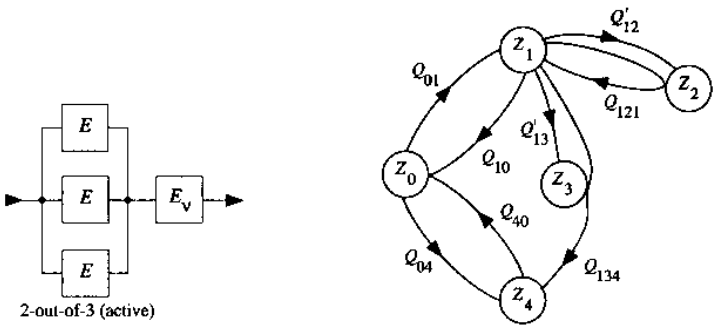


Figure 6.16 Reliability block diagram and state transition diagram for a 2-out-of-3 majority redundancy (constant failure rates λ for E and λ_v for E_v, repair time distributed according to $G(t)$ with density $g(t)$, one repair crew, no repair priority, no further failures at system down; Z_0, Z_1, and Z_4 constitute an embedded semi-Markov process, Z_0 and Z_1 are up states)

Solution

The diagram of transition probabilities in $(t, t+\delta t]$ of Fig. 6.15 can be used by changing the transition from state Z_3 to state Z_2 to one from Z_3 to Z_1 and μ_v in μ. The reliability function is still given by Eq. (6.156), then states Z_1, Z_3, and Z_4 are absorbing states for reliability calculations. For the asymptotic and steady-state value of the point and average availability $PA_S = AA_S$, Eq. (6.159) is modified to

$$P_0 = \frac{(\mu_v P_1 + \mu P_2)}{2\lambda + \lambda_v}, \quad P_1 = \frac{\lambda_v}{\mu_v} P_0 + \frac{\mu}{\mu_v} P_3, \quad P_2 = \frac{1}{\lambda + \lambda_v + \mu}(\mu P_4 + 2\lambda P_0),$$

$$P_3 = \frac{\lambda_v}{\mu} P_2, \qquad P_4 = \frac{\lambda}{\mu} P_2, \qquad P_0 + P_1 + p_2 + P_3 + P_4 = 1,$$

and the solution leads to

$$\mathrm{PA}_S = \mathrm{AA}_S = \frac{1}{1 + \frac{\lambda_v}{\mu_v} + \frac{2\lambda(\lambda+\lambda_v)/\mu^2}{1+(2\lambda+\lambda_v)/\mu}} \approx 1 - \frac{\lambda_v}{\mu_v} - \frac{2\lambda(\lambda+\lambda_v)/\mu^2}{1+(2\lambda+\lambda_v)/\mu} \le 1 - \frac{\lambda_v}{\mu_v}. \tag{6.162}$$

As a second example let us consider a 2-out-of-3 *majority redundancy* (2-out-of-3 active redundancy in series with a voter E_v). Assumptions (6.1) to (6.7) also hold here, in particular Assumption (6.2), i.e. *no further failures at system down*. The system has constant failure rates, λ for the three redundant elements and λ_v for the series element E_v, and repair time distributed according to $G(t)$ with $G(0) = 0$ and density $g(t)$. Figure 6.16 shows the corresponding reliability diagram and the *state transition diagram*. Z_0 and Z_1 are up states. Z_0, Z_1 and Z_4 are *regeneration states* and constitute a *semi-Markov process embedded* in the original process. This property will be used for the investigations. From Fig. 6.16 and Table 6.2 there follows for the *semi-Markov transition probabilities* $Q_{01}(x), Q_{10}(x), Q_{04}(x), Q_{40}(x)$, $Q_{121}(x)$, and $Q_{134}(x)$ the expressions (Fig. A7.10)

$$Q_{01}(x) = \Pr\{\tau_{01} \le x \cap \tau_{04} > \tau_{01}\} = \int_0^x 3\lambda e^{-3\lambda y} e^{-\lambda_\nu y} dy = \frac{3\lambda(1 - e^{-(3\lambda+\lambda_\nu)x})}{3\lambda + \lambda_\nu}$$

$$Q_{10}(x) = \Pr\{\tau_{10} \le x \cap (\tau_{12} > \tau_{10} \cap \tau_{13} > \tau_{10})\} = \int_0^x g(y) e^{-(2\lambda+\lambda_\nu)y} dy$$

$$= G(x) e^{-(2\lambda+\lambda_\nu)x} + \int_0^x (2\lambda + \lambda_\nu) e^{-(2\lambda+\lambda_\nu)y} G(y) dy$$

$$Q_{121}(x) = \Pr\{\tau_{121} \le x\}$$

$$= \int_0^x g(y) \int_0^y 2\lambda e^{-(2\lambda+\lambda_\nu)z} dz\, dy = \int_0^x g(y) \frac{2\lambda}{2\lambda + \lambda_\nu}(1 - e^{-(2\lambda+\lambda_\nu)y}) dy$$

$$Q_{134}(x) = \Pr\{\tau_{134} \le x = \int_0^x g(y) \int_0^y \lambda_\nu e^{-(2\lambda+\lambda_\nu)z} dz\, dy = \frac{\lambda_\nu}{2\lambda} Q_{121}(x)$$

$$Q_{04}(x) = \Pr\{\tau_{04} \le x \cap \tau_{01} > \tau_{04}\} = \int_0^x \lambda_\nu e^{-(\lambda_\nu+3\lambda)y} dy = \frac{\lambda_\nu}{3\lambda} Q_{01}(x)$$

$$Q_{40}(x) = \Pr\{\tau_{40} \le x\} = G(x). \tag{6.163}$$

$Q_{121}(x)$ is used to compute the *point availability*. It accounts for the process returning from state Z_2 to state Z_1 and that Z_2 is *not a regeneration state* (probability for the transition $Z_1 \rightarrow Z_2 \rightarrow Z_1$, see also Fig. A7.10a), similarly for $Q_{134}(x)$. $Q'_{12}(x)$ and $Q'_{13}(x)$ as given in Fig 6.16 are *not* semi-Markov transition probabilities (Z_2 and Z_3 are *not regeneration states*), however they are useful in this example for the computation of the *reliability function*

$$Q'_{12}(x) = \Pr\{\tau_{12} \le x \cap (\tau_{13} > \tau_{12} \cap \tau_{10} > \tau_{12})\} = \int_0^x 2\lambda e^{-2\lambda y} e^{-\lambda_\nu y}(1 - G(y)) dy$$

$$Q'_{13}(x) = \Pr\{\tau_{13} \le x \cap (\tau_{12} > \tau_{13} \cap \tau_{10} > \tau_{13})\} = \frac{\lambda_\nu}{2\lambda} Q'_{12}(x).$$

Considering that Z_0 and Z_1 are up states and at the same time *regeneration states*, as well as the above expressions, the following system of integral equations can be established for the *reliability functions* $R_{S0}(t)$ and $R_{S1}(t)$

$$R_{S0}(t) = e^{-(3\lambda+\lambda_\nu)t} + \int_0^t 3\lambda e^{-(3\lambda+\lambda_\nu)x} R_{S1}(t - x) dx$$

$$R_{S1}(t) = e^{-(2\lambda+\lambda_\nu)t}(1 - G(t)) + \int_0^t g(x) e^{-(2\lambda+\lambda_\nu)x} R_{S0}(t - x) dx. \tag{6.164}$$

The solution of Eq. (6.163) yields

$$\tilde{R}_{S0}(s) = \frac{s + 5\lambda + \lambda_\nu - 3\lambda\tilde{g}(s + 2\lambda + \lambda_\nu)}{(s + 2\lambda + \lambda_\nu)[s + \lambda_\nu + 3\lambda(1 - \tilde{g}(s + 2\lambda + \lambda_\nu))]} \tag{6.165}$$

and

$$MTTF_{S0} = \frac{5\lambda + \lambda_\nu - 3\lambda\tilde{g}(2\lambda + \lambda_\nu)}{(2\lambda + \lambda_\nu)[\lambda_\nu + 3\lambda(1 - \tilde{g}(2\lambda + \lambda_\nu))]}. \tag{6.166}$$

$\tilde{R}_{S0}(s)$ and thus also $MTTF_{S0}$ could be directly obtained as for Eqs. (6.156) and (6.157) by setting $s = s + \lambda_\nu$ in Eq (6.151).

For the *point availability*, calculation of the transition probabilities $P_{ij}(t)$ with Table 6.2 and Eq. (6.163) leads to

$$P_{00}(t) = e^{-(3\lambda+\lambda_\nu)t} + \int_0^t 3\lambda e^{-(3\lambda+\lambda_\nu)x} P_{10}(t - x) dx + \int_0^t \lambda_\nu e^{-(3\lambda+\lambda_\nu)x} P_{40}(t - x) dx$$

$$P_{10}(t) = \int_0^t g(x) e^{-(2\lambda+\lambda_\nu)x} P_{00}(t - x) dx$$

$$+ \int_0^t \frac{2\lambda}{2\lambda + \lambda_\nu}(1 - e^{-(2\lambda+\lambda_\nu)x}) g(x) P_{10}(t - x) dx$$

$$+ \int_0^t \frac{\lambda_\nu}{2\lambda + \lambda_\nu}(1 - e^{-(2\lambda+\lambda_\nu)x}) g(x) P_{40}(t - x) dx$$

$$P_{40}(t) = \int_0^t g(x) P_{00}(t - x) dx, \tag{6.167}$$

and

$$P_{01}(t) = \int_0^t 3\lambda e^{-(3\lambda+\lambda_\nu)x} P_{11}(t - x) dx + \int_0^t \lambda_\nu e^{-(3\lambda+\lambda_\nu)x} P_{41}(t - x) dx$$

$$P_{11}(t) = e^{-(2\lambda+\lambda_\nu)t}(1 - G(t)) + \int_0^t g(x) e^{-(2\lambda+\lambda_\nu)x} P_{01}(t - x) dx$$

$$+ \int_0^t \frac{1}{2\lambda + \lambda_\nu}(1 - e^{-(2\lambda+\lambda_\nu)x}) g(x)[2\lambda P_{11}(t - x) + \lambda_\nu P_{41}(t - x)] dx \lambda_\nu$$

$$P_{41}(t) = \int_0^t g(x) P_{01}(t - x) dx. \tag{6.168}$$

From Eqs. (6.167) and (6.168) it follows that the *point availability* $PA_{S0}(t) = P_{00}(t) + P_{01}(t)$ and from this (using the Laplace transform), the *steady-state value*

$$PA_S = AA_S = \frac{2\lambda + \lambda_\nu + \lambda(1 - \tilde{g}(2\lambda + \lambda_\nu))}{(2\lambda + \lambda_\nu)[1 + (3\lambda + \lambda_\nu) MTTR] + \lambda(\lambda_\nu MTTR - 2)(1 - \tilde{g}(2\lambda + \lambda_\nu))} \tag{6.169}$$

with *MTTR* as per Eq. (6.110).

For the asymptotic and steady-state value of the *interval reliability*, the following *approximate expression* can be used for practical applications (Eq. (6.111))

$$\begin{aligned} \mathrm{IR}_S(\theta) &\approx P_0 \, \mathrm{R}_{S0}(\theta) \\ &= \frac{[(2\lambda + \lambda_\nu) - 2\lambda(1 - \tilde{g}(2\lambda + \lambda_\nu))] \mathrm{R}_{S0}(\theta)}{(2\lambda + \lambda_\nu)[1 + (3\lambda + \lambda_\nu) MTTR] + \lambda(\lambda_\nu MTTR - 2)(1 - \tilde{g}(2\lambda + \lambda_\nu))}. \end{aligned} \tag{6.170}$$

In Eq. (6.170), $P_0 = \lim_{t \to \infty} \mathrm{P}_{00}(t)$ with $\mathrm{P}_{00}(t)$ from Eq. (6.167). For $\tilde{g}(2\lambda + \lambda_\nu) \approx 1$ $\mathrm{IR}_S(\theta) \approx \mathrm{R}_{S0}(\theta)$ can also be used.

Example 6.12

(i) Compute from Eqs. (6.166) and (6.169) the mean time to failure $MTTF_{S0}$ and the asymptotic and steady-state value of the point and average availability $PA_S = AA_S$ for the case of a constant repair rate μ. (ii) Compare for the case of constant repair rate the true value of the interval reliability $\mathrm{IR}_S(\theta)$ with the approximate expression given by Eq. (6.170).

Solution

(i) With $G(t) = 1 - e^{-\mu t}$ it follows that $\tilde{g}(2\lambda + \lambda_\nu) = \mu / (2\lambda + \lambda_\nu + \mu)$ and thus from Eq. (6.166)

$$MTTF_{S0} = \frac{5\lambda + \lambda_\nu + \mu}{(3\lambda + \lambda_\nu)(2\lambda + \lambda_\nu) + \mu\lambda_\nu} = \frac{1}{\lambda_\nu + 6\lambda^2 /(5\lambda + \lambda_\nu + \mu)} \approx \frac{1}{\lambda_\nu + 6\lambda^2 / \mu}, \tag{6.171}$$

and from Eq. (6.169)

$$\begin{aligned} PA_S = AA_S &= \frac{\mu(3\lambda + \lambda_\nu + \mu)}{(3\lambda + \lambda_\nu + \mu)(\lambda_\nu + \mu) + 3\lambda(2\lambda + \lambda_\nu)} = \frac{1}{1 + \frac{\lambda_\nu}{\mu} + \frac{3\lambda(2\lambda + \lambda_\nu)}{\mu(\mu + 3\lambda + \lambda_\nu)}} \\ &\approx 1 - \frac{\lambda_\nu}{\mu} - \frac{3\lambda(2\lambda + \lambda_\nu)}{\mu^2}. \end{aligned} \tag{6.172}$$

(ii) With $\mathrm{P}_{00}(t)$ and $\mathrm{P}_{01}(t)$ from Eqs. (6.167) and (6.168) it follows for the asymptotic and steady-state value of the internal reliability (Table 6.2) that

$$\mathrm{IR}_S(\theta) = \frac{\mu(\lambda_\nu + \mu)\mathrm{R}_{S0}(\theta) + 3\lambda\mu \mathrm{R}_{S1}(\theta)}{(3\lambda + \lambda_\nu + \mu)(\lambda_\nu + \mu) + 3\lambda(2\lambda + \lambda_\nu)}. \tag{6.173}$$

The *approximate expression* according to Eq. (6.170) yields.

$$\mathrm{IR}_S(\theta) \approx \frac{\mu(\lambda_\nu + \mu)\mathrm{R}_{S0}(\theta)}{(3\lambda + \lambda_\nu + \mu)(\lambda_\nu + \mu) + 3\lambda(2\lambda + \lambda_\nu)},$$

which, for λ, $\lambda_\nu << \mu$, gives values very near to those given by Eq. (6.173). The approximation for $\mathrm{IR}_S(\theta)$ takes into account that $\mathrm{R}_{S1}(\theta) \le \mathrm{R}_{S0}(\theta)$ and assumes $\mu >> 3\lambda$.

To give a better feeling for the mutual influence of the different parameters involved, Figs. 6.17 and 6.18 compare the mean time to failure $MTTF_{S0}$ and the asymptotic and steady-state unavailability $1 - PA_S$ of some basic *series/parallel structures*. The equations used are obtained using Table 6.10 which summarizes the results of Sections 6.2 to 6.6 (approximate expressions are taken to simplify calculations). A comparison with Figs. 2.8 and 2.9 (nonrepairable case) confirms that the most important gain is obtained by the first step (structure b), and shows that the influence of series elements is much greater in the repairable than in the nonrepairable case. Referring to the structures a), b), and c) of Figs. 6.17 and 6.18 the following *design guideline* can be established:

The failure rate of the series element in a repairable 1-out-of-2 active redundancy should not be greater than 1% (0.1% for $\mu / \lambda_1 > 500$*) of the failure rate of the redundant elements, i.e. with respect to Fig. 6.17*

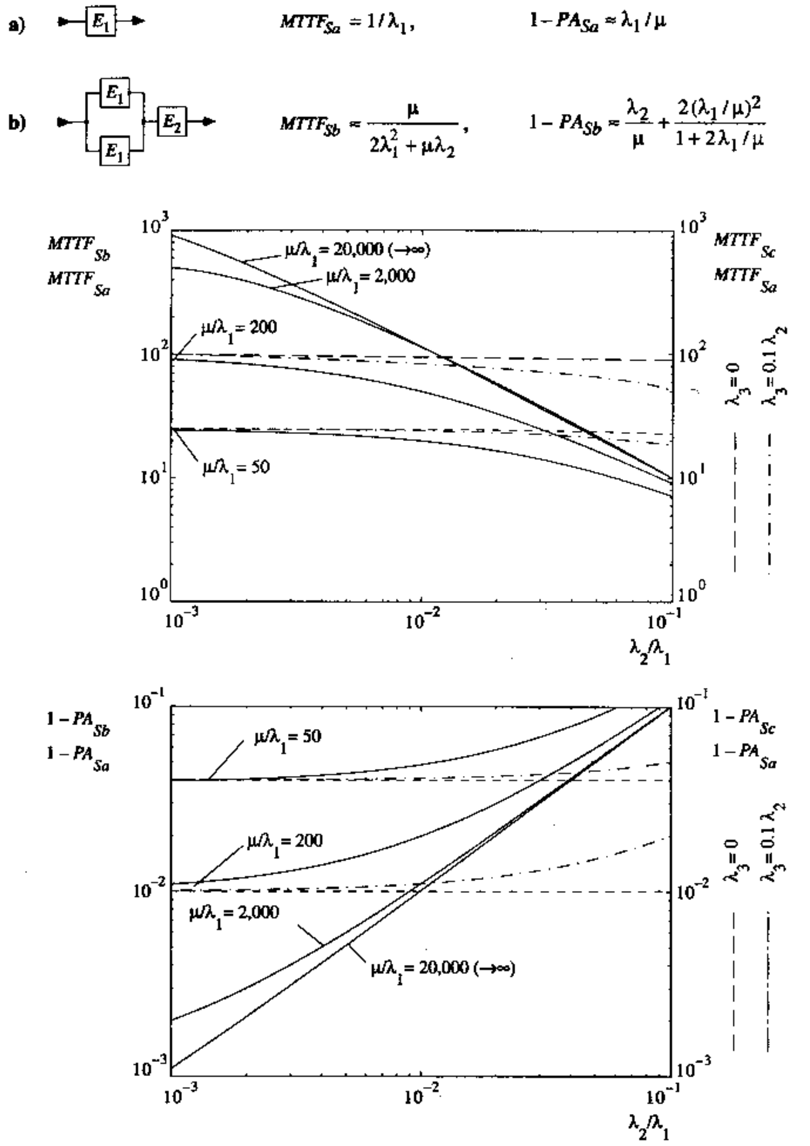
$$\lambda_2 \le 0.01\lambda_1 \text{ in general, and } \lambda_2 \le 0.002\lambda_1 \text{ for } \mu/\lambda_1 > 500. \tag{6.174}$$

6.7 Approximate Expressions for Large Series/Parallel Structures

6.7.1 Introduction

Reliability and availability computations of *large series/parallel structures* rapidly becomes time consuming, even if a *constant failure rate* λ_i and a *constant repair rate* μ_i are assumed for each element E_i of the reliability block diagram, and only the mean time to failure $MTTF_{S0}$ or the steady-state availability $PA_S = AA_S$ is required (solution of algebraic equations). This is because of the large number of states involved, comprised between 2^n and $1 + \sum_{i=1}^{n} \prod_{k=n-i+1}^{n} k = n! \sum_{i=0}^{n} 1/i! \approx e \cdot n!$ for a reliability block diagram with n elements, often close to $e \cdot n!$. In such cases the use of *approximate expressions* is necessary. Assuming for each element E_i that $\lambda_i << \mu_i$ holds, approximate expressions for the system reliability and availability can be obtained by one of the following methods:

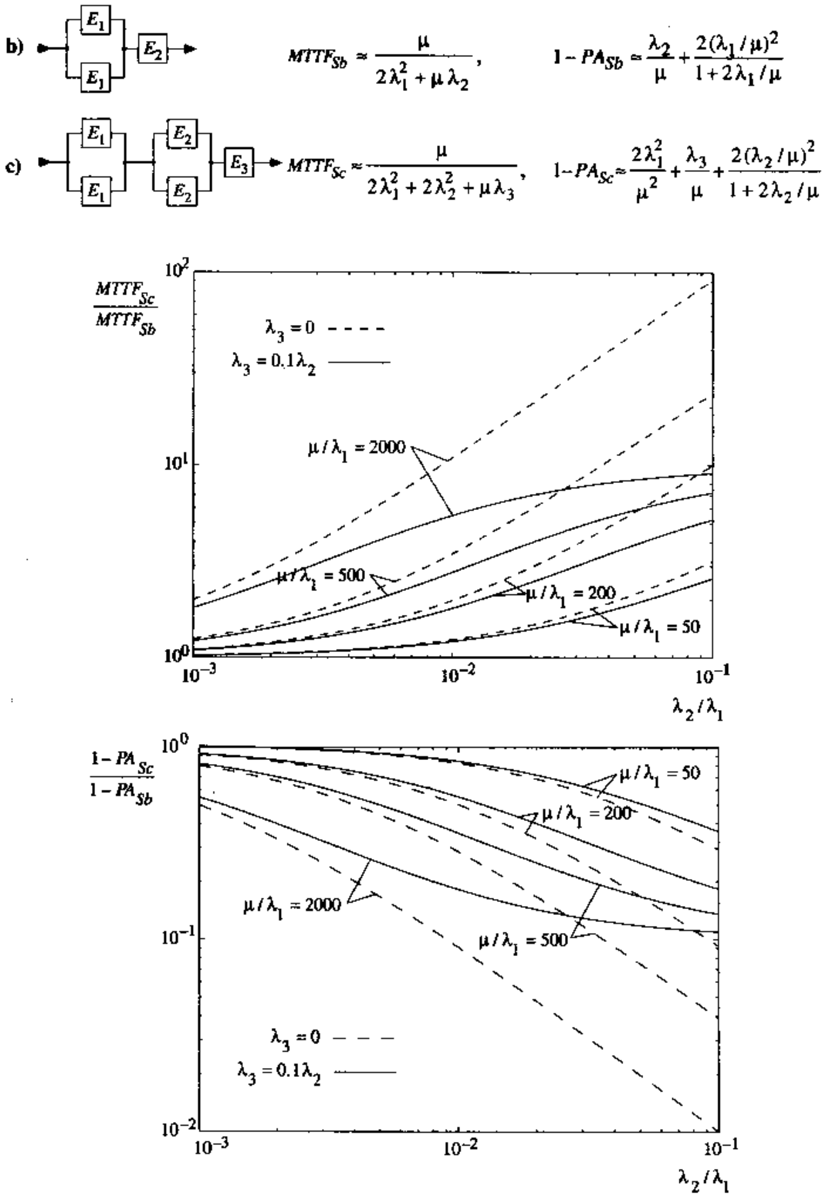
1. *Independent elements in operation and repair*: If each element of the reliability block diagram operates independently from every other element (active redundancy, independent elements, one repair crew for each element), series/parallel structures can be reduced to one-item structures, which are themselves successively integrated into further series/parallel structures up to the



a) $MTTF_{Sa} = 1/\lambda_1, \qquad 1 - PA_{Sa} \approx \lambda_1/\mu$

b) $MTTF_{Sb} \approx \dfrac{\mu}{2\lambda_1^2 + \mu\lambda_2}, \qquad 1 - PA_{Sb} \approx \dfrac{\lambda_2}{\mu} + \dfrac{2(\lambda_1/\mu)^2}{1 + 2\lambda_1/\mu}$

Figure 6.17 Comparison between a one-item structure and a 1-out-of-2 active redundancy with a series element (repairable, one repair crew with repair priority for E_2, no further failure at system down, constant failure rates λ_1 and λ_2, constant repair rate μ, λ_1 remains the same in both structures, equations according to Table 6.10; also given (right-hand side) are $MTTF_{Sc}/MTTF_{Sa}$ and $(1-PA_{Sc})/(1-PA_{Sa})$ with $MTTF_{Sc}$ and $1-PA_{Sc}$ from Fig. 6.18 to show the smaller dependency on λ_2/λ_1; see Fig. 2.8 for the nonrepairable case)



b) $MTTF_{Sb} \approx \dfrac{\mu}{2\lambda_1^2 + \mu\lambda_2}, \qquad 1 - PA_{Sb} \approx \dfrac{\lambda_2}{\mu} + \dfrac{2(\lambda_1/\mu)^2}{1 + 2\lambda_1/\mu}$

c) $MTTF_{Sc} \approx \dfrac{\mu}{2\lambda_1^2 + 2\lambda_2^2 + \mu\lambda_3}, \qquad 1 - PA_{Sc} \approx \dfrac{2\lambda_1^2}{\mu^2} + \dfrac{\lambda_3}{\mu} + \dfrac{2(\lambda_2/\mu)^2}{1 + 2\lambda_2/\mu}$

Figure 6.18 Comparison between basic *series/parallel structures* (repairable, one repair crew with repair priority for E_3, active redundancy, no further failure at system down, constant failure rates λ_1 to λ_3, λ_1 and λ_2 remain the same in both structures, equations according to Table 6.10; see Fig. 2.9 for the nonrepairable case) system level. To each of the one-item structure obtained, the mean time to failure $MTTF_{S0}$ and the steady-state availability PA_S, calculated for the underlying series/parallel structure, are used to find an equivalent $MTTR_S$ from the relationship $PA_S = MTTF_S / (MTTF_S + MTTR_S)$, with $MTTF_S = MTTF_{S0}$. To simplify calculations and considering the remarks given above Eq. (6.93), *constant failure rate* $\lambda_S = 1/MTTF_{S0}$ and *constant repair rate* $\mu_S = 1/MTTR_S$ are assumed for each of the one-item structures obtained. Table 6.9 summarizes the basic series/parallel structures useful for practical applications, see Section 6.7.2 for an example.

2. *Macro-structures.* A macro-structure is a series, parallel, or simple series/parallel structure which is considered as a one-item structure for calculations at higher levels (integration into further macro structures). It satisfies Assumptions (6.1) to (6.7), in particular one repair crew for each macro-structure and no further failures during a repair at the macro-structure level. The procedure is simiar to that of point 1 above. Table 6.10 summarizes the basic macro-structures useful for practical applications, see Sections 6.2 to 6.6 for the results and Section 6.7.2 for an example.
3. *One repair crew and no further failures at system down*: Assumptions (6.3) and (6.2), valid for all models investigated in Sections 6.3 to 6.6, are true for many practical applications. No further failures at system down means that failures during a repair at system level are neglected. This assumption has *no influence on the reliability function* at system level and its *influence on the availability is limited*, because of $\lambda_i << \mu_i$.
4. *Cutting states*: Removing the states with more than k failures from the diagram of transition probabilities in $(t, t+\delta t]$, or from the state transition diagram, produces in general an important reduction of the number of states in the state diagram. The choice of k (often $k = 2$) is based on the required precision. An upper bound of the error for the asymptotic and steady-state value of the point and average availability $PA_S = AA_S$ (based on the mapping of states with k failures at the system level in the state Z_k of a birth and death process and using the property $P_k \geq \sum_{i=k+1}^{n} P_i$, valid for $2(\lambda_1 + \ldots + \lambda_n) < \min\{\mu_1, \ldots, \mu_n\}$, see Eq. (A7.143)) has been developed in [2.51 (1992)].
5. *Clustering of states*: Grouping of elements in the reliability block diagram (series elements for example) or of states in the diagram of transition probabilities in $(t, t+\delta t]$ produces in general an important reduction of the number of states in the state diagram.

A combination of the above methods is possible. Considering that the steady-state probability for states with more than one failure at the system level decreases rapidly as the number of failures increases ($P_{i+1}/P_i \sim \lambda/\mu$ in general), the above methods yield good *approximate expressions* for $MTTF_{S0}$ and PA_S in practical applications.

However, referring to the ***unavailability*** $1 - PA_S$, **method 1 above can deliver lower** values, for instance a factor 2 with an order of magnitude $(\lambda/\mu)^2$ for a 1-out-of-2 active redundancy (Tables 6.9 and 6.10). An analytical comparison of the above methods is in general difficult. Numerical calculations show a close convergence of the results given by the different methods, as illustrated in Section 6.7.2 for a practical example with a very large λ/μ ratio (up to 0.05).

6.7.2 Application to a Practical Example

To illustrate how methods 1 to 3 of Section 6.7.1 work, let us consider the system with a reliability block diagram as in Fig. 6.19, and assume active redundancy, constant failure rates λ_1 to λ_3 and constant repair rates μ_1 to μ_3, repair priority in the sequence E_1, E_3 and E_2, system new at $t = 0$, and (as for Sections 6.3 to 6.6 and Method 3 of Section 6.7.1) only one repair crew and no further failures at system down. Figure 6.20 gives the corresponding diagram of transition probabilities in $(t, t+\delta t]$. Besides some series elements, the reliability block diagram of Fig. 6.19 describes an uninterruptible power supply as used, for example, to buffer power network failures in computer systems (E_1 represents the power network). Although limited to 4 elements, the process describing the system of Fig. 6.19 would contain more than 50 states if the assumption of no further failures at system down were dropped. Before establishing the results for the model described by Fig. 6.20 using Table 6.2, let us develop *approximate expressions* for the failure rate $\lambda_S = 1/MTTF_{S0}$ and the asymptotic and steady-state value of the point and average availability $PA_S = AA_S$ as per Methods 1 and 2 of Section 6.7.1.

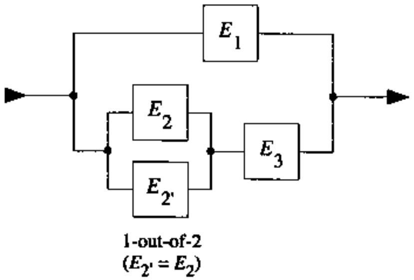
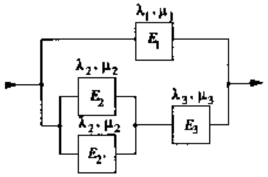
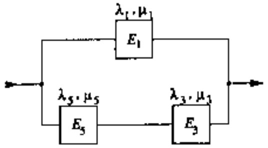


Figure 6.19 Basic part of the reliability block diagram of an uninterruptible power supply

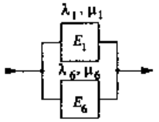
Method 1 of Section 6.7.1 yields, using Table 6.9,



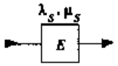
System



$$\lambda_5 \approx \frac{2\lambda_2^2}{\mu_2}, \qquad \mu_5 = 2\mu_2, \tag{6.175}$$



$$\lambda_6 \approx \lambda_3 + \lambda_5, \qquad \mu_6 \approx \frac{\lambda_5 + \lambda_3}{\lambda_5 / \mu_5 + \lambda_3 / \mu_3}, \tag{6.176}$$



$$\lambda_S \approx \frac{\lambda_1 \lambda_6 (\mu_1 + \mu_6)}{\mu_1 \mu_6}, \qquad \mu_S \approx \mu_1 + \mu_6. \tag{6.177}$$

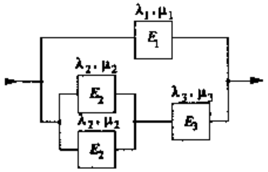
From Eqs. (6.175) – (6.177) it follows that

$$\lambda_S \approx \lambda_1 \left[\frac{\lambda_3}{\mu_1} + \frac{2\lambda_2^2}{\mu_1 \mu_2} + \frac{\lambda_3}{\mu_3} + (\frac{\lambda_2}{\mu_2})^2\right] \tag{6.178}$$

and

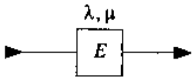
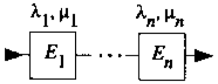
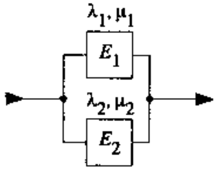
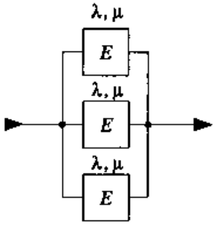
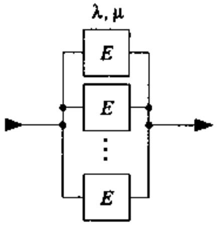
$$PA_S \approx 1 - \frac{\lambda_S}{\mu_S} \approx 1 - \frac{\lambda_1}{\mu_1}\left[\frac{\lambda_3}{\mu_3} + (\frac{\lambda_2}{\mu_2})^2\right]. \tag{6.179}$$

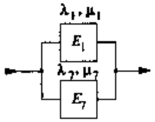
Method 2 of Section 6.7.1 yields, using Table 6.10,



System

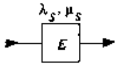
Table 6.9 Basic structures for the investigation of ***large series/parallel systems*** by assuming ***independend elements*** (each element operates and is repaired independently of every other element), constant failure rate (λ) and constant repair rate (μ), active redundancy, one repair crew for each element (see Eqs. (6.48), (2.48) and (6.60), (2.48) and (6.99), (2.48) and (6.171) with $\lambda_v = 0$, (6.148) and (6.149)); $\lambda_S = 1/MTTF_{S0}$ and $\mu_S = 1/MTTR_S$ are used to simplify the notation; approximations valid for $\lambda_i << \mu_i$

	$\lambda_S = \lambda, \quad \mu_S = \mu, \quad PA_S = \frac{1}{1 + \lambda_S / \mu_S} \approx 1 - \frac{\lambda_S}{\mu_S}$ $\Rightarrow \mu_S = \frac{\lambda_S PA_S}{1 - PA_S} \approx \frac{\lambda_S}{1 - PA_S}$
	$PA_S = PA_1 \dots PA_n \approx 1 - (\frac{\lambda_1}{\mu_1} + \dots + \frac{\lambda_n}{\mu_n})$ $\lambda_S = \lambda_1 + \dots + \lambda_n \quad \Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx \frac{\lambda_1 + \dots + \lambda_n}{\lambda_1/\mu_1 + \dots + \lambda_n/\mu_n}$
 1-out-of-2 (active)	$PA_S = PA_1 + PA_2 - PA_1 PA_2 \approx 1 - \frac{\lambda_1 \lambda_2}{\mu_1 \mu_2}$ $MTTF_{S0} \approx \frac{1}{\lambda_S} \approx \frac{\mu_1 \mu_2}{\lambda_1 \lambda_2 (\mu_1 + \mu_2)} \quad \rightarrow \lambda_S \approx \frac{\lambda_1 \lambda_2 (\mu_1 + \mu_2)}{\mu_1 \mu_2}$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} = \mu_1 + \mu_2$
 2-out-of-3 (active)	$PA_S = 3PA^2 - 2PA^3 \approx 1 - \frac{3(\lambda/\mu)^2}{1 + 3\lambda/\mu} \approx 1 - 3(\frac{\lambda}{\mu})^2$ $MTTF_{S0} \approx \frac{1}{\lambda_S} = \frac{5\lambda + \mu}{6\lambda^2} \approx \frac{\mu}{6\lambda^2} \quad \rightarrow \lambda_S \approx \frac{6\lambda^2}{\mu}$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx 2\mu$
 *k*-out-of-*n* (active)	$PA_S \approx 1 - \frac{k}{n-k+1} \binom{n}{k} (\frac{\lambda}{\mu})^{n-k+1}$ $MTTF_{S0} \approx \frac{1}{\lambda_S} \approx \frac{1}{k\lambda\binom{n}{k}} (\frac{\mu}{\lambda})^{n-k} \quad \rightarrow \lambda_S \approx k\lambda \binom{n}{k} (\frac{\lambda}{\mu})^{n-k}$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx (n-k+1)\mu$



$$\lambda_7 \approx \lambda_3 + 2\lambda_2^2 / \mu_2,$$

$$\mu_7 \approx \frac{\mu_3 (2\lambda_2^2 + \mu_2 \lambda_3)(1 + 2\lambda_2 / \mu_2)}{\mu_2 \lambda_3 + 2\lambda_2 \lambda_3 + 2\lambda_2^2 \mu_3 / \mu_2}, \tag{6.180}$$



$$\lambda_S \approx \frac{\lambda_1 \lambda_7 (\mu_1 + \mu_7)}{\mu_1 \mu_7}, \quad \mu_S \approx \mu_1 \mu_7 \frac{\mu_1 + \mu_7}{\mu_1^2 + \mu_7^2}. \tag{6.181}$$

From Eqs. (6.180) and (6.181) it follows that

$$\lambda_S \approx \lambda_1 \Big(\frac{2\lambda_2^2 + \mu_2 \lambda_3}{\mu_1 \mu_2} + \frac{\mu_2 \lambda_3 + 2\lambda_2 \lambda_3 + 2\mu_3 \lambda_2^2 / \mu_2}{\mu_2 \mu_3 (1 + 2\lambda_2 / \mu_2)}\Big) \tag{6.182}$$

and

$$PA_S \approx 1 - \frac{\lambda_S}{\mu_S} \approx 1 - \frac{2\lambda_2^2 + \mu_2 \lambda_3}{\mu_2} \Big(\frac{\lambda_1}{\mu_1^2} + \frac{\lambda_1 (\mu_2 \lambda_3 + 2\lambda_2 \lambda_3 + 2\mu_3 \lambda_2^2 / \mu_2)^2}{(2\lambda_2^2 + \mu_2 \lambda_3)^2 (1 + 2\lambda_2 / \mu_2)^2 \mu_3^2}\Big)$$

$$= 1 - \frac{\lambda_1}{\mu_1} \Big(\frac{\lambda_3}{\mu_3} + \frac{2\lambda_2^2}{\mu_2 \mu_3}\Big) \Big(\frac{\mu_3}{\mu_1} + \frac{(\mu_2 \lambda_3 + 2\lambda_2 \lambda_3 + 2\mu_3 \lambda_2^2 / \mu_2)^2 \mu_1 / \mu_3}{(2\lambda_2^2 + \mu_2 \lambda_3)^2 (1 + 2\lambda_2 / \mu_2)^2}\Big). \tag{6.183}$$

Method 3 of Section 6.7.1 yields, using Table 6.2 and Fig. 6.20, the following system of algebraic equations for the *mean time to failure* ($M_i = MTTF_{Si}$)

$$\begin{aligned}
&\rho_0 M_0 = 1 + \lambda_1 M_1 + 2\lambda_2 M_2 + \lambda_3 M_3, &&\rho_1 M_1 = 1 + \mu_1 M_0 + 2\lambda_2 M_7, \\
&\rho_2 M_2 = 1 + \mu_2 M_0 + \lambda_3 M_4 + \lambda_2 M_6 + \lambda_1 M_7, &&\rho_3 M_3 = 1 + \mu_3 M_0 + 2\lambda_2 M_4, \\
&\rho_4 M_4 = 1 + \mu_3 M_2 + \lambda_2 M_5, &&\rho_5 M_5 = 1 + \mu_3 M_6, \\
&\rho_6 M_6 = 1 + \mu_2 M_2 + \lambda_3 M_5, &&\rho_7 M_7 = 1 + \mu_1 M_2,
\end{aligned} \tag{6.184}$$

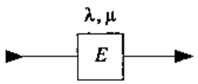
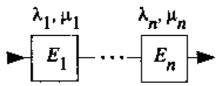
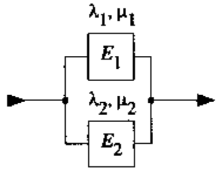
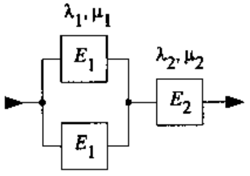
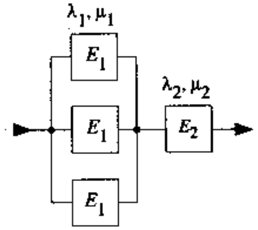
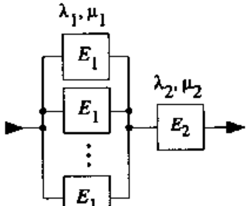
where

$$\begin{aligned}
&\rho_0 = \lambda_1 + 2\lambda_2 + \lambda_3, &&\rho_1 = \mu_1 + 2\lambda_2 + \lambda_3, &&\rho_2 = \mu_2 + \lambda_1 + \lambda_2 + \lambda_3, \\
&\rho_3 = \mu_3 + 2\lambda_2 + \lambda_1, &&\rho_4 = \mu_3 + \lambda_2 + \lambda_1, &&\rho_5 = \mu_3 + \lambda_1, \\
&\rho_6 = \mu_2 + \lambda_3 + \lambda_1, &&\rho_7 = \mu_1 + \lambda_3 + \lambda_2, &&\rho_8 = \mu_1, \\
&\rho_9 = \mu_1, &&\rho_{10} = \mu_1, &&\rho_{11} = \mu_1.
\end{aligned} \tag{6.185}$$

From Eqs. (6.184) and (6.185) it follows that

$$MTTF_{S0} = \frac{a_5 + a_6 (a_8 + a_9 a_{10}) + a_7 a_{10}}{1 - a_6 a_{12} - a_{11} (a_7 + a_6 a_9)}, \tag{6.186}$$

Table 6.10 Basic *macro-structures* for the investigation of large series/parallel systems by successive building of macro-structures from the bottom up to the system level, constant failure rate (λ) and constant repair rate (μ), active redundancy, one repair crew and no further failure at system down for every macro-structure (see Eqs. (6.48), (6.65) and (6.60), (6.103) and (6.99), (6.160) and (6.158), (6.172) and (6.171), (6.65) & (6.60) and Tab. 6.8); $\lambda_S = 1 / MTTF_{S0}$ and $\mu_S = 1 / MTTR_S$ are used to simplify the notation; approximations valid for $\lambda_i << \mu_i$

	$\lambda_S \equiv \lambda, \quad \mu_S \equiv \mu, \qquad PA_S = 1/(1 + \lambda_S / \mu_S) \approx 1 - \lambda_S / \mu_S$ $\Rightarrow \quad \mu_S = \frac{\lambda_S PA_S}{1 - PA_S} \approx \frac{\lambda_S}{1 - PA_S}$
	$PA_S \approx 1 - (\lambda_1 / \mu_1 + \ldots + \lambda_n / \mu_n); \qquad \lambda_S = \lambda_1 + \ldots + \lambda_n$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx \frac{\lambda_1 + \ldots + \lambda_n}{\lambda_1 / \mu_1 + \ldots + \lambda_n / \mu_n}$
 1-out-of-2 (active)	$PA_S \approx 1 - \frac{\lambda_1 \lambda_2}{\mu_1^2 \mu_2^2} (\mu_1^2 + \mu_2^2)$ $MTTF_{S0} \equiv \frac{1}{\lambda_S} \approx \frac{\mu_1 \mu_2}{\lambda_1 \lambda_2 (\mu_1 + \mu_2)}$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx \mu_1 \mu_2 \frac{\mu_1 + \mu_2}{\mu_1^2 + \mu_2^2}$
 1-out-of-2 (active) repair priority on E_2	$PA_S \approx 1 - \frac{\lambda_2}{\mu_2} - \frac{2 (\lambda_1 / \mu_1)^2}{1 + 2\lambda_1 / \mu_1}$ $MTTF_{S0} \equiv \frac{1}{\lambda_S} = \frac{1}{\lambda_2 + 2\lambda_1^2 / (\mu_1 + 3\lambda_1 + \lambda_2)} \approx \frac{1}{\lambda_2 + 2\lambda_1^2 / \mu_1}$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx \mu_2 \frac{\lambda_2 + 2\lambda_1^2 / \mu_1}{\lambda_2 + \frac{2\lambda_1^2 / \mu_1}{1 + 2\lambda_1 / \mu_1} \cdot \frac{\mu_2}{\mu_1}} \approx \mu_2 \quad$ for $\mu_2 \approx \mu_1$
 2-out-of-3 (active), rep. priority on E_2	$PA_S \approx 1 - \frac{\lambda_2}{\mu_2} - \frac{6 (\lambda_1 / \mu_1)^2}{1 + 3\lambda_1 / \mu_1}$ $MTTF_{S0} \equiv \frac{1}{\lambda_S} \approx \frac{1}{\lambda_2 + 6\lambda_1^2 / \mu_1}$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx \mu_2 \frac{\lambda_2 + 6\lambda_1^2 / \mu_1}{\lambda_2 + \frac{6\lambda_1^2 / \mu_1}{1 + 3\lambda_1 / \mu_1} \cdot \frac{\mu_2}{\mu_1}} \approx \mu_2 \quad$ for $\mu_2 \approx \mu_1$
 k-out-of-n (active)	$PA_S \approx 1 - \frac{\lambda_2}{\mu_2} - \frac{n!}{(k-1)!} \big(\frac{\lambda_1}{\mu_1}\big)^{n-k+1}$ $MTTF_{S0} \equiv \frac{1}{\lambda_S} \approx \frac{1}{\lambda_2 + \lambda_1 \frac{n!}{(k-1)!} \big(\frac{\lambda_1}{\mu_1}\big)^{n-k}}$ $\Rightarrow \quad \mu_S \approx \frac{\lambda_S}{1 - PA_S} \approx \mu_2 \quad$ for $\mu_2 \approx \mu_1$

with

$$a_1 = \frac{1}{\rho_4} + \frac{\lambda_2}{\rho_4 \rho_5}\left(1 + \mu_3 \frac{\lambda_3 + \rho_5}{\rho_5 \rho_6 - \lambda_3 \mu_3}\right), \qquad a_2 = \frac{\lambda_2 \mu_2 \mu_3}{\rho_4(\rho_5 \rho_6 - \lambda_3 \mu_3)} + \frac{\mu_3}{\rho_4},$$

$$a_3 = \frac{1}{\rho_3}(1 + 2\lambda_2 a_1), \qquad a_4 = \frac{2\lambda_2}{\rho_3} a_2, \qquad a_5 = \frac{1 + \lambda_3 a_3}{\rho_0 - \lambda_3 \mu_3 / \rho_3},$$

$$a_6 = \frac{\lambda_1}{\rho_0 - \lambda_3 \mu_3 / \rho_3}, \qquad a_7 = \frac{2\lambda_2 + \lambda_3 a_4}{\rho_0 - \lambda_3 \mu_3 / \rho_3}, \qquad a_8 = \frac{1 + 2\lambda_2 / \rho_7}{\rho_1},$$

$$a_9 = \frac{2\lambda_2 \mu_1}{\rho_1 \rho_7}, \qquad a_{10} = \frac{1 + \lambda_3 a_1 + \dfrac{\lambda_2 \lambda_3 + \lambda_2 \rho_5}{\rho_5 \rho_6 - \lambda_3 \mu_3} + \dfrac{\lambda_1}{\rho_7}}{\rho_2 - \lambda_3 a_2 - \dfrac{\lambda_2 \mu_2 \rho_5}{\rho_5 \rho_6 - \lambda_3 \mu_3} - \dfrac{\lambda_1 \mu_1}{\rho_7}},$$

$$a_{11} = \frac{\mu_2}{\rho_2 - \lambda_3 a_2 - \dfrac{\lambda_2 \mu_2 \rho_5}{\rho_5 \rho_6 - \lambda_3 \mu_3} - \dfrac{\lambda_1 \mu_1}{\rho_7}}, \qquad a_{12} = \frac{\mu_1}{\rho_1}. \tag{6.187}$$

Similarly, for the *asymptotic* and *steady-state* value of the point and average availability $PA_S = AA_S$ the following system of algebraic equations, can be obtained using Table 6.2 and Fig. 6.20

$$\begin{aligned}
&\rho_0 P_0 = \mu_1 P_1 + \mu_2 P_2 + \mu_3 P_3, && \rho_1 P_1 = \lambda_1 P_0,\\
&\rho_2 P_2 = 2\lambda_2 P_0 + \mu_3 P_4 + \mu_2 P_6 + \mu_1 P_7, && \rho_3 P_3 = \lambda_3 P_0 + \mu_1 P_{10},\\
&\rho_4 P_4 = \lambda_3 P_2 + 2\lambda_2 P_3 + \mu_1 P_8, && \rho_5 P_5 = \lambda_2 P_4 + \lambda_3 P_6 + \mu_1 P_9,\\
&\rho_6 P_6 = \lambda_2 P_2 + \mu_3 P_5 + \mu_1 P_{11}, && \rho_7 P_7 = 2\lambda_2 P_1 + \lambda_1 P_2,\\
&\rho_8 P_8 = \lambda_1 P_4 + \lambda_3 P_7, && \rho_9 P_9 = \lambda_1 P_5,\\
&\rho_{10} P_{10} = \lambda_3 P_1 + \lambda_1 P_3, && \rho_{11} P_{11} = \lambda_1 P_6 + \lambda_2 P_7,\\
&P_0 + P_1 + \ldots + P_{11} = 1.
\end{aligned} \tag{6.188}$$

with ρ_i as in Eq. (6.185). One (arbitrarily chosen, because of the implicitly assumed ergodicity) of the first 12 equations must be dropped and replaced by the 13th one, $\sum P_i = 1$, because of linear dependence. The solution yields P_0 to P_{11}, from which

$$PA_S = P_0 (1 + b_1 + b_2 + b_3 + b_4 + b_5 + b_6 + b_7), \tag{6.189}$$

with

$$P_0 = 1 / (1 + \sum_{i=1}^{11} b_i) \tag{6.190}$$

and

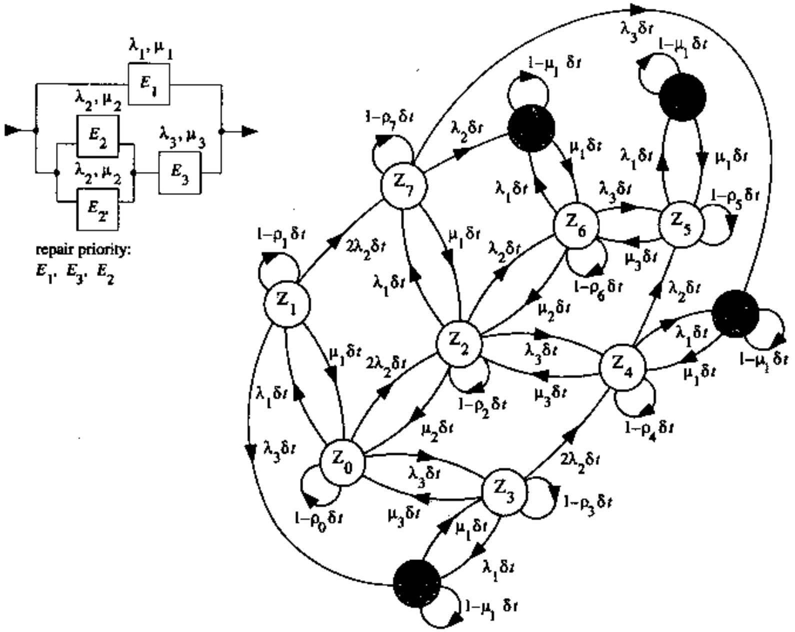
$$b_1 = \frac{\lambda_1}{\rho_1}, \qquad b_2 = \frac{\rho_0 - \lambda_1 \mu_1 / \rho_1}{\mu_2} - \frac{\mu_3 \lambda_3 (1 + \lambda_1 / \rho_1)}{(\mu_3 + 2\lambda_2)\mu_2},$$

$$b_3 = \frac{\lambda_3 (1 + \lambda_1 / \rho_1)}{\mu_3 + 2\lambda_2}, \qquad b_4 = \frac{\lambda_3 b_2 (1 + \lambda_1 / \rho_7) + 2\lambda_2 b_3 + \dfrac{2\lambda_1 \lambda_2 \lambda_3}{\rho_7 \rho_1}}{\rho_4 - \lambda_1},$$

$$b_5 = \frac{\lambda_2 b_4 + \dfrac{\lambda_2 \lambda_3}{\rho_7 (\mu_2 + \lambda_3)} (b_2 (\rho_7 + \lambda_1) + 2\lambda_1 \lambda_2 / \rho_1)}{\rho_5 - \lambda_1 - \dfrac{\mu_3 \lambda_3}{\mu_2 + \lambda_3}}, \qquad b_7 = \frac{2\lambda_1 \lambda_2}{\rho_1 \rho_7} + \frac{\lambda_1}{\rho_7} b_2,$$

$$b_6 = \frac{\lambda_2}{\mu_2 + \lambda_3}\left(b_2 + \frac{2\lambda_1 \lambda_2}{\rho_1 \rho_7} + \frac{\lambda_1}{\rho_7} b_2 + \frac{\mu_3}{\lambda_2} b_5\right), \qquad b_8 = \frac{\lambda_1}{\mu_1} b_4 + \frac{\lambda_3}{\mu_1} b_7$$

$$b_9 = \frac{\lambda_1}{\mu_1} b_5, \qquad b_{10} = \frac{\lambda_3 \lambda_1}{\mu_1 \rho_1} + \frac{\lambda_1}{\mu_1} b_3, \qquad b_{11} = \frac{\lambda_1}{\mu_1} b_6 + \frac{\lambda_2}{\mu_1} b_7. \tag{6.191}$$



$\rho_{01} = \lambda_1$; $\rho_{02} = 2\lambda_2$; $\rho_{03} = \lambda_3$; $\rho_{10} = \mu_1$; $\rho_{17} = 2\lambda_2$; $\rho_{1\,10} = \lambda_3$; $\rho_{20} = \mu_2$;
$\rho_{24} = \lambda_3$; $\rho_{26} = \lambda_2$; $\rho_{27} = \lambda_1$; $\rho_{30} = \mu_3$; $\rho_{34} = 2\lambda_2$; $\rho_{3\,10} = \lambda_1$; $\rho_{42} = \mu_3$;
$\rho_{45} = \lambda_2$; $\rho_{48} = \lambda_1$; $\rho_{56} = \mu_3$; $\rho_{59} = \lambda_1$; $\rho_{62} = \mu_2$; $\rho_{65} = \lambda_3$; $\rho_{6\,11} = \lambda_1$;
$\rho_{72} = \mu_1$; $\rho_{78} = \lambda_3$; $\rho_{7\,11} = \lambda_2$; $\rho_{84} = \mu_1$; $\rho_{95} = \mu_1$; $\rho_{10\,3} = \mu_1$; $\rho_{11\,6} = \mu_1$

Figure 6.20 Reliability block diagram and diagram of transition probabilities in $(t, t + \delta t]$ of the system described by Fig. 6.19 (active redundancy, one repair crew, repair priority in the sequence E_1, E_3, and E_2, no further failures at syst, down, arbitrary t, $\delta t \downarrow 0$, Markov proc., $\rho_i = \sum_j \rho_{ij}$, $j \neq i$)

An analytical comparison of Eqs. (6.186) with Eqs. (6.178) and (6.182) or of Eq. (6.189) with Eqs. (6.179) and (6.183) is difficult. Numerical evaluation yields (λ and μ in h^{-1}, *MTTF* in h)

λ_1	1/100	1/100	1/1,000	1/1,000
λ_2	1/1,000	1/1,000	1/10,000	1/10,000
λ_3	1/10,000	1/10,000	1/100,000	1/100,000
μ_1	1	1/5	1	1/5
μ_2	1/5	1/5	1/5	1/5
μ_3	1/5	1/5	1/5	1/5
$MTTF_{S0}$ (Eq. (6.178), IE)	$1.575 \cdot 10^{+5}$	$9.302 \cdot 10^{+4}$	$1.657 \cdot 10^{+7}$	$9.926 \cdot 10^{+6}$
$MTTF_{S0}$ (Eq. (6.182), MS)	$1.528 \cdot 10^{+5}$	$9.136 \cdot 10^{+4}$	$1.652 \cdot 10^{+7}$	$9.906 \cdot 10^{+6}$
$MTTF_{S0}$ (Eq. (6.186), no FF)	$1.589 \cdot 10^{+5}$	$9.332 \cdot 10^{+4}$	$1.658 \cdot 10^{+7}$	$9.927 \cdot 10^{+6}$
$MTTF_{S0}$ (Method 4, Cutting)	$1.487 \cdot 10^{+5}$	$9.294 \cdot 10^{+4}$	$1.645 \cdot 10^{+7}$	$9.917 \cdot 10^{+6}$
$MTTF_{S0}$ (only one repair crew)	$1.596 \cdot 10^{+5}$	$9.327 \cdot 10^{+4}$	$1.657 \cdot 10^{+7}$	$9.922 \cdot 10^{+6}$
$1 - PA_S$ (Eq. (6.179), IE)	$5.250 \cdot 10^{-6}$	$2.625 \cdot 10^{-5}$	$5.025 \cdot 10^{-8}$	$2.513 \cdot 10^{-7}$
$1 - PA_S$ (Eq. (6.183), MS)	$2.806 \cdot 10^{-5}$	$5.446 \cdot 10^{-5}$	$2.621 \cdot 10^{-7}$	$5.045 \cdot 10^{-7}$
$1 - PA_S$ (Eq. (6.189), no FF)	$6.574 \cdot 10^{-6}$	$5.598 \cdot 10^{-5}$	$6.060 \cdot 10^{-8}$	$5.062 \cdot 10^{-7}$
$1 - PA_S$ (Method 4, Cutting)	$2.995 \cdot 10^{-5}$	$5.556 \cdot 10^{-5}$	$2.647 \cdot 10^{-7}$	$5.059 \cdot 10^{-7}$
$1 - PA_S$ (only one repair crew)	$6.574 \cdot 10^{-6}$	$5.627 \cdot 10^{-5}$	$6.061 \cdot 10^{-8}$	$5.062 \cdot 10^{-7}$

Also given in the above numerical comparison are the results obtained by method 4 of Section 6.7.1 (for a given precision of 10^{-8} on the unavailability $1 - PA_S$) and by dropping the assumption of no further failures at system down in method 3. These results confirm that for $\lambda_i << \mu_i$ good approximate expressions for practical applications can be obtained from all the methods presented in Section 6.7.1. The influence of λ_i / μ_i appears clearly when comparing columns 1 with 2 and 3 with 4. The results obtained with method 1 of Section 6.7.1 (Eqs. (6.178) and (6.179)) should still give higher values for $MTTF_{S0}$ and PA_S, because of the assumption that each element has its own repair crew (independent elements in operation and repair). Comparing the results form Eqs. (6.186) and (6.189) with those for the case in which the assumption of no further failures at system down is dropped shows (at least for these examples) the small influence of this assumption on the final results.

6.8 Systems with Complex Structure

6.8.1 General Considerations

For reliability investigations, a structure is considered to be *complex* if its reliability block diagram either cannot be reduced to a *series/parallel structure* or does not exist. Examples of such structures are given in Table 2.1 (Examples *7 to 9*) and Section 2.3. Because of the interdependence between performance and reliability as well as of their *reconfiguration* capability, *fault tolerant distributed systems* (e.g. networks) are *complex structures* in the above sense.

If the *reliability block diagram exists*, but it is not possible to reduce it to a series/parallel structure (structure based on series and parallel models with independent elements), reliability and availability analyses can still be performed using *one or more* of the following assumptions, as appropriate:

1. For each element in the reliability block diagram, the failure-free operating time and the repair time are statistically *independent*.
2. Failure and repair rates of each element are *constant* (time independent).
3. Each element in the reliability block diagram has a *constant failure rate*.
4. The flow of failures is a *Poisson process* (homogeneous or inhomogeneous).
5. *No further failures at system down* (i.e. failures during a repair at system level are neglected).
6. Redundant elements are repaired *without interrupting operarion* at system level.
7. After each repair, the *repaired element is as-good-as-new*.
8. After each repair, the *entire system is as-good-as-new*.
9. *Only one repair crew* is available, repair is either started as soon as the repair crew is freed from a running repair (*first-in first-out*) or according to a given *repair strategy* when a failure occurs (running repair stopped, if necessary).
10. Each element works (operates and is repaired) *independently* of every *other element* (*n* repair crews for a reliability block diagram with *n* elements).
11. Failure detection is 100% reliable, i.e. no *hidden failures* exist.
12. All failure-free operating times and repair times are positive, continuous, and have a *finite mean and variance*.
13. For each element, the mean time to repair is *much lower* than the mean time to failure ($MTTR_i << MTTF_i$).
14. Switches and *switching operations* are 100% reliable and have no influence on the reliability of the system.
15. *Preventive maintenance* is not considered.

A list of the assumpions made is important to *fix the validity of the results* obtained for the case under consideration. It is often tacitly assumed that each element has just 2 states (good/failed), only one *failure mode* (e.g. shorts or opens), a time-invariant required function, and that the system is in *continuous operation* (each element is operating when not under repair or waiting for repair). A time dependent operation and/or required function (performance) can often be easily investigated when constant failure rates are assumed.

The following is a brief discussion of the above assumptions. With assumptions 1 and 2, the time behavior of the system can be described by a (time) *homogeneous Markov process* with finitely many states. Equations can be established using the *diagram* of *transition probabilities in* $(t, t+\delta t]$ and Table 6.2. Difficulties can arise because of the *large number of states involved*, often close to $e \cdot n!$ (Section 6.7). In such cases, a first possibility is to limit the investigations to the computation of the mean time to failure $MTTF_{S0}$ and of the asymptotic and steady-state values of the point and average availability $PA_S = AA_S$, i.e. to the solution of *algebraic equations*. A second possibility is to use *approximate expressions* as described in Section 6.7 or special software tools (Section 6.8.2). Assumptions 7 and 8 are satisfied if either assumption 2 or 3 holds. Assumption 7 is frequently used, its *validity must be checked*. Assumption 8 is seldom used. Assumption 4 often applies to systems with a large number of elements. As shown in Sections 6.3 to 6.6, assumption 5 simplifies the calculation of the point availability and interval reliability. It gives useful *approximate expressions*, particularly when assumption 13 applies (see Section 6.7.2 for an example). Assumption 6 must be met during the *system design*. If it is not satisfied, the improvement in reliability and availability given by redundancy is poor (see Example 6.15 and Figs. 6.17 and 2.8). As a minimum, *fault detection* must be required and implemented. Assumption 9 simplifies the calculations. It is useful for obtaining *approximate expressions*, especially if assumption 13 holds. Together with assumption3, the time behavior of the system can be described by a *semi-regenerative process* (process with an embedded semi-Markov process). Assumption 3 alone can only assure that the process is *regenerative* (often with only one regeneration state). With assumption 10, point availability can be computed using the reliability equation for the non repairable case (Eqs. (2.47) and (2.48)). This assumption rarely applies in practical applications. However, it does allow a simple computation of an *upper bound* for the point availability. Assumption 13 is often met. It leads to *approximate expressions*, as illustrated in Section 6.7 or by using asymptotic expansions, see e.g. [6.11, A7.26]. As shown in Examples 6.8 and 6.9, the shape of the distribution function of the repair time has a small influence on final results ($MTTF_{S0}$, PA_S, $\mathrm{IR}_S(\theta)$) if assumption 13 holds. Assumptions 14 and 15 simplify investigations. They are *valid for all models discussed in Sections 6.2 to 6.7* (examples of systems with *preventive maintenance*, *imperfect switching* and *hidden failures* are considered in Sections 6.9 and 6.10).

In general, investigation of *large series/parallel structures* or of complex structures is time consuming and can become analytically intractable. As a first step it is therefore often useful to work with *Markov or semi-Markov models* and develop either exact solutions (analytically as in Sections 6.2 to 6.6 or numerically with computer programs as in Section 6.8.2) or approximate expressions (Section 6.7). In a second step, refinements can be considered on a case-by-case basis. For difficult situations, Monte Carlo simulations may be the only way to get results.

For complex fault-tolerant distributed systems, *graceful degradation* at system level must be considered (beside the method discussed in this book and based on the concept of a given (fixed) required function). Investigations on the combination of performance and reliability to *performability* are in progress, see e.g. [6.5, 6.6, 6.16].

If a *reliability block diagram* does not *exists*, stochastic processes and tools introduced in Appendix A7 and Chapter 6 can also be used to investigate the reliability and availability of *fault tolerant systems*, on the basis of an extended reliability *state transition diagram* and of the method of *integral equations* (a publication on this subject is in preparation).

6.8.2 Computer Aided Reliability and Availability Prediction

Computation of the reliability of a complex nonrepairable system, or of the reliability and availability of a *repairable system* can become time-consuming. Software tools exist to solve this kind of problems [2.51 to 2.57]. From such a software package one generally expects high degrees of *completeness*, *usability*, *robustness*, *integrity*, and *portability* (Tab. 5.4). The following is a comprehensive list of specific requirements:

General requirements:

1. Possibility for a design engineer to work with a software interface similar to that of CAD/CAE tools.
2. Provide a well structured data bank with at least 10,000 components, possibility for manufacturer and company-specific labeling, and permanent storage of non application-specific data.
3. Accept input file for part lists with interface to *configuration management* software packages.
4. Provide component failure rate calculations for different failure rate models (e. g. Bellcore , CNET , IEC [2.21 to 2.29]).
5. Have flexible output (regarding medium, sorting capability, weighting), graphic interface, single and multi-user capability, high usability and integrity.
6. Be portable to different platforms by keeping one source of the program that can be compiled on the other platforms.

Specific for nonrepairable systems:

7. Consider reliability block diagrams (RBD) of *arbitrary complexity* and with a large number of elements (≥ 1000) and levels (≥ 10); possibility for any element to appear more than once in the RBD; automatically edition of series and parallel models; powerful method to handle complex structures; constant or time dependent failure rate for each element; possibility to handle as element (with given reliability function) items with more than one failure mode, macro structures, etc.; easy edition of *application-specific* data, with user features such as:
   - automatic computation of the ambient temperature at component level, starting from the ambient temperature at system level and the freely selectable temperature difference between elements,
   - freely selectable duty cycle from the system level downwards,
   - global change of environmental factor (π_E) by set-instructions, manual selection of stress factors (S) for tradeoff studies and risk assessment, and manual introduction of field data and of *default values* for component families or assemblies.
8. Allow reuse of elements with arbitrary complexity in a RBD (libraries).

Specific for repairable systems:

9. Consider elements with *constant failure rate* and const. or *arbitrary repair rate*, i.e. be applicable to Markov, semi-Markov, and semi-regenerative processes.
10. Have *automatic* generation of the transition rates ρ_{ij} for Markov model and of the involved semi Markov transition probabilities $Q_{ij}(x)$ for systems with constant failure rates, one repair crew, and arbitrary repair rate, starting from a given set of *succesful paths* (ideally, directly from a given reliability block diagram); automatic generation and solution of the equations describing the system's behavior (algebraic, differential, or integral equations).
11. Allow *different repair strategies*, for instance first-in first-out, only one repair crew, user-defined priority.
12. Use sophisticated *algorithms* for quick inversion of sparse matrices.
13. Consider at least 20,000 states for the *exact solution* of the steady-state availability $PA_S = AA_S$, mean time to system failure $MTTF_S$, and steady-state interval reliability $\mathrm{IR}_S(\theta)$.
14. Deliver *solutions* for some important approximation methods (for instance methods 1 to 5 in Section 6.7.1), if possible with indication of error bounds.

A scientific software package satisfying many of the above requirements has been developed at the Reliability Lab. of the ETH [2.51 (1995 & 1997)]. For basic series/parallel structures, commercial programs are available on the market [2.52 to 2.58], among these are for example Relex and RAC PRISM in the USA and RAMTOOL in Europe.

6.9 Influence of Imperfect Switching

For reliability analyses, *switching* is necessary for powering down failed elements or powering up repaired elements. In many cases it is sufficient to locate the *switching element* in series with the redundancy on the reliability block diagram, yielding series/parallel structures as investigated in Section 6.6. However, such an approach is often too simple to cover real situations. Figure 6.21 gives an example in which measurement points M_1 and M_2, switches S_1 and S_2, as well as a control unit C must be considered. To simplify, let us consider here only the reliability function in the nonrepairable case (up to system failure). From a reliability point of view, switch S_i, element E_i, and measurement point M_i in Fig. 6.21 are in series ($i = 1, 2$). Let τ_{b1} and τ_{b2} be the corresponding failure-free operating times with distribution function $\mathrm{F}_b(t)$ and density $\mathrm{f}_b(t)$. τ_c is the failure-free operating time of the control device with distribution function $\mathrm{F}_c(t)$ and density $\mathrm{f}_c(t)$.

Consider first the case of *standby redundancy* and assume that at $t = 0$ element E_1 is switched on. A system failure in the interval $(0, t]$ occurs with one of the following mutually exclusive events

$$\tau_c > \tau_{b1} \cap (\tau_{b1} + \tau_{b2}) \leq t$$

or

$$\tau_c < \tau_{b1} \leq t.$$

It is implicitly assumed here that a failure of the control device has no influence on the operating element, and does not lead to a commutation to E_2. A verification of these conditions by a *FMEA*, as introduced in Section 2.7, is necessary. With these assumptions, the *reliability function* $\mathrm{R}_S(t)$ of the system described by Fig. 6.21 is given by (nonrepairable case)

$$\mathrm{R}_S(t) = 1 - [\int_0^t \mathrm{f}_b(x)(1 - \mathrm{F}_c(x))\mathrm{F}_b(t - x)dx + \int_0^t \mathrm{f}_b(x)\mathrm{F}_c(x)dx]. \tag{6.192}$$

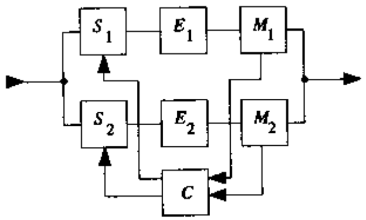


Figure 6.21 Functional block diagram for a 1-out-of-2 redundancy with switches S_1 and S_2, measurement points M_1 and M_2, and control device C

From Eq. (6.192), $f_b(t)=\lambda_b e^{-\lambda_b t}$ and $f_c(t)=\lambda_c e^{-\lambda_c t}$ would yield

$$R_S(t)=e^{-\lambda_b t}+(1-e^{-\lambda_c t})\frac{\lambda_b}{\lambda_c}e^{-\lambda_b t} \qquad (6.193)$$

and the *mean time to failure*

$$MTTF_S=\frac{2\lambda_b+\lambda_c}{\lambda_b(\lambda_b+\lambda_c)}. \qquad (6.194)$$

$\lambda_c\equiv 0$ or $F_c(t)\equiv 0$ leads to the results of Section 2.3.5 for the 1-out-of-2 standby redundancy.

Assuming now an *active redundancy* (at $t=0$, E_1 is put into operation and E_2 into the reserve state), a system failure occurs in the interval $(0,t]$ with one of the following *mutually exclusive events*

$$\tau_{b1}\le t\cap\tau_c>\tau_{b1}\cap\tau_{b2}\le t$$

or

$$\tau_c<\tau_{b1}\le t.$$

The *reliability function* $R_S(t)$ is then given by (nonrepairable case)

$$R_S(t)=1-[F_b(t)\int_0^t f_b(x)(1-F_c(x))dx+\int_0^t f_b(x)F_c(x)dx]. \qquad (6.195)$$

From Eq.(6.195) and assuming

$$f_b(t)=\lambda_b e^{-\lambda_b t} \qquad \text{and} \qquad f_c(t)=\lambda_c e^{-\lambda_c t}$$

it follows that

$$R_S(t)=\frac{2\lambda_b+\lambda_c}{\lambda_b+\lambda_c}e^{-\lambda_b t}-\frac{\lambda_b}{\lambda_b+\lambda_c}e^{-(2\lambda_b+\lambda_c)t} \qquad (6.196)$$

and

$$MTTF_S=\frac{2\lambda_b+\lambda_c}{\lambda_b(\lambda_b+\lambda_c)}-\frac{\lambda_b}{(\lambda_b+\lambda_c)(2\lambda_b+\lambda_c)}. \qquad (6.197)$$

$\lambda_c\equiv 0$ or $F_c(t)\equiv 0$ leads to the results of Section 2.2.6.3 for the 1-out-of-2 active redundancy.

6.10 Influence of Preventive Maintenance

Preventive maintenance is necessary to avoid *wearout failures* and to identify and repair *hidden failures* (i.e. failures of redundant elements which cannot be detected during normal operation). This section investigates some basic situations.

6.10.1 One-item Repairable Structure

Let us first consider a *one-item repairable* structure for which preventive maintenance is performed at periodic time intervals T_{PM}. The item is new at $t=0$. Its failure-free operating time is distributed according to $F(t)$ with density $f(t)$, the repair time has distribution $G(t)$ with density $g(t)$. Preventive maintenance is of negligible time duration (performed by specialized personnel with all the necessary tools) and restores the item to an *as-good-as-new* state. If a preventive maintenance is due at a time in which the item is under repair, one of the following cases will apply:

1. Preventive maintenance will not be performed (included in the repair, because it is assumed as in Section 6.2 that after each repair the item is as-good-as-new).
2. Preventive maintenance is performed, i.e. the running repair is terminated with the preventive maintenance in a negligible time span.

Both situations can occur in practical applications. In case 2, the times 0, T_{PM}, $2T_{PM}$, ... are *renewal points* (Fig. A7.1). This case will be considered in the following.

The *reliability function* $R_{PM}(t)$ can be computed from

$$R_{PM}(t)=R(t), \qquad \text{for } 0\le t<T_{PM}$$

$$R_{PM}(t)=R^n(T_{PM})R(t-nT_{PM}), \qquad \text{for } nT_{PM}\le t<(n+1)T_{PM},\quad n\ge 1, \qquad (6.198)$$

with $R(t)=1-F(t)$, where $F(t)$ is the distribution function of the failure-free operating time of the one-item structure considered. Figure 6.22 shows the shape of $R(t)$ and $R_{PM}(t)$ for an arbitrary $F(t)$, and for $F(t)=1-e^{-\lambda t}$. Because of the *memoryless property* which characterizes exponentially distributed failure-free operating times, $R_{PM}(t)=R(t)=e^{-\lambda t}$ holds for $F(t)=1-e^{-\lambda t}$ *independently* of T_{PM}. From Eq. (6.198), the mean time to failure with preventive maintenance $MTTF_{PM}$ follows as

$$MTTF_{PM} = \int_0^{\infty} \mathrm{R}_{PM}(t)dt = [1 + \sum_{n=1}^{\infty} \mathrm{R}^n(T_{PM})] \int_0^{T_{PM}} \mathrm{R}(t)dt = \frac{\int_0^{T_{PM}} \mathrm{R}(t)dt}{1 - \mathrm{R}(T_{PM})}. \tag{6.199}$$

For $\mathrm{F}(t) = 1 - e^{-\lambda t}$, Eq. (6.199) yields $MTTF_{PM} = 1/\lambda$ *independently* of the period T_{PM} of the preventive maintenance. Determination of the *optimal preventive maintenance period* must consider Eq. (6.199) as well as *cost* and *logistical support* aspects (for $\mathrm{F}(0) = 0$, $MTTF_{PM} \to \infty$ for $T_{PM} \to 0$).

Calculation of the *point availability* is easy if preventive maintenance is performed at the times $0, T_{PM}, 2T_{PM}, \ldots$ (case 2 above) and leads to

$$\mathrm{PA}_{PM}(t) = \mathrm{PA}_0(t) \qquad \text{for } 0 \le t < T_{PM}$$

$$\mathrm{PA}_{PM}(t) = \mathrm{PA}_0(t - nT_{PM}) \qquad \text{for } nT_{PM} \le t < (n+1)T_{PM}, \ n \ge 1 \tag{6.200}$$

with $\mathrm{PA}_0(t)$ from Eq. (6.17). Figure 6.23 shows the shape of $\mathrm{PA}_{PM}(t)$ as given by Eq. (6.200). Contrary to $\mathrm{R}_{PM}(t)$, $\mathrm{PA}_{PM}(t)$ goes to 1 at $0, T_{PM}, 2T_{PM}, \ldots$, i.e. at each renewal point.

If the time duration for the preventive maintenance is not negligible, it is useful to define, in addition to the availabilities introduced in Section 6.2.1, the *overall availability OA*, defined for $t \to \infty$ as the ratio of the total up time in $(0, t]$ to the sum of total up and down time in $(0, t]$, i.e. to t. Defining $MTTF$ = mean time to failure and MDT = mean down time (with $MTTR$ = mean time to repair, $MTTPM$ = mean time to carry out preventive maintenance, MLD = *mean logistic delay*, and T_{PM} = preventive maintenance period) it follows that

$$OA = \frac{MTTF}{MTTF + MDT} = \frac{MTTF}{MTTF + MTTR + MLD + \frac{MTTF}{T_{PM}} MTTPM}. \tag{6.201}$$

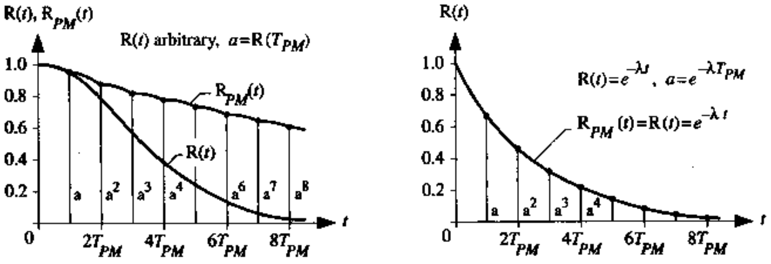


Figure 6.22 Reliability functions of a one-item structure with preventive maintenance (of negligible duration) at times T_{PM}, $2T_{PM}$, ... for two distribution functions $\mathrm{F}(t)$ of the failure-free operating times (new at $t = 0$)

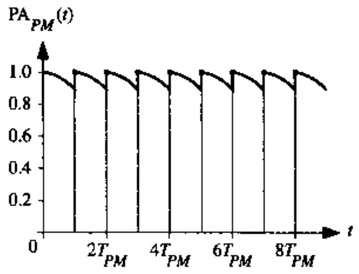


Figure 6.23 Point availability of a one-item structure with repair at every failure and preventive maintenance (of negligible duration) at times T_{PM}, $2T_{PM}$, ... (new at $t = 0$)

Example 6.13

Assume a nonrepairable (up to system failure) 1-out-of-2 active redundancy with two identical elements with constant failure rate λ. Compute the mean time to failure $MTTF_{PM}$ by assuming a preventive maintenance with period $T_{PM} << MTBF = 1/\lambda$. The preventive maintenance is performed in a negligible time span and restores the 1-out-of-2 active redundancy as-good-as-new.

Solution

For a nonrepairable (up to system failure) 1-out-of-2 active redundancy with two identical elements with constant failure rate λ, the relibility function is given by Eq. (2.20)

$$\mathrm{R}(t) = 2e^{-\lambda t} - e^{-2\lambda t}$$

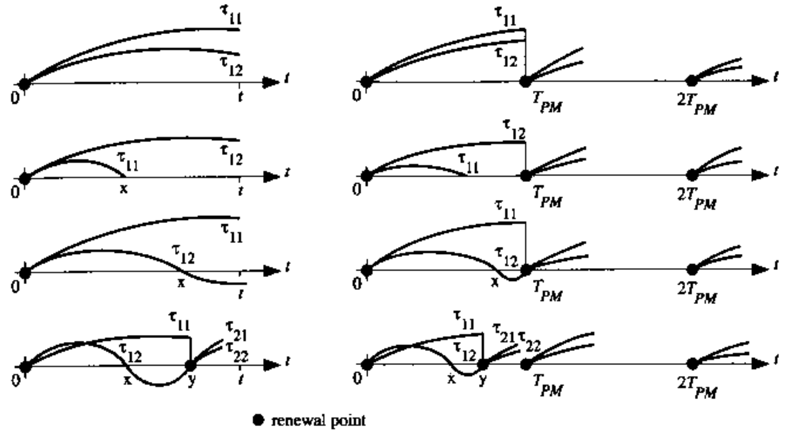
The mean time to failure with preventve maintenance follows then from Eq. (6.199) as

$$MTTF_{PM} = \frac{\int_0^{T_{PM}} \mathrm{R}(t)dt}{1 - \mathrm{R}(T_{PM})} = \frac{\int_0^{T_{PM}} (2e^{-\lambda t} - e^{-2\lambda t})dt}{1 - 2e^{-\lambda T_{PM}} + e^{-2\lambda T_{PM}}} = \frac{\frac{2}{\lambda}(1 - e^{-\lambda T_{PM}}) - \frac{1}{2\lambda}(1 - e^{-2\lambda T_{PM}})}{1 - 2e^{-\lambda T_{PM}} + e^{-2\lambda T_{PM}}}$$

$$\approx \frac{2T_{PM} - T_{PM}}{\lambda^2 T_{PM}^2} = \frac{1}{\lambda^2 T_{PM}} = MTBF \frac{MTBF}{T_{PM}} \qquad (\text{using } e^{-x} \approx 1 - x + x^2/2). \tag{6.202}$$

6.10.2 1-out-of-2 Active Redundancy with Hidden Failures in one Element

Let us consider now a 1-out-of-2 active redundancy with two different elements E_1 and E_2, and assume that failures of E_1 can be detected only during the repair of E_2 or at a preventive maintenance (*hidden failures* in E_1). Elements E_1 and E_2 have



a) Without preventive maintenance b) With preventive maintenance (period T_{PM})

Figure 6.24 Time schedule of a repairable 1-out-of-2 parallel redundancy with hidden failures in element E_1 (new at $t = 0$)

constant failure rates λ_1 and λ_2, the repair time of E_2 is distributed according to $\mathrm{G}(t)$ with a density $\mathrm{g}(t)$, and the repair of E_1 takes a negligible time.

If no *preventive maintenance* is performed, Fig. 6.24a shows a possible time schedule of the system (new at $t = 0$), yielding for the *reliability function*

$$\mathrm{R}_{S0}(t) = e^{-(\lambda_1+\lambda_2)t} + \int_0^t \lambda_1 e^{-\lambda_1 x} e^{-\lambda_2 t} dx + \int_0^t \lambda_2 e^{-\lambda_2 x} e^{-\lambda_1 t} (1 - \mathrm{G}(t-x)) dx$$
$$+ \int_0^t \int_0^y \lambda_2 e^{-\lambda_2 x} e^{-\lambda_1 y} \mathrm{g}(y-x) \mathrm{R}_{S0}(t-y) dx\, dy. \qquad (6.203)$$

The Laplace transform of $\mathrm{R}_{S0}(t)$ follows as

$$\tilde{\mathrm{R}}_{S0}(s) = \frac{(s+\lambda_1)(s+\lambda_1+\lambda_2) + \lambda_2 (s+\lambda_2)(1-\tilde{\mathrm{g}}(s+\lambda_1))}{(s+\lambda_1)(s+\lambda_2)(s+\lambda_1+\lambda_2) - (s+\lambda_1)(s+\lambda_2)\lambda_2 \tilde{\mathrm{g}}(s+\lambda_1)}, \qquad (6.204)$$

and thus the *mean time to failure* becomes

$$MTTF_{S0} = \frac{\lambda_1(\lambda_1+\lambda_2) + \lambda_2^2 (1-\tilde{\mathrm{g}}(\lambda_1))}{\lambda_1 \lambda_2 (\lambda_1+\lambda_2) - \lambda_1 \lambda_2^2 \tilde{\mathrm{g}}(\lambda_1)}. \qquad (6.205)$$

If *preventive maintenance* is performed at times $T_{PM}, 2T_{PM}, \ldots$ independently of the state of element E_2, and if after each preventive maintenance (assumed to be of negligible duration as in Section 6.10.1) the entire system is *as-good-as-new*, then the times $0, T_{PM}, 2T_{PM}, \ldots$ are *renewal points* for the system. For the *reliability function* $\mathrm{R}_{S0_{PM}}(t)$ it follows that (considering Eq. (6.198) and Fig. 6.24b)

$$\mathrm{R}_{S0_{PM}}(t) = \mathrm{R}_{S0}(t), \qquad \text{for} \quad 0 \le t < T_{PM}$$
$$\mathrm{R}_{S0_{PM}}(t) = \mathrm{R}_{S0}^n(T_{PM}) \mathrm{R}_{S0}(t - nT_{PM}), \qquad \text{for} \quad nT_{PM} \le t < (n+1)T_{PM}, \quad n \ge 1 \qquad (6.206)$$

and the *mean time to failure*

$$MTTF_{S0_{PM}} = \frac{\int_0^{T_{PM}} \mathrm{R}_{S0}(t)\, dt}{1 - \mathrm{R}_{S0}(T_{PM})}. \qquad (6.207)$$

In Eqs. (6.206) and (6.207), $\mathrm{R}_{S0}(t)$ is as in Eq. (6.203). The time T_{PM} between two consecutive preventive maintenance operations can now be *optimized* considering Eqs. (6.205) and (6.207) as well as cost and logistical aspects.

Example 6.14

Derive *approximate expressions* for the mean time to failure $MTTF_{S0}$ given by Eq. (6.205).

Solution

For $\tilde{\mathrm{g}}(\lambda_1) \to 1$, it follows from Eq. (6.205) that

$$MTTF_{S0} \approx \frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}. \qquad (6.208)$$

A better approximation is obtained by considering $\tilde{\mathrm{g}}(\lambda_1) = 1 - \lambda_1 MTTR$

$$MTTF_{S0} \approx \frac{\lambda_1 + \lambda_2 + \lambda_2^2 MTTR}{\lambda_1 \lambda_2 (1 + \lambda_2 MTTR)} \qquad (6.209)$$

with $MTTR$ as in Eq. (6.110).

Example 6.15

Investigate $\mathrm{R}_{S0}(t)$ as in Eq. (6.203) and $\mathrm{R}_{S0_{PM}}(t)$ as in Eq. (6.206) as well as $MTTF_{S0}$ as in Eq. (6.205) and $MTTF_{S0_{PM}}$ as in Eq. (6.207) for the case of constant repair rate μ, i.e. by assuming $\mathrm{g}(t) = \mu e^{-\mu t}$.

Solution

With $\tilde{\mathrm{g}}(s+\lambda_1) = \mu / (s + \lambda_1 + \mu)$ it follows from Eq. (6.204) that

$$\tilde{\mathrm{R}}_{S0}(s)=\frac{(s+\lambda_1+\lambda_2)(s+\lambda_1+\mu)+\lambda_2(s+\lambda_2)}{(s+\lambda_1)(s+\lambda_2)(s+\lambda_1+\lambda_2+\mu)} \tag{6.210}$$

and thus $\mathrm{R}_{S0}(t)=A\,e^{-\lambda_1 t}+B\,e^{-\lambda_2 t}+C\,e^{-(\lambda_1+\lambda_2+\mu)t}$ with

$$A=\frac{\lambda_2(\lambda_2-\lambda_1+\mu)}{(\lambda_2-\lambda_1)(\lambda_2+\mu)},\qquad B=\frac{-\lambda_1(\lambda_1-\lambda_2+\mu)}{(\lambda_2-\lambda_1)(\lambda_1+\mu)},\qquad C=\frac{-\lambda_1\lambda_2}{(\lambda_1+\mu)(\lambda_2+\mu)}.$$

The mean time to failure $MTTF_{S0}$ follows then from Eq. (6.210)

$$MTTF_{S0}=\tilde{\mathrm{R}}_{S0}(0)=\frac{(\lambda_1+\lambda_2)(\lambda_1+\mu)+\lambda_2^2}{\lambda_1\lambda_2(\lambda_1+\lambda_2+\mu)}\approx\frac{\lambda_1+\lambda_2}{\lambda_1\lambda_2}=\frac{1}{\lambda_1}+\frac{1}{\lambda_2}. \tag{6.211}$$

Assuming, $\lambda_1+\lambda_2<<\mu$ yields

$$\mathrm{R}_{S0}(t)\approx\frac{\lambda_2\,e^{-\lambda_1 t}-\lambda_1\,e^{-\lambda_2 t}}{\lambda_2-\lambda_1} \tag{6.212}$$

and

$$MTTF_{S0}=\frac{\lambda_1+\lambda_2}{\lambda_1\lambda_2}=\frac{1}{\lambda_1}+\frac{1}{\lambda_2}. \tag{6.213}$$

Equations (6.211) to (6.213) show that the *repairable* 1-out-of-2 active redundancy with *hidden failures* in one element *behaves like a nonrepairable* 1-out-of-2 standby redundancy. This result shows how important it is in the presence of redundancy to investigate *failure detection* and *failure modes*. In the case of periodic preventive maintenance (period T_{PM}), Eq. (6.207) yields

$$MTTF_{S0_{PM}}\approx\frac{\frac{\lambda_2}{\lambda_1}(1-e^{-\lambda_1 T_{PM}})-\frac{\lambda_1}{\lambda_2}(1-e^{-\lambda_2 T_{PM}})}{\lambda_2(1-e^{-\lambda_1 T_{PM}})-\lambda_1(1-e^{-\lambda_2 T_{PM}})}\approx\frac{2}{\lambda_1\lambda_2 T_{PM}}. \tag{6.214}$$

The last part of Eq. (6.214) has been obtained using $e^{-\lambda x}\approx 1-\lambda x+(\lambda x)^2/2$. The optimization of the time T_{PM} between two consecutive preventive maintenance operations can be performed using Eq. (6.214), also paying attention to cost and logistical aspects. It should be noted that Eq. (6.210) could also have been obtained directly using Table 6.2 (Markov process) with up states Z_0, Z_1, and Z_2, absorbing state Z_3, and transition rates $\rho_{01}=\lambda_1$, $\rho_{02}=\lambda_2$, $\rho_{13}=\lambda_2$, $\rho_{20}=\mu$, and $\rho_{23}=\lambda_1$, as well as $\mathrm{P}_0(0)=1$.

---

7 Statistical Quality Control and Reliability Tests

Statistical quality control and *reliability tests* are performed to estimate or demonstrate quality and reliability characteristics (figures) on the basis of data collected from sampling tests. *Estimation* leads to a point or interval estimate of an unknown characteristic, *demonstration* is a test of a given *hypothesis* on the unknown characteristic in an acceptance test. Estimation and demonstration of an *unknown probability* is investigated in Section 7.1 for the case of a defective probability p and applied in Section 7.2.1 to reliability, maintainability, and availability figures. Estimation and demonstration of a *constant failure rate* λ (or $MTBF=1/\lambda$) and of an $MTTR$ are discussed in depth in Sections 7.2.2 and 7.3. Basic models for *accelerated tests* and for *goodness-of-fit tests* are considered in Sections 7.4 and 7.5, respectively. To simplify the notation, the term *sample* will be used instead of *random sample*. Theoretical foundations for this chapter are given in Appendix A8. Empirical and graphical methods are considered in Section 7.5 and Appendices A8.1 and A9.8.

7.1 Statistical Quality Control

One of the main purposes of *statistical quality control* is to use *sampling tests* to *estimate* or *demonstrate* the *defective probability* p of a given item, to a required accuracy and often on the basis of tests by *attributes* (i.e. tests of type good/bad). However, considering p as an *unknown probability*, a broader field of applications can be covered by the methods used in statistical quality control. Other tasks of statistical quality control, such as *tests by variables* and *statistical processes control* will not be considered in this book. For these one may refer e.g. to [7.1 to 7.5].

In this section, p will be considered as a *defective probability*. It will be assumed that p is the same for each element in the (random) sample considered, and that each sample element is statistically independent from each other. These assumptions presuppose that the lot is ***homogeneous*** and ***much larger*** than the sample. They allow the use of the ***binomial distribution*** (Appendix A6.10.7).

7.1.1 Estimation of a Defective Probability p

Let n be the size of a (random) sample from a large homogeneous lot. If k defective items have been observed within the sample of size n, then

$$\hat{p} = \frac{k}{n} \tag{7.1}$$

is the *maximum likelihood point estimate* of the defective probability p for an item in the lot under consideration, see Eq. (A8.29). For a given *confidence level* $\gamma = 1 - \beta_1 - \beta_2 \ (0 < \beta_1 < 1 - \beta_2 < 1)$, the *lower* $\hat{p}_l$ and *upper* $\hat{p}_u$ limit of the *confidence interval* of p can be obtained from

$$\sum_{i=k}^{n} \binom{n}{i} \hat{p}_l^i (1-\hat{p}_l)^{n-i} = \beta_2 \qquad \text{and} \qquad \sum_{i=0}^{k} \binom{n}{i} \hat{p}_u^i (1-\hat{p}_u)^{n-i} = \beta_1 \tag{7.2}$$

for $0 < k < n$, and from

$$\hat{p}_l = 0 \qquad \text{and} \qquad \hat{p}_u = 1 - \sqrt[n]{\beta_1} \qquad \text{for } k = 0 \quad (\gamma = 1 - \beta_1), \tag{7.3}$$

or from

$$\hat{p}_l = \sqrt[n]{\beta_2} \qquad \text{and} \qquad \hat{p}_u = 1 \qquad \text{for } k = n \quad (\gamma = 1 - \beta_2), \tag{7.4}$$

see Eqs. (A8.37) to (A8.40) and the remarks given there. β_1 is the risk that the true value of p is larger than $\hat{p}_u$ and β_2 the risk that the value of p is smaller than $\hat{p}_l$. The *confidence level* is nearly equal to (but not less than) $\gamma = 1 - \beta_1 - \beta_2$. It can be explained as the relative frequency of cases in which the interval $[\hat{p}_l, \hat{p}_u]$ *overlaps* the true value of p, in an increasing series of repetitions of the experiment of taking a random sample of size n.

In many practical applications, a graphical determination of $\hat{p}_l$ and $\hat{p}_u$ is sufficient. The upper diagram in Fig. 7.1 can be used for $\beta_1 = \beta_2 = 0.05$, the lower diagram for $\beta_1 = \beta_2 = 0.1$ ($\gamma = 0.9$ and $\gamma = 0.8$). The continuous lines in Fig. 7.1 are the envelopes of staircase functions (k, n integer) given by Eq. (7.2). They converge rapidly, for $\min(np, n(1-p)) > 5$, to the *confidence ellipses* (dashed lines in Fig. 7.1). Using the confidence ellipses (Eq. (A8.42)), $\hat{p}_l$ and $\hat{p}_u$ can be computed from

$$\hat{p}_{u,l} = \frac{k + 0.5b^2 \pm b\sqrt{k(1-k/n) + b^2/4}}{n + b^2}. \tag{7.5}$$

b is the $(1+\gamma)/2$ quantile of the standard normal distribution $\Phi(t)$, given for some typical values of γ by (Table A9.1)

$\gamma =$	0.6	0.8	0.9	0.95	0.98	0.99
$b =$	0.84	1.28	1.64	1.96	2.33	2.58

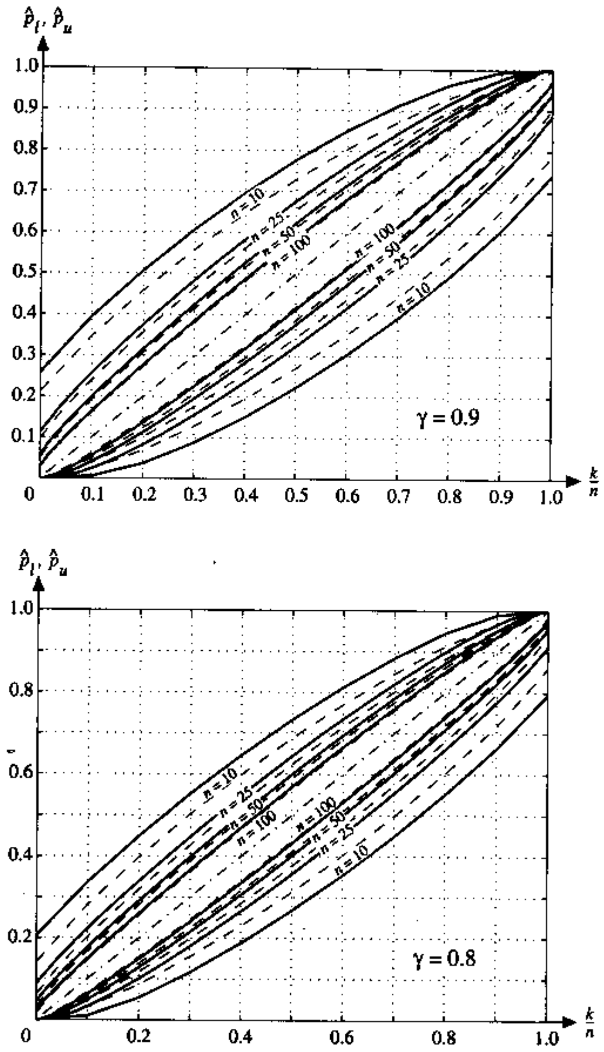


Figure 7.1 Confidence limits $\hat{p}_l$ and $\hat{p}_u$ for an unknown probability p (e.g. defective probability) as a function of the observed relative frequency k/n (n = sample size, γ = confidence level = $1 - \beta_1 - \beta_2$ with $\beta_1 = \beta_2$; continuous lines are the exact solution according to Eq. (7.2), dashed are the approximations according to Eq. (7.5))

Example: $n = 25$, $k = 5$ gives $\hat{p} = k/n = 0.2$ and for $\gamma = 0.9$ the confidence interval [0.08, 0.38] ([0.0823, 0.3754] using Eq. (7.2), and [0.1011, 0.3572] using Eq. (7.5))

The confidence limits $\hat{p}_l$ and $\hat{p}_u$ can also be used as *one-sided confidence intervals*. In this case,

$$
\begin{aligned}
&0 \le p \le \hat{p}_u \quad (\text{or simply } p \le \hat{p}_u), &&\text{with} \quad \gamma = 1-\beta_1\\
&\hat{p}_l \le p \le 1 \quad (\text{or simply } p \ge \hat{p}_l), &&\text{with} \quad \gamma = 1-\beta_2 .
\end{aligned}
\tag{7.6}
$$

Example 7.1

In a sample of size $n = 25$, exactly $k = 5$ items were found to be defective. Determine for the underlying defective probability p, (i) the point estimate, (ii) the interval estimate for $\gamma = 0.8$ $(\beta_1 = \beta_2 = 0.1)$, (iii) the upper bound of p for a one sided confidence interval with $\gamma = 0.9$.

Solution

(i) Equation (7.1) yields the point estimate $\hat{p} = 5/25 = 0.2$. (ii) For the interval estimate, the lower part of Fig. 7.1 leads to the confidence interval [0.10, 0.34], [0.1006, 0.3397] using Eq. (7.2) and [0.1175, 0.3194] using Eq. (7.5). (iii) With $\gamma = 0.9$ it holds $p \le 0.34$.
Supplmentary result: The upper part of Fig. 7.1, would lead to $p \le 0.38$ with $\gamma = 0.95$.

The role of k/n and p can be *reversed* and Eq. (A8.42) can be used to compute the limits k_1 and k_2 of the number of observations k in n *independent trials* (e.g. the number k of defective items in a sample of size n) for a *given* probability γ $(\gamma = 1-\beta_1-\beta_2$ with $\beta_1 = \beta_2)$ and *known* values of p and n (Eq. (A8.45))

$$k_{2,1} = np \pm b\sqrt{np(1-p)} . \tag{7.7}$$

As in Eq. (7.5), the quantity b in Eq. (7.7) is the $(1+\gamma)/2$ quantile of the standard normal distribution (e.g. $b = 1.64$ for $\gamma = 0.9$, Table A9.1). For a graphical solution, Fig. 7.1 can be used by taking the ordinate p as known, and by reading k_1/n and k_2/n from the abscissa.

7.1.2 Simple Two-sided Sampling Plans for the Demonstration of a Defective Probability *p*

In the context of *acceptance testing*, the *demonstration* of a defective probability p is often required, instead of its estimation as considered in Section 7.1.1. The main concern of this test is to check a *zero hypothesis* $H_0 : p < p_0$ against an *alternative hypothesis* $H_1 : p > p_1$, on the basis of the following agreement between producer and consumer:

> *The lot should be accepted with a probability nearly equal to (but not less than) $1-\alpha$ if the true (unknown) defective probability p is lower than p_0 but rejected with a probability nearly equal to (but not less than) $1-\beta$ if p is greater than p_1 (p_0 and $p_1 > p_0$* are given *(fixed)* values).

p_0 is the *specified* defective probability and p_1 is the *maximum acceptable* defective probability. α is the allowed *producer's risk* (type I error), i.e. the probability of rejecting a true hypothesis $H_0 : p < p_0$. β is the allowed *consumer's risk* (type II error), i.e. the probability of accepting the hypothesis $H_0 : p < p_0$ although the alternative hypothesis $H_1 : p > p_1$ is true (in the following it will be tacitly assumed that $0 < \alpha < 1-\beta < 1$). Verification of the agreement stated above is a problem of statistical hypothesis testing (Appendix A8.3) and can be performed, for instance, with a *simple two-sided sampling plan* or a *sequential test*. In both cases, the basic model is the sequence of *Bernoulli trials*, as introduced in Appendix A6.10.7.

7.1.2.1 Simple Two-sided Sampling Plans

The procedure (*test plan*) for the *simple two-sided sampling plan* is as follows (Appendix A8.3.1.1):

1. From p_0, p_1, α, and β, determine the smallest integers c and n for which

$$\sum_{i=0}^{c} \binom{n}{i} p_0^i (1-p_0)^{n-i} \ge 1-\alpha \tag{7.8}$$

and

$$\sum_{i=0}^{c} \binom{n}{i} p_1^i (1-p_1)^{n-i} \le \beta . \tag{7.9}$$

2. Take a sample of size n, determine the number k of defective items in the sample, and

$$
\begin{aligned}
&\bullet \text{ reject } H_0 \ (p < p_0) \quad &&\text{if} \quad k > c\\
&\bullet \text{ accept } H_0 \ (p < p_0) \quad &&\text{if} \quad k \le c .
\end{aligned}
\tag{7.10}
$$

The graph of Fig. 7.2 helps to clarify the validity of the above rule (see Appendix A8.3.1.1 for a proof). It satisfies the inequalities (7.8) and (7.9), and is known as an *operating characteristic curve*. For each value of p, it gives the probability of having no more than c defective items in a sample of size n. Since the operating characteristic curve as a function of p decreases monotonically, the risk of a false decision decreases for $p < p_0$ and $p > p_1$, respectively. It can be shown that the quantities c and $n p_0$ depend only on α, β, and the ratio p_1/p_0 (*discrimination ratio*). Table 7.3 in Section 7.2.2.2 gives c and $n p_0$ for some important values of α, β and p_1/p_0 for the case where the *Poisson approximation* (Eq. (7.12)) applies.

Using the operating characteristic, the *Average Outgoing Quality (AOQ)* can be calculated. *AOQ* represents the percentage of defective items that reach the customer, assuming that all rejected samples have been 100% inspected, and that the defective items have been replaced by good ones, and is given by

$$AOQ = p\Pr\{\text{acceptance} \mid p\} = p\sum_{i=0}^{c}\binom{n}{i}p^i(1-p)^{n-i}. \tag{7.11}$$

The maximum value of AOQ is the *Average Outgoing Quality Limit* [7.4, 7.5].

Obtaining the solution of the inequalities (7.8) and (7.9) is time consuming. For small values of p_0 and p_1 (up to a few %), the *Poisson approximation*

$$\binom{n}{i}p^i(1-p)^{n-i} \approx \frac{(n\,p)^i}{i!}e^{-n\,p} \tag{7.12}$$

can be used (Eq. (A6.129)). Substituting the approximate value obtained by Eq. (7.12) in Eqs. (7.8) and (7.9) leads to a Poisson distribution with parameters $m_1 = n\,p_1$ and $m_0 = n\,p_0$, which can be solved using a table of the χ^2 distribution (Table A9.2). Alternatively, the curves of Fig. 7.3 provide graphical solutions, sufficiently good for many practical applications. Exact solutions for some important cases are given in Table 7.3 of Section 7.2.2.2.

Example 7.2

Determine the sample size n and the number of allowed defective items c to test the null hypothesis $H_0 : p < p_0 = 1\%$ against the alternative hypothesis $H_1 : p > p_1 = 2\%$ with producer and consumer risks $\alpha = \beta = 0.1$ (which means $\alpha \approx \beta \lessapprox 0.1$).

Solution

For $\alpha = \beta = 0.1$, Table A9.2 yields $\nu = 30$ (value of ν for which $t_{\nu,q_1} / t_{\nu,q_2} = 2$ with $q_1 \gtrapprox 1-\alpha = 0.9$ and $q_2 \lessapprox \beta = 0.1$) and, with a linear interpolation, $F(20.4) \approx 0.095 < \beta$ and $F(40.8) \approx 0.908 > 1-\alpha$ ($\nu = 28$ falls just short). Thus $c = \nu/2 - 1 = 14$ and $n = 20.4/(2\cdot 0.01) = 1020$. The values of c and n according to Table 7.3 would be $c = 14$ and $n = 10.17/0.01 = 1017$.

Using the graph of Fig. 7.3 yields practically the same result: $c = 14$, $m_0 \approx 10.2$ and $m_1 \approx 20.4$ for $\alpha \approx \beta \approx 0.1$. Both the analytical and graphical methods require a solution by successive approximation (choice of c and check of the conditions for α and β by considering the ratio p_1 / p_0).

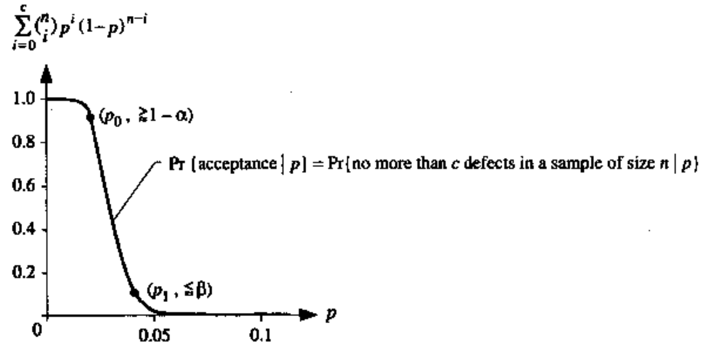


Figure 7.2 Operating characteristic curve as a function of the defective probability p for a given n and c ($p_0 = 2\%$, $p_1 = 4\%$, $\alpha \approx \beta \approx 0.1$; $n = 510$ and $c = 14$ as per Table 7.3)

7.1.2.2 Sequential Tests

The procedure for a *sequential test* is as follows (Appendix A8.3.1.2):

1. In a Cartesian coordinate system draw the *acceptance line*

$$y_1(n) = a\,n - b_1 \tag{7.13}$$

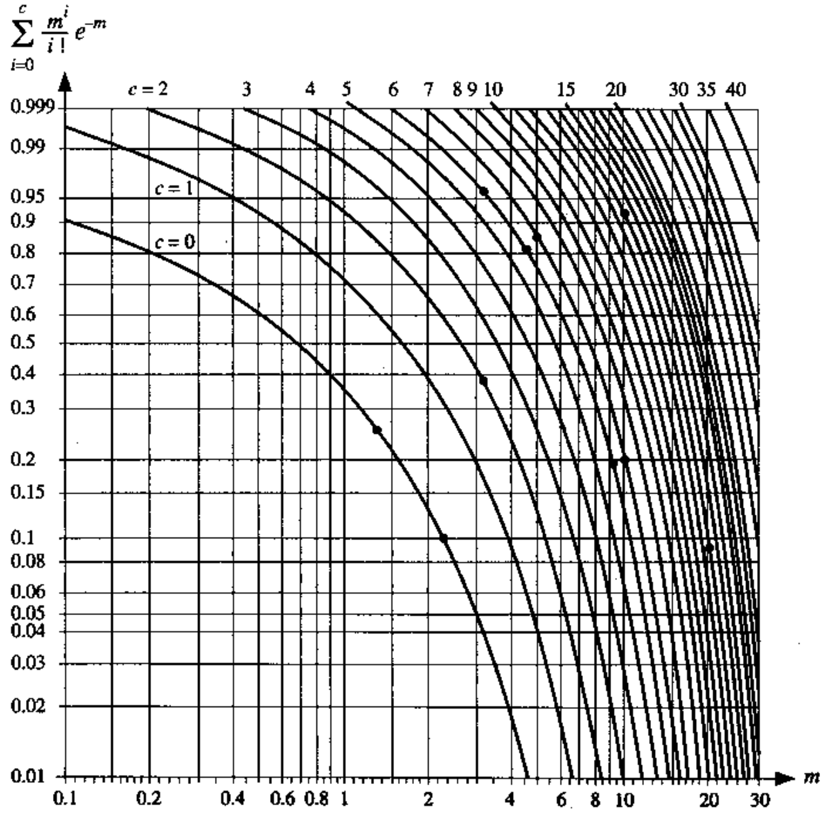


Figure 7.3 Poisson distribution (• results for Examples 7.2, 7.4, 7.5, 7.6, 7.9; $c = 14, 7, 0, 2$ & $0, 6$)

and the *rejection line*

$$y_2(n) = a n + b_2, \tag{7.14}$$

with

$$a = \frac{\ln \frac{1-p_0}{1-p_1}}{\ln \frac{p_1}{p_0} + \ln \frac{1-p_0}{1-p_1}}, \quad b_1 = \frac{\ln \frac{1-\alpha}{\beta}}{\ln \frac{p_1}{p_0} + \ln \frac{1-p_0}{1-p_1}}, \quad b_2 = \frac{\ln \frac{1-\beta}{\alpha}}{\ln \frac{p_1}{p_0} + \ln \frac{1-p_0}{1-p_1}}. \tag{7.15}$$

2. Select one item after another from the lot, test the item, enter the test result in the diagram drawn in step 1, and stop the test as soon as either the rejection or the acceptance line is crossed.

Figure A8.8 shows acceptance and rejection lines for $p_0 = 1\%$, $p_1 = 2\%$ and $\alpha \approx \beta \approx 0.2$. The advantage of the sequential test is that on *average* it requires a smaller sample size than the corresponding simple two-sided sampling plan (see Example 7.10 or Fig. 7.8). A disadvantage is that the test duration (i.e. the sample size involved) is random.

7.1.3 One-sided Sampling Plans for the Demonstration of a Defective Probability *p*

The two-sided sampling plans of Section 7.1.2 are fair in the sense that for $\alpha = \beta$, both producer and consumer run the *same risk* of making a false decision. In practical applications however, *one-sided sampling plans* are often used, i.e. only p_0 and α or p_1 and β are specified. In these cases, the operating characteristic curve is not completely defined. For every value of c ($c = 0, 1, \ldots$) a largest n ($n = 1, 2, \ldots$) exists which satisfies inequality (7.8) for a given p_0 and α, or a smallest n exists which satisfies inequality (7.9) for a given p_1 and β. It can be shown that the operating characteristic curves become steeper as the value of c increases (see e.g. Figs. 7.4 or A8.9). Hence, for small values of c, the producer (if p_0 and α are given) or the consumer (if p_1 and β are given) *can be favored.* For example, for a given p_0 and α, the operating characteristic curve remains close to $1-\alpha$ for $p > p_0$ for small values of c, and thus the producer can realize an advantage.

When only p_0 and α or p_1 and β are given, it is usual to set

$$p_0 = AQL \qquad \text{and} \qquad p_1 = LTPD, \tag{7.16}$$

where *AQL* is the *Acceptable Quality Level* and *LTPD* is the *Lot Tolerance Percent Defective* (Eqs. (A8.69) to (A8.74)).

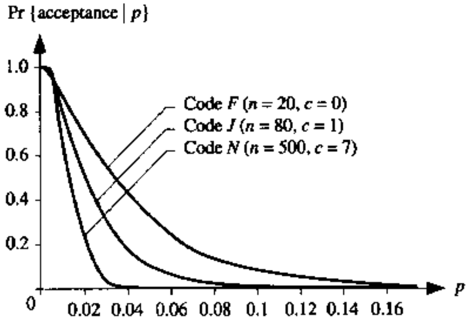


Figure 7.4 Operating characteristic curves for the demonstration of an $AQL = 0.65\%$ with sample sizes $n = 20$, 80, and 500 as per Table 7.1

A large number of one-sided sampling plans for the demonstration of *AQL* values are given in national and international standards (IEC 60410, ISO 2859, MIL-STD-105, DIN 40080 [7.2, 7.3]). Many of these plans have been established empirically. The following remarks can be useful in evaluating such plans:

1. *AQL* values are given in %.
2. The values for n and c are in general obtained using the *Poisson approximation.*
3. Not all values of c are listed, the value of α often decreases with increasing c.
4. Sample size is related to lot size, and this relationship is empirical.
5. A distinction is made between reduced tests (level I), normal tests (level II) and tightened tests (level III); level II is normally used.
6. Transition from level II to level III is necessary if 2 out of 5 successive independent lots have been rejected, a return to level II follows if 5 successive independent lots are passed (transitions between levels II and I are also defined).
7. The value of α is not given explicitly (for $c = 0$, for example, α is approximately 0.05 for level I, 0.1 for level II, and 0.2 for level III).

Table 7.1 presents some test procedures from IEC 60410 [7.2] and Fig. 7.4 shows the corresponding operating characteristic curves for $AQL = 0.65\%$ and sample sizes $n = 20, 80$, and 500.

Test procedures for the demonstration of *LTPD* values with a customer risk $\beta \approx 0.1$ are given in Table 7.2 (based on the Poisson approximation given by Eq. (7.12)).

In addition to the simple one-sided sampling plans described above, *multiple one-sided sampling plans* are often used to demonstrate *AQL* values. Fig. 7.5 shows the procedure for a *double one-sided sampling plan.* The operating characteristic curve or acceptance probability for this plan can be calculated as

$$\Pr\{\text{acceptance} \mid p\} = \sum_{i=0}^{c_1} \binom{n_1}{i} p^i (1-p)^{n_1-i} + \sum_{i=c_1+1}^{d_1-1} [\binom{n_1}{i} p^i (1-p)^{n_1-i} \sum_{j=0}^{c_2-i} \binom{n_2}{i} p^j (1-p)^{n_2-j}]. \qquad (7.17)$$

Multiple one-sided sampling plans are also given in national and international standards, see for example IEC 60410 [7.2] for the following double one-sided sampling plan to demonstrate $AQL = 1\%$

Sample Size	n_1	n_2	c_1	d_1	c_2
281 - 500	32	32	0	2	1
501 - 1,200	50	50	0	3	3
1,201 - 3,200	80	80	1	4	4
3,201 - 10,000	125	125	2	5	6

The advantage of multiple one-sided sampling plans is that on average they require smaller sample sizes than would be necessary for simple one-sided sampling plans.

Table 7.1 Test procedures for AQL demonstration (test level II, from IEC 60410 [7.2])

Code	Lot size	Sam-ple size	*AQL* in %											
			0.04	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5
	N	*n*	*c*	*c*	*c*	*c*	*c*	*c*	*c*	*c*	*c*	*c*	*c*	*c*
A	2 - 8	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0
B	9 - 15	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	0	↑
C	16 - 25	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	0	↑	↓
D	26 - 50	8	↓	↓	↓	↓	↓	↓	↓	↓	0	↑	↓	1
E	51 - 90	13	↓	↓	↓	↓	↓	↓	↓	0	↑	↓	1	2
F	91 - 150	20	↓	↓	↓	↓	↓	↓	0	↑	↓	1	2	3
G	151 - 280	32	↓	↓	↓	↓	↓	0	↑	↓	1	2	3	5
H	281 - 500	50	↓	↓	↓	↓	0	↑	↓	1	2	3	5	7
J	501 - 1200	80	↓	↓	↓	0	↑	↓	1	2	3	5	7	10
K	1.2k - 3.2k	125	↓	↓	0	↑	↓	1	2	3	5	7	10	14
L	3.2k - 10k	200	↓	0	↑	↓	1	2	3	5	7	10	14	21
M	10k - 35k	315	0	↑	↓	1	2	3	5	7	10	14	21	↑
N	35k - 150k	500	↑	↓	1	2	3	5	7	10	14	21	↑	↑
P	150k - 500k	800	↓	1	2	3	5	7	10	14	21	↑	↑	↑
Q	over 500k	1250	1	2	3	5	7	10	14	21	↑	↑	↑	↑

Use the first sampling plan above for ↑ or below for ↓, c = number of allowed defects

Table 7.2 Test procedures for *LTPD* demonstration (MIL-S-19500 [3.11])

Number of allowed defects	*LTPD* in %										
	1	1.5	2	3	5	7	10	15	20	30	50
c	*n*	*n*	*n*	*n*	*n*	*n*	*n*	*n*	*n*	*n*	*n*
0	231	153	116	76	45	32	22	15	11	8	5
1	390	258	195	129	77	55	38	25	18	13	8
2	533	354	266	176	105	75	52	34	25	18	11
3	668	444	333	221	132	94	65	43	32	22	13
4	798	531	398	265	158	113	78	52	38	27	16
5	927	617	462	308	184	131	91	60	45	31	19
6	1054	700	528	349	209	149	104	68	51	35	21
7	1178	783	589	390	234	166	116	77	57	39	24
8	1300	864	648	431	258	184	128	85	63	43	26
9	1421	945	709	471	282	201	140	93	69	47	28
10	1541	1025	770	511	306	218	152	100	75	51	31
11	1664	1109	832	555	330	238	166	111	83	54	33
12	1781	1187	890	594	356	254	178	119	89	59	36
13	1896	1264	948	632	379	271	190	126	95	63	38
14	2015	1343	1007	672	403	288	201	134	101	67	40
15	2133	1422	1066	711	426	305	213	142	107	71	43
16	2249	1499	1124	750	450	321	225	150	112	74	45
17	2364	1576	1182	788	473	338	236	158	118	79	47
18	2478	1652	1239	826	496	354	248	165	124	83	50
19	2591	1728	1296	864	518	370	259	173	130	86	52
20	2705	1803	1353	902	541	386	271	180	135	90	54
25	3259	2173	1629	1086	652	466	326	217	163	109	65

n = sample size, c = number of allowed defects in the sample of size n, β = 10%

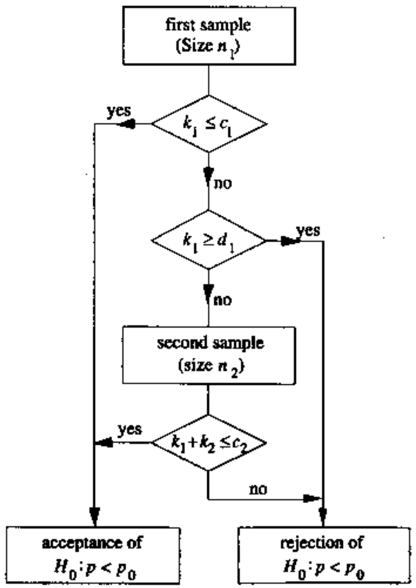


Figure 7.5 Flowchart for a double one-sided sampling plan

7.2 Statistical Reliability Tests

Reliability tests are necessary to evaluate the reliability actually *achieved* in a given item. Early initiation of such testing permits quick identification and cost-effective correction of weaknesses not discovered by reliability analyses. This supports a learning process which leads to a mature product. Since reliability tests are generally time consuming and expensive, they must be coordinated with other tests. Test conditions should be as close as possible to those experienced in the field (real-world tests). As with quality control, a distinction is made between *estimation* and *demonstration* of a specific reliability characteristic.

7.2.1 Estimation and Demonstration of a Reliability or Availability Value

Reliability and availability are defined as probabilities. Their estimation and demonstration can thus be performed using the methods described in Section 7.1 for an unknown defective probability p. If R is the reliability and PA the availability, it is helpful to set

$$p = 1 - R \tag{7.18}$$

or

$$p = 1 - PA. \tag{7.19}$$

For a demonstration (acceptance test), the null hypothesis $H_0 : p < p_0$ is then converted to $H_0 : R > R_0$ or $H_0 : PA > PA_0$, which corresponds better to the concept of reliability or availability. Obviously, the same also holds for *maintainability* and for any other reliability characteristic expressed as a *probability*. Thus, results and considerations from Section 7.1 apply (Examples 7.3 to 7.6).

Example 7.3

A reliability test examines 100 subassemblies, and 95 of them pass. Determine the confidence interval for the reliability R with a confidence level $\gamma = 0.9$ ($\beta_1 = \beta_2 = 0.05$).

Solution

With $p = 1 - R$ and $\hat{R} = 0.95$ the confidence interval for p follows from Fig. 7.1 as [0.03, 0.10]. The confidence interval for R is then [0.9, 0.97]. Calculation using Eq. (7.5) leads to the interval [0.025, 0.099] for p and [0.901, 0.975] for R.

Example 7.4

The reliability of a given subassembly was $R = 0.9$ and should have been improved through constructive measures. In a test of 100 subassemblies, 94 of them pass the test. Check with a type I error $\alpha = 20\%$ the hypothesis $H_0 : R > 0.95$.

Solution

For $p_0 = 1 - R_0 = 0.05$, $\alpha = 20\%$, and $n = 100$, Eq. 7.8 delivers $c = 7$ (see also the graphical solution from Fig. 7.3 with $m = n\,p_0 = 5$ and acceptance probability $\geq 1 - \alpha = 0.8$, yielding $\alpha \approx 0.15$ for $m = 5$ and $c = 7$). As just $k = 6$ subassemblies have failed the test, the hypothesis $H_0 : R > 0.95$ can be accepted at the level $1 - \alpha = 0.8$.
Supplementary result: Assuming as an alternative hypothesis $H_1 : R < 0.90$, or $p > p_1 = 0.1$, the type II error β can be computed from Eq. (7.9) with $c = 7$ and $n = 100$ or graphically from Fig. 7.3 with $m = n\,p_1 = 10$, yielding $\beta \approx 0.2$.

Example 7.5

Determine the minimum number of tests n that must be repeated to verify the hypothesis $H_0 : R > R_1 = 0.95$ with a consumer risk $\beta = 0.1$. What is the allowed number of failures c?

Solution

The inequality (7.9) must be fulfilled with $p_1 = 1 - R_1 = 0.05$ and $\beta = 0.1$, n and c must thus satisfy

$$\sum_{i=0}^{c} \binom{n}{i} 0.05^i \cdot 0.95^{n-i} \leq 0.1.$$

The number of tests n is a minimum for $c = 0$. From $0.95^n \approx 0.1$, it follows that $n = 45$ (calculation with the Poisson approximation (Eq. (7.12)) yields $n = 46$, graphical solution with Fig. 7.3 leads to $m \approx 2.3$ and then $n = m / p \approx 46$).

Example 7.6

Continuing with Example 7.5, (i) find n for $c = 2$ and (ii) how large would the producer risk be for $c = 0$ and $c = 2$ if the true reliability were $R = 0.97$?

Solution

(i) From Eq. (7.9),

$$\sum_{i=0}^{2} \binom{n}{i} 0.05^i \cdot 0.95^{n-i} \lessapprox 0.1$$

and thus $n = 105$ (Fig. 7.3 yields $m \approx 5.3$ and $n \approx 106$; from Table A9.2, $\nu = 6$, $t_{2,0.9} = 10.645$ and $n = 107$).

(ii) The producer risk is

$$\alpha = 1 - \sum_{i=0}^{c} \binom{n}{i} 0.03^i \cdot 0.97^{n-i},$$

hence, $\alpha \approx 0.75$ for $c = 0$ and $n = 45$, $\alpha \approx 0.61$ for $c = 2$ and $n = 105$ (Fig. 7.3, yields $\alpha \approx 0.75$ for $c = 0$ and $m = 1.35$, $\alpha \approx 0.62$ for $c = 2$ and $m = 3.15$; from Table A9.2, $\alpha \approx 0.73$ for $\nu = 2$ and $t_{2,\alpha} = 2.7$, $\alpha \approx 0.61$ for $\nu = 6$ and $t_{6,\alpha} = 6.3$).

7.2.2 Estimation and Demonstration of a Constant Failure Rate λ or of *MTBF* = 1/λ

MTBF is the mean (expected value) of the failure-free operating time of an item exhibiting a *constant failure rate* λ. For such an item, $R(t)=e^{-\lambda t}$ and $MTBF = 1/\lambda$ (Eq. (A6.84)). This section discusses the estimation and demonstration of λ or of $MTBF = 1/\lambda$ (see Appendices A8.2 and A8.3 for basic considerations and A8.1 for empirical and graphical methods). In particular, the case of a given, *fixed, cumulative operating time T* (*Type I censoring*) is considered when repair times are ignored (immediate renewal) and individual failure-free operating times are assumed to be *independent*. Due to the relationship between exponentially distributed failure-free operating times and the (homogeneous) *Poisson process* (Eq. (A7.38)) as well as the *additive property of Poisson processes* (Example 7.7), the fixed *cumulative operating time T* can be partitioned in a quite *arbitrary way* from failure-free operating times of (s-) identical items. The following are some examples:

1. Operation of a single item that is immediately renewed after every failure (renewal time = 0), here $T = t$ = calendar time.
2. Operation of m identical items, each of them being immediately renewed after every failure (renewal time = 0), here

$$T = mt, \qquad m = 1, 2, \ldots .$$

As stated above, in the case of a constant failure rate λ, the failure process is a (homogeneous) *Poisson process* with intensity λ (in the case $m=1$) over the fixed time interval $[0, T]$. Hence, the probability of k failures occurring within the cumulative operating time T is given by (Eq. (A7.39))

$$\Pr\{k \text{ failures within } T \mid \lambda\} = \frac{(\lambda T)^k}{k!} e^{-\lambda T}. \tag{7.20}$$

Statistical procedures for the estimation and demonstration of a failure rate λ or of $MTBF = 1/\lambda$ can thus be based on the statistical evaluation of the parameter $(m = \lambda T)$ of a Poisson distribution (see applications in Sections 7.2.2.1 to 7.2.2.3).

In addition to the case of a given fixed *cumulative operating time T* and immediate renewal, further possibilities are known. Assuming m identical items at $t = 0$, and labeling the *individual failure times* as $t_1 < t_2 < \ldots$, the following cases are important for practical applications :

1. *Fixed number k of failures*, the test is stopped at the k-th failure (*Type II censoring*) and failed items are *instantaneously renewed* (renewal time = 0); an unbiased point estimate of the failure rate λ is (for $k>1$)

$$\hat{\lambda} = \frac{k-1}{m t_k}. \tag{7.21}$$

2. *Fixed number k of failures*, the test is stopped at the k-th failure (*Type II censor.*), failed items are *not renewed*; an unbiased point estimate of the failure rate λ is

$$\begin{aligned}\hat{\lambda} &= (k-1)/[m t_1 + (m-1)(t_2 - t_1) + \ldots + (m-k+1)(t_k - t_{k-1})] \\ &= (k-1)/[t_1 + \ldots + t_k + (m-k)t_k].\end{aligned} \tag{7.22}$$

3. *Fixed test time t* (*Type I censoring*), failed items are *not renewed*; a (biased) point estimate of the failure rate λ (given k items have failed) is

$$\begin{aligned}\hat{\lambda} &= k/[m t_1 + (m-1)(t_2 - t_1) + \ldots + (m-k)(t - t_k)] \\ &= k/[t_1 + \ldots + t_k + (m-k)t].\end{aligned} \tag{7.23}$$

4. *Fixed test time t* (*Type I censoring*), failed items are *not renewed* and only the total number k of failures is known (not the individual failure times t_i); a good approximation for the point estimate of the failure rate λ is (for $k << m$)

$$\hat{\lambda} \approx \frac{k}{mt}\left(1 + \frac{k}{2m}\right). \tag{7.24}$$

Example 7.7

An item with constant failure rate λ operates first for a fixed time period T_1 and then for a fixed time period T_2. Repair times are neglected. Find the probability that k failures will occur in the time period $T = T_1 + T_2$.

Solution

The item's behavior within each of the time periods T_1 and T_2 can be described by a (homogeneous) Poisson process with intensity λ. From Eq. (A7.39) follows that

$$\Pr\{i \text{ failures in the time period } T_1 \mid \lambda\} = \frac{(\lambda T_1)^i}{i!} e^{-\lambda T_1}$$

and, because of the *memoryless property* of the Poisson process

$$\begin{aligned}\Pr\{k \text{ failures in } T = T_1 + T_2 \mid \lambda\} &= \sum_{i=0}^{k} \frac{(\lambda T_1)^i}{i!} e^{-\lambda T_1} \cdot \frac{(\lambda T_2)^{k-i}}{(k-i)!} e^{-\lambda T_2} \\ &= e^{-\lambda T} \sum_{i=0}^{k} \lambda^k \frac{T_1^i}{i!} \cdot \frac{T_2^{k-i}}{(k-i)!} = \frac{(\lambda T)^k}{k!} e^{-\lambda T}.\end{aligned} \tag{7.25}$$

The last part of Eq. (7.25) follows from the binomial expansion of $(T_1 + T_2)^k$. Example 7.7 shows that the *cumulative operating time T* can be partitioned in any *arbitrary way*. Necessary and sufficient is that the failure rate λ be *constant*. The same procedure can be used to prove that the *sum of two independent Poisson processes* with intensities λ_1 and λ_2 is a *Poisson process* with intensity $\lambda_1 + \lambda_2$

$$\begin{aligned}&\Pr\{k \text{ failures in } (0,T] \mid \lambda_1, \lambda_2\} \\ &\quad = \sum_{i=0}^{k} \frac{(\lambda_1 T)^i}{i!} e^{-\lambda_1 T} \frac{(\lambda_2 T)^{k-i}}{(k-i)!} e^{-\lambda_2 T} = \frac{((\lambda_1+\lambda_2)T)^k}{k!} e^{-(\lambda_1+\lambda_2)T}.\end{aligned} \tag{7.26}$$

7.2.2.1 Estimation of a constant Failure Rate λ or of $MTBF = 1/\lambda$ *)

Let us consider an item with a constant failure rate λ. When during the *given* (fixed) *cumulative operating time* T, k failures have occurred, the maximum likelihood *point estimate* for the unknown parameter λ or $MTBF = 1/\lambda$ follows from Eq. (A8.46) as

$$\hat{\lambda} = \frac{k}{T} \qquad \text{or} \qquad M\hat{T}BF = \frac{T}{k}. \tag{7.27}$$

For a given *confidence level* $\gamma = 1 - \beta_1 - \beta_2$ $(0 < \beta_1 < 1 - \beta_2 < 1$, i.e. $\beta_1, \beta_2 > 0$ and $\beta_1 + \beta_2 < 1)$ and $k > 0$, the lower $\hat{\lambda}_l$ and upper $\hat{\lambda}_u$ limits of the *confidence interval* for the failure rate λ can be obtained from (Eqs. (A8.47) to (A8.51))

$$\sum_{i=k}^{\infty} \frac{(\hat{\lambda}_l T)^i}{i!} e^{-\hat{\lambda}_l T} = \beta_2 \qquad \text{and} \qquad \sum_{i=0}^{k} \frac{(\hat{\lambda}_u T)^i}{i!} e^{-\hat{\lambda}_u T} = \beta_1, \tag{7.28}$$

or from the quantile of the χ^2 distribution (Table A9.2) as

$$\hat{\lambda}_l = \frac{\chi^2_{2k,\beta_2}}{2T} \qquad \text{and} \qquad \hat{\lambda}_u = \frac{\chi^2_{2(k+1),1-\beta_1}}{2T}, \tag{7.29}$$

the corresponding limits for $MTBF = 1/\lambda$ being

$$M\hat{T}BF_l = \frac{2T}{\chi^2_{2(k+1),1-\beta_1}} \qquad \text{and} \qquad M\hat{T}BF_u = \frac{2T}{\chi^2_{2k,\beta_2}}.$$

For $k = 0$, Eq. (A8.49) yields

$$\hat{\lambda}_l = 0 \qquad \text{and} \qquad \hat{\lambda}_u = \frac{\ln(1/\beta_1)}{T}, \tag{7.30}$$

or

$$M\hat{T}BF_l = \frac{T}{\ln(1/\beta_1)} \qquad \text{and} \qquad M\hat{T}BF_u = \infty,$$

the confidence level is here $\gamma = 1 - \beta_1$.

For many practical applications, a graphical solution is often sufficiently good. Figure 7.6 gives the relationship between γ, k, T, $\hat{\lambda}_l$, and $\hat{\lambda}_u$ for the case $\beta_1 = \beta_2 = (1-\gamma)/2$, allowing also the visualisation of the confidence interval width $(\hat{\lambda}_u - \hat{\lambda}_l)$ as function of the number k of failures.

---

*) The case considered in Sections 7.2.2.1 to 7.2.2.3 corresponds to a sampling plan with n elements *with replacement* and k failures in the *fixed* (given) time interval $[0, T/n]$, *Type I (time) censoring*; the underlying process is a (hom.) Poisson process with constant intensity $n\lambda$.

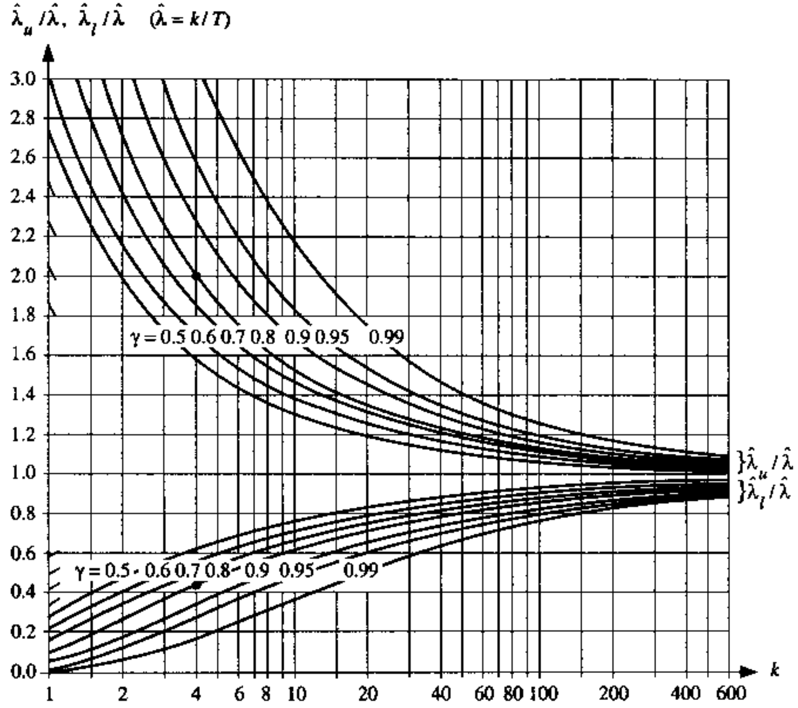


Figure 7.6 Confidence limits $\hat{\lambda}_l$ and $\hat{\lambda}_u$ for an unknown constant failure rate λ or $MTBF = 1/\lambda$ (T=given (fixed) cumulative operating time (time censoring), k = number of failures during T, γ = confidence level $=1-\beta_1-\beta_2$ (here with $\beta_1 = \beta_2$), $M\hat{T}BF_l = 1/\hat{\lambda}_u$ and $M\hat{T}BF_u = 1/\hat{\lambda}_l$
• results for Examples 7.8 and 7.13)

The confidence limits $\hat{\lambda}_l$ and $\hat{\lambda}_u$ or $M\hat{T}BF_l$ and $M\hat{T}BF_u$ can also be used to give *one-sided confidence intervals*. In this case

$$\lambda \le \hat{\lambda}_u \quad \text{or} \quad MTBF \ge M\hat{T}BF_l, \qquad \text{with} \quad \beta_2 = 0 \quad \text{and} \quad \gamma = 1 - \beta_1,$$

or

$$\lambda \ge \hat{\lambda}_l \quad \text{or} \quad MTBF \le M\hat{T}BF_u, \qquad \text{with} \quad \beta_1 = 0 \quad \text{and} \quad \gamma = 1 - \beta_2. \tag{7.31}$$

Example 7.8

In testing a subassembly with constant failure rate λ, 4 failures occur during $T = 10^4$ cumulative operating hours. Find the confidence interval of λ for a confidence level $\gamma = 0.8$ $(\beta_1 = \beta_2 = 0.1)$.

Solution

From Fig. 7.6 it follows that for $k = 4$ and $\gamma = 0.8$, $\hat{\lambda}_l / \hat{\lambda} \approx 0.43$ and $\hat{\lambda}_u / \hat{\lambda} \approx 2$. With $T = 10^4$ h, $k = 4$, and $\hat{\lambda} = 4 \cdot 10^{-4}\,\text{h}^{-1}$, the confidence limits are $\hat{\lambda}_l \approx 1.7 \cdot 10^{-4}\,\text{h}^{-1}$ and $\hat{\lambda}_u \approx 8 \cdot 10^{-4}\,\text{h}^{-1}$. The corresponding limits for the $MTBF = 1/\lambda$ are $\hat{MTBF}_l \approx 1250$ h and $\hat{MTBF}_u \approx 5814$ h.
Supplementary result: The above results would also lead to the one-sided confidence interval $\lambda \le 8 \cdot 10^{-4}\,h^{-1}$ or $MTBF \ge 1250$ h with $\gamma = 0.9$.

In the above considerations (Eqs. (7.27) to (7.31)), the *cumulative operating time* T was fixed (given), independent of the individual failure-free operating times and of the number m of items involved (*Type I censoring*). The situation is different when *the number of failures* k is fixed, i.e. when the test is stopped at the occurrence of the k-th failure (*Type II censoring*). Here, the cumulative operating time is a *random variable*, given by the term $(k-1)/\hat{\lambda}$ of Eqs. (7.21) and (7.22). Using the memoryless property of (homogeneous) Poisson processes, it can be shown that the quantities

$$m\,(t_i^* - t_{i-1}^*) \text{ for renewal, } \quad \text{and} \quad (m-i+1)\,(t_i^* - t_{i-1}^*) \text{ for no renewal,} \qquad (7.32)$$

with $i = 1, \ldots, k$ and $t_0^* = 0$, are independent observations of a random variable distributed according to $F(t) = 1 - e^{-\lambda t}$. This is necessary and sufficient to prove that the $\hat{\lambda}$ given by Eqs. (7.21) and (7.22) are maximum likelihood estimates for λ. For confidence intervals, results of Appendix A8.2.2.3 can be used.

7.2.2.2 Simple Two-sided Test for the Demonstration of a Constant Failure Rate λ or of *MTBF* = 1/λ

In the context of an acceptance test, the demonstration of λ or of $MTBF = 1/\lambda$ is generally required, and not merely its estimation as in Section 7.2.2.1. The main concern of this test is to check a zero hypothesis $H_0 : \lambda < \lambda_0$ or $MTBF > MTBF_0$ against an alternative hypothesis $H_1 : \lambda > \lambda_1$ or $MTBF < MTBF_1$, on the basis of the following agreement between producer and consumer:

> *Items should be accepted with a probability nearly equal to (but not less than)* $1-\alpha$, *if the true (unknown) MTBF is greater than* $MTBF_0$, *but rejected with a probability nearly equal to (but not less than)* $1-\beta$, *if the MTBF is less than* $MTBF_1$ ($MTBF_0 > MTBF_1$ *are given fixed values).*

$MTBF_0$ is the *specified MTBF* and $MTBF_1$ is the *minimum acceptable MTBF* (m_0 and m_1 in IEC 60605 [7.12], or θ_0 and θ_1 in MIL-STD-781, [7.15]). α is the producer's risk (type I error), i.e. the probability of rejecting a true hypothesis $H_0 : MTBF > MTBF_0$). β is the corresponding consumer's risk (type II error), i.e. the probability of accepting H_0 although the alternative hypothesis $H_1 : MTBF < MTBF_1$ is true. In the following it will tacitly be assumed that $0 < \alpha < 1 - \beta < 1$. Evaluation of the above agreement is a problem of statistical hypothesis testing (Appendix A8.3), and can be performed for instance with a *simple two-sided test* or a *sequential test*. The methods and results given here are also valid for testing the hypothesis $H_0 : \lambda < \lambda_0 = 1/MTBF_0$ against the alternative hypothesis $H_1 : \lambda > \lambda_1 = 1/MTBF_1$.

With the *simple two-sided test* (also known as the *fixed length test*), the cumulative operating time T and the number of allowed failures c during T are fixed quantities. The procedure (test plan) is as follows:

1. From $MTBF_0$, $MTBF_1$, α, and β determine the smallest integer c and the value of T for which

$$\sum_{i=0}^{c} \frac{(T/MTBF_0)^i}{i!} e^{-T/MTBF_0} \ge 1-\alpha \qquad (7.33)$$

and

$$\sum_{i=0}^{c} \frac{(T/MTBF_1)^i}{i!} e^{-T/MTBF_1} \le \beta. \qquad (7.34)$$

2. Perform a test with a total *cumulative operating time* T, determine the number of failures k during the test, and

   - reject H_0 : $MTBF > MTBF_0$ if $k > c$
   - accept H_0 : $MTBF > MTBF_0$ if $k \le c$. (7.35)

Example 7.9

The following conditions have been specified for the *MTBF* demonstration (acceptance test) of an assembly: $MTBF_0 = 2000$ h (specified *MTBF*), $MTBF_1 = 1000$ h (minimum acceptable *MTBF*), producer risk $\alpha = 0.2$, consumer risk $\beta = 0.2$. Determine: (i) the cumulative test time T and the allowed number of failures c during T; (ii) the probability of acceptance, if the true *MTBF* were 3000 h.

Solution

(i) From Fig. 7.3, $c = 6$ and $m \approx 4.6$ for Pr{acceptance} ≈ 0.82, $c = 6$ and $m \approx 9.2$ for Pr{acceptance} ≈ 0.19; thus $c = 6$ and $T = 9200$ h. These values agree well with those obtained from Table A9.2 ($\nu = 14$) and are given in Table 7.3. (ii) For $MTBF = 3000$ h, $T = 9200$ h, and $c = 6$

$$\Pr\{\text{acceptance} \mid MTBF = 3000\text{ h}\}$$
$$= \Pr\{\text{no more than 6 failures in } T = 9200\text{ h} \mid MTBF = 3000\text{ h}\} = \sum_{i=0}^{6} \frac{3.07^i}{i!} e^{-3.07} \approx 0.96,$$

see also Fig. 7.3 for a graphical solution.

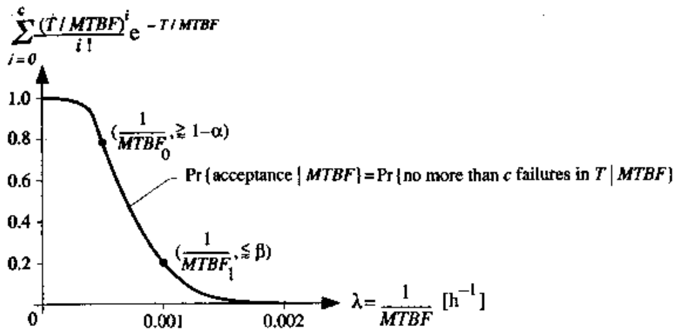


Figure 7.7 Operating characteristic curve (acceptance probability curve) as a function of $\lambda = 1/MTBF$ for fixed T and c ($MTBF_0 = 2000$ h, $MTBF_1 = 1000$ h, $\alpha \approx \beta \approx 0.2$; $T = 9200$ h and $c = 6$ as in Table 7.3)

The graph of Fig. 7.7 clarifies the validity of the above agreement between producer and consumer (customer). It satisfies the inequalities (7.33) and (7.34), and is known as the *operating characteristic curve* (acceptance probability curve). For each value of $\lambda = 1/MTBF$, it gives the probability of having not more than c failures during a cumulative operating time T. Since the operating characteristic curve as a function of $\lambda = 1/MTBF$ is monotonically decreasing, the risk of a false decision decreases for $MTBF > MTBF_0$ and $MTBF < MTBF_1$, respectively. It can be shown that the quantities c and $T/MTBF_0$ depend only on α, β, and the ratio $MTBF_0/MTBF_1$ (*discrimination ratio*).

Table 7.3 gives c and $T/MTBF_0$ for some important values of α, β and $MTBF_0/MTBF_1$. With $MTBF_0/MTBF_1 = p_1/p_0$ and $T/MTBF_0 = n\,p_0$, Table 7.3 can also be used for the demonstration of an *unknown probability* p (Eqs. (7.8) and (7.9)) in the case where the *Poisson approximation* (Eq. (7.12)) applies.

In addition to the simple two-sided test described above, a *sequential test* is often used (see Appendix A8.3.1.2 and Section 7.1.2.2 for basic considerations and Fig. 7.8 for an example). In this test, neither the cumulative operating time T, nor the number c of allowed failures during T are specified before the test begins. The number of failures is recorded as a function of the cumulative operating time (normalized to $MTBF_0$). As soon as the resulting staircase curve crosses the *acceptance line* or the *rejection line* the test is stopped. Sequential tests offer the advantage that on average the test duration is shorter than with simple two-sided tests. Using Eqs. (7.13) to (7.15) with $p_0 = 1 - e^{-\delta t/MTBF_0}$, $p_1 = 1 - e^{-\delta t/MTBF_1}$, $n = T/\delta t$, and $\delta \to 0$ (continuous in time), the acceptance and rejection lines are obtained as

Table 7.3 Number of allowed failures c during the cumulative operating time T and value of $T/MTBF_0$ to demonstrate $MTBF > MTBF_0$ against $MTBF_0 < MTBF_1$ for various values of α, β, and $MTBF_0/MTBF_1$ (can also be used for demonstration of a constant failure rate $\lambda = 1/MTBF$ or of a probability p by setting $MTBF_0/MTBF_1 = p_1/p_0$ and $T/MTBF_0 = n\,p_0$))

	$\frac{MTBF_0}{MTBF_1} = 1.5$	$\frac{MTBF_0}{MTBF_1} = 2$	$\frac{MTBF_0}{MTBF_1} = 3$
$\alpha \approx \beta \leq 0.1$	$c = 40$ $T/MTBF_0 \approx 32.98$ ($\alpha \approx \beta \approx 0.098$)	$c = 14^*$ $T/MTBF_0 \approx 10.17$ ($\alpha \approx \beta \approx 0.093$)	$c = 5$ $T/MTBF_0 \approx 3.12$ ($\alpha \approx \beta \approx 0.096$)
$\alpha \approx \beta \leq 0.2$	$c = 17$ $T/MTBF_0 \approx 14.33$ ($\alpha \approx \beta \approx 0.197$)	$c = 6$ $T/MTBF_0 \approx 4.62$ ($\alpha \approx \beta \approx 0.185$)	$c = 2$ $T/MTBF_0 \approx 1.47$ ($\alpha \approx \beta \approx 0.184$)
$\alpha \approx \beta \leq 0.3$	$c = 6$ $T/MTBF_0 \approx 5.41$ ($\alpha \approx \beta \approx 0.2997$)	$c = 2$ $T/MTBF_0 \approx 1.85$ ($\alpha \approx \beta \approx 0.284$)	$c = 1$ $T/MTBF_0 \approx 0.92$ ($\alpha \approx \beta \approx 0.236$)

* $c = 13$ yields $T/MTBF_0 = 9.48$ and $\alpha \approx \beta \approx 0.1003$

- acceptance line: $y_1(x) = a\,x - b_1$, (7.36)
- rejection line: $y_2(x) = a\,x + b_2$, (7.37)

with $x = T/MTBF_0$, and

$$a = \frac{\frac{MTBF_0}{MTBF_1} - 1}{\ln\frac{MTBF_0}{MTBF_1}}, \qquad b_1 = \frac{\ln\frac{1-\alpha}{\beta}}{\ln\frac{MTBF_0}{MTBF_1}}, \qquad b_2 = \frac{\ln\frac{1-\beta}{\alpha}}{\ln\frac{MTBF_0}{MTBF_1}}. \tag{7.38}$$

Sequential tests used in practical applications are given in national and international standards [7.12, 7.15]. To limit testing effort, restrictions are often placed on the test duration and the number of allowed failures. Figure 7.8 shows two *truncated* sequential test plans for $\alpha \approx \beta \approx 0.2$ and $MTBF_0/MTBF_1 = 1.5$ and 2, respectively. The lines defined by Eqs. (7.36) to (7.38) are shown dashed in Fig. 7.8a.

Example 7.10

Continuing with Example 7.9, determine the expected test duration given that the true $MTBF$ equals $MTBF_0$ and a sequential test as per Fig. 7.8 is used.

Solution

From Fig. 7.8 with $MTBF_0/MTBF_1 = 2$ it follows that E[test duration | $MTBF = MTBF_0$] $\approx 2.4\,MTBF_0 = 4800$ h.

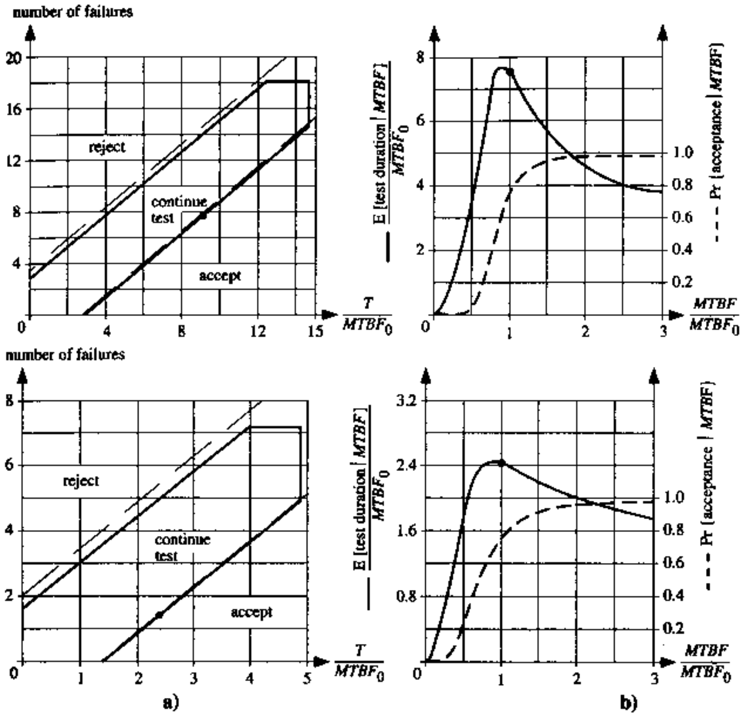


Figure 7.8 a) Sequential test plan to demonstrate $MTBF > MTBF_0$ against $MTBF < MTBF_1$ for $\alpha \approx \beta \approx 0.2$ and $MTBF_0 / MTBF_1 = 1.5$ (top) and 2 (down), as per IEC 60605 and MIL-HDBK-781 [7.12, 7.15], the lines as per Eqs. (7.36) to (7.38) are shown dashed; b) Expected test duration until acceptance and operating characteristic curve (dashed) as a function of $MTBF / MTBF_0$

7.2.2.3 Simple One-sided Test for the Demonstration of a Constant Failure Rate λ or of $MTBF = 1/\lambda$

Simple two-sided tests (Fig. 7.7) and sequential tests (Fig. 7.8) have the advantage that, for $\alpha = \beta$, producer and consumer run the same risk of making a false decision. In practical applications, often only $MTBF_0$ and α or only $MTBF_1$ and β, i.e. *simple one-sided tests*, are used. The considerations of Section 7.1.3 also apply here. Care should be taken with small values of c, as operating with only $MTBF_0$ and α (or $MTBF_1$ and β) the producer (or the consumer) can be favored. Figure 7.9 shows the operating characteristic curves for various values of c as a function of $1/MTBF$ for the demonstration of the hypotesis H_1: $MTBF < 1000\text{h}$ (against H_0: $MTBF > 1000\text{h}$) with a type II error $\beta \approx 0.2$ for $MTBF = 1000\text{h}$.

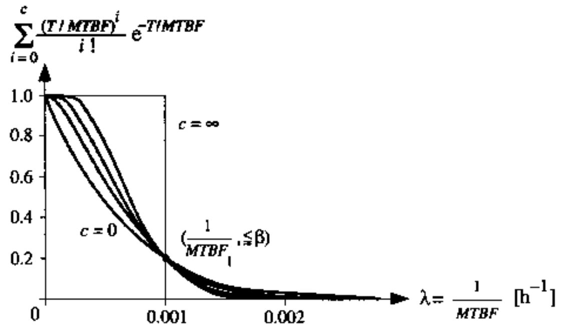


Figure 7.9 Operating characteristic curves (acceptance probability curves) for $MTBF_1 = 1000\text{h}$, $\beta = 0.2$, and $c = 0$ ($T = 1610\text{h}$), $c = 1$ ($T = 2995\text{h}$), $c = 2$ ($T = 4280\text{h}$), $c = 5$ ($T = 7905\text{h}$), and $c = \infty$ ($T = \infty$)

7.3 Statistical Maintainability Tests

Maintainability is generally expressed as a *probability*. In this case, the results of Sections 7.1 and 7.2.1 can be used directly to estimate or demonstrate maintainability. However, estimation and demonstration of specific parameters, like *MTTPM* (Mean Time To Preventive Maintenance) or *MTTR* (Mean Time To Repair), is of particular importance in many practical applications. If the underlying random values are exponentially distributed, the results of Section 7.2.2 for a *constant failure rate* λ or $MTBF = 1/\lambda$ can be used. This section deals with the estimation and demonstration of an *MTTR* by assuming that repair times are *lognormally* distributed and that the collected data are obtained from a *representative sample*.

7.3.1 Estimation of an *MTTR*

Let $t_1, \ldots, t_n$ be independent observations (realizations) of the repair time τ' of a given item. From Eqs. (A8.6) and (A8.10), the *empirical mean* and *variance* of τ' are given by

$$\hat{E}[\tau'] = \frac{1}{n}\sum_{i=1}^{n} t_i, \tag{7.39}$$

$$\hat{\mathrm{Var}}[\tau'] = \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \hat{\mathrm{E}}[\tau'])^2 = \frac{1}{n-1}[\sum_{i=1}^{n} t_i^2 - \frac{1}{n}(\sum_{i=1}^{n} t_i)^2]. \qquad (7.40)$$

For these estimates it holds that, $\mathrm{E}[\hat{\mathrm{E}}[\tau']] = \mathrm{E}[\tau'] = MTTR$, $\mathrm{Var}[\hat{\mathrm{E}}[\tau']] = \mathrm{Var}[\tau']/n$, and $\mathrm{E}[\hat{\mathrm{Var}}[\tau']] = \mathrm{Var}[\tau']$. As stated above, the repair time τ' is assumed to be lognormally distributed with a distribution function (Eq. (A6.110))

$$\mathrm{F}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln(\lambda t)}{\sigma}} e^{-\frac{x^2}{2}} dx, \qquad (7.41)$$

and with mean and variance given by (Eqs. (A6.112) and (A6.113))

$$\mathrm{E}[\tau'] = MTTR = \frac{e^{\sigma^2/2}}{\lambda}, \qquad \mathrm{Var}[\tau'] = \frac{e^{2\sigma^2} - e^{\sigma^2}}{\lambda^2} = MTTR^2(e^{\sigma^2} - 1). \qquad (7.42)$$

Using Eqs. (A8.24) and (A8.27), the *Maximum-Likelihood* estimation of λ and σ^2 is obtained from

$$\hat{\lambda} = [\prod_{i=1}^{n} \frac{1}{t_i}]^{\frac{1}{n}} \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[\ln(\hat{\lambda} t_i)]^2. \qquad (7.43)$$

A *point estimate* for λ and σ can also be obtained by the method of quantiles. The idea is to substitute some particular quantiles with the corresponding empirical quantiles to obtain estimates for λ or σ. For $t = 1/\lambda$, $\ln(\lambda t) = 0$ and $\mathrm{F}(1/\lambda) = 0.5$, therefore, $1/\lambda$ is the 0.5 quantile (median) $t_{0.5}$ of the distribution function $\mathrm{F}(t)$ given by Eq. (7.41). From the empirical 0.5 quantile $\hat{t}_{0.5} = \inf(t : \hat{\mathrm{F}}_n(t) \ge 0.5)$ an estimate for λ follows as

$$\hat{\lambda} = \frac{1}{\hat{t}_{0.5}}. \qquad (7.44)$$

Moreover, $t = e^{\sigma}/\lambda$ yields $\mathrm{F}(e^{\sigma}/\lambda) = 0.8413$; thus $e^{\sigma}/\lambda = t_{0.8413}$ is the 0.8413 quantile of $\mathrm{F}(t)$ given by Eq. (7.41). Using $\lambda = 1/\hat{t}_{0.5}$ and $\sigma = \ln(\lambda t_{0.8413}) = \ln(t_{0.8413}/t_{0.5})$, an estimate for σ is obtained as

$$\hat{\sigma} = \ln(\hat{t}_{0.8413}/\hat{t}_{0.5}). \qquad (7.45)$$

Considering $\mathrm{F}(e^{-\sigma}/\lambda) = 1 - 0.8413 = 0.1587$, i.e. $t_{0.1587} = e^{-\sigma}/\lambda$, one has $e^{2\sigma} = \lambda t_{0.8413}/\lambda t_{0.1587}$ and thus Eq. (7.45) can be replaced by

$$\hat{\sigma} = \frac{1}{2}\ln(\hat{t}_{0.8413}/\hat{t}_{0.1587}). \qquad (7.46)$$

The possibility of representing a lognormal distribution function as a straight line, to simplify the interpretation of data, is discussed in Section 7.5.1 (Fig. 7.14, Appendix A9.8.1).

To obtain *interval estimates* for the parameters λ and σ, note that the logarithm of a log normally distributed variable is normally distributed with mean $\ln(1/\lambda)$ and variance σ^2. Applying the transformation $t_i \to \ln t_i$ to the individual observations $t_1, \ldots, t_n$ and using the results known for the interval estimation of the parameters of a normal distribution [A6.1, A6.4], the confidence intervals

$$[\frac{n\hat{\sigma}^2}{\chi^2_{n-1,\frac{1+\gamma}{2}}}, \frac{n\hat{\sigma}^2}{\chi^2_{n-1,\frac{1-\gamma}{2}}}] \qquad (7.47)$$

for σ^2, and

$$[\hat{\lambda} e^{-\varepsilon}, \hat{\lambda} e^{\varepsilon}] \qquad \text{with} \qquad \varepsilon = \frac{\hat{\sigma}}{\sqrt{n-1}} t_{n-1,\frac{1+\gamma}{2}} \qquad (7.48)$$

for λ can be found with $\hat{\lambda}$ and $\hat{\sigma}$ as in Eq. (7.43). $\chi^2_{n-1,q}$ and $t^2_{n-1,q}$ are the q quantiles of the χ^2- and t-distribution with $n-1$ degrees of freedom, respectively (Tables A9.2 and A9.3).

Example 7.11

Let 1.1, 1.3, 1.6, 1.9, 2.0, 2.3, 2.4, 2.7, 3.1, and 4.2 h be 10 independent observations (realizations) of a lognormally distributed repair time. Determine the maximum likelihood estimate and, for $\gamma = 0.9$, the confidence interval for the parameters λ and σ^2, as well as the maximum likelihood estimate for *MTTR*.

Solution

Equation (7.43) yields $\hat{\lambda} \approx 0.476\,\mathrm{h}^{-1}$ and $\hat{\sigma}^2 \approx 0.146$ as maximum likelihood estimates of λ and σ^2. From Eq. (7.42), $\hat{MTTR} \approx e^{0.073}/0.476\,\mathrm{h}^{-1} \approx 2.26\,\mathrm{h}$. Using Eqs. (7.47) and (7.48), as well as Tables A9.2 and A9.3, the confidence intervals are $[1.46/16.919, 1.46/3.325] \approx [0.086, 0.44]$ for σ^2 and $[0.476 e^{-0.127 \cdot 1.833}, 0.476 e^{0.127 \cdot 1.833}]\mathrm{h}^{-1} \approx [0.38, 0.60]\mathrm{h}^{-1}$ for λ, respectively.

7.3.2 Demonstration of an *MTTR*

The demonstration of an *MTTR* (in an acceptance test) will be investigated here by assuming that the repair time τ' is *lognormally* distributed with known σ^2 (method 1A of MIL-STD-471 [7.15]). A rule is sought to test the null hypothesis $H_0 : MTTR = MTTR_0$ against the alternative hypothesis $H_1 : MTTR = MTTR_1$ for given type I error α and type II error β (Appendix A8.3). The procedure (test plan) is as follows:

1. From α and β $(0 < \alpha < 1-\beta < 1)$, determine the quantiles t_β and $t_{1-\alpha}$ of the normal distribution (Table A9.1)

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t_\beta} e^{-\frac{x^2}{2}}\,dx = \beta \qquad \text{and} \qquad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t_{1-\alpha}} e^{-\frac{x^2}{2}}\,dx = 1-\alpha. \tag{7.49}$$

From $MTTR_0$ and $MTTR_1$, compute then the sample size n (next highest integer)

$$n = \frac{(t_{1-\alpha}\, MTTR_0 - t_\beta\, MTTR_1)^2}{(MTTR_1 - MTTR_0)^2}\,(e^{\sigma^2} - 1). \tag{7.50}$$

2. Perform n independent repairs and record the observed repair times $t_1, \ldots, t_n$ (representative sample of repair times).

3. Compute $\hat{\mathrm{E}}[\tau']$ according to Eq. (7.39) and reject $H_0: MTTR = MTTR_0$ if

$$\hat{\mathrm{E}}[\tau'] > MTTR_0\,(1 + t_{1-\alpha}\sqrt{\frac{e^{\sigma^2}-1}{n}}), \tag{7.51}$$

otherwise accept H_0.

The proof of the above rule implies a sample size $n \geq 30$, so that the quantity $\hat{\mathrm{E}}[\tau']$ can be assumed to have a normal distribution with mean $MTTR$ and variance $\mathrm{Var}[\tau']/n$ (Eq. (A6.148)). Considering the type I and type II errors

$$\alpha = \Pr\{\hat{\mathrm{E}}[\tau'] > c \mid MTTR = MTTR_0\}, \qquad \beta = \Pr\{\hat{\mathrm{E}}[\tau'] < c \mid MTTR = MTTR_1\},$$

and using Eqs. (A6.105) and (7.49), the relationship

$$c = MTTR_0 + t_{1-\alpha}\sqrt{\frac{\mathrm{Var}_0[\tau']}{n}} = MTTR_1 + t_\beta\sqrt{\frac{\mathrm{Var}_1[\tau']}{n}} \tag{7.52}$$

can be found, with $\mathrm{Var}_0[\tau'] = (e^{\sigma^2} - 1)\,MTTR_0^2$ and $\mathrm{Var}_1[\tau'] = (e^{\sigma^2} - 1)\,MTTR_1^2$ according to Eq. (7.42). The sample size n (Eq. (7.50)) follows then from Eq. (7.52) and the right hand side of Eq. (7.51) is equal to the constant c as per Eq. (7.52).

The *operating characteristic curve* can be computed from

$$\Pr\{\text{acceptance} \mid MTTR\} = \Pr\{\hat{\mathrm{E}}[\tau'] \leq c \mid MTTR\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d} e^{-\frac{x^2}{2}}\,dx, \tag{7.53}$$

with

$$d = \frac{MTTR_0}{MTTR}\,t_{1-\alpha} - (1 - \frac{MTTR_0}{MTTR})\sqrt{\frac{n}{e^{\sigma^2}-1}}.$$

Replacing in d the quantity $n/(e^{\sigma^2} - 1)$ from Eq. (7.50) one recognizes that the operating characteristic curve is independent of σ^2 (rounding of n neglected).

Example 7.12

Determine the rejection conditions (Eq. (7.51)) and the related operating characteristic curve for the demonstration of $MTTR = MTTR_0 = 2\,\text{h}$ against $MTTR = MTTR_1 = 2.5\,\text{h}$ with $\alpha = \beta = 0.1$. σ^2 is assumed to be 0.2.

Solution

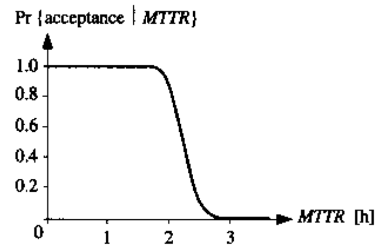
For $\alpha = \beta = 0.1$, Eq. (7.49) and Table A9.1 yield $t_{1-\alpha} = 1.28$ and $t_\beta = -1.28$. From Eq. (7.50), $n = 30$ and the rejection condition is therefore

$$\sum_{i=1}^{30} t_i > 2\,\text{h}\,(1 + 1.23\sqrt{\frac{e^{0.2}-1}{30}})\,30 = 66.6\,\text{h}.$$

From Eq. (7.53), the operating characteristic curve is given by

$$\Pr\{\text{acceptance} \mid MTTR\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d} e^{-\frac{x^2}{2}}\,dx,$$

with $d \approx 25.84\,\text{h}\,/\,MTTR - 11.64$ (see graph).



7.4 Accelerated Testing

The failure rate λ of electronic components lies typically between 10^{-10} and $10^{-7}\,\text{h}^{-1}$, and that of assemblies in the range of 10^{-7} to $10^{-5}\,\text{h}^{-1}$. With such figures, cost and scheduling considerations demand the use of accelerated testing for λ or $MTBF = 1/\lambda$ estimation and demonstration, for instance when reliable *field data* are not readily available. An *accelerated test* is a test in which the applied stress is chosen to exceed that encountered in field operation, in order to shorten the time to failure of the item considered, but still below the technological limits, to *avoid alteration* of the involved *failure mechanism* (genuine acceleration). In accelerated tests, failure mechanisms are assumed to be activated selectively by increased stress. The quantitative relationship between degree of activation and extent of stress, i.e. the *acceleration factor A*, is determined via specific tests. Generally it is assumed that the stress will not have any influence on the *type* of the failure-free operating time *distribution function* of the item under test, but only modify the *parameters*. In the following, this hypothesis is assumed to be valid.

Many electronic component *failure mechanisms* are activated through an increase in *temperature*. Calculating the acceleration factor A, the *Arrhenius model* can often be applied over a reasonably large temperature range (about 0 to 150°C for ICs). The *Arrhenius model* is based on the Arrhenius rate law [3.42], which states that the rate v of a simple (first-order) chemical reaction depends on temperature T as

$$\nu = \nu_0 \, e^{-\frac{E_a}{kT}}. \qquad (7.54)$$

E_a and ν_o are parameters, k is the Boltzmann constant ($k = 8.6 \cdot 10^{-5}$ eV/K), and T the absolute temperature in Kelvin degrees. E_a is the *activation energy* and is expressed in eV. Assuming that the event considered (for example the diffusion between two liquids) occurs when the chemical reaction has reached a given *threshold*, and the reaction time dependence is given by a function $\mathrm{r}(t)$, then the relationship between the times t_1 and t_2 necessary to reach at two temperatures T_1 and T_2 a given level of the chemical reaction considered can be espressed as

$$\nu_1 \, \mathrm{r}(t_1) = \nu_2 \, \mathrm{r}(t_2).$$

Furthermore, assuming $\mathrm{r}(t) \sim t$, i.e. a linear time dependence, it follows that

$$\nu_1 t_1 = \nu_2 t_2.$$

Substituting in Eq. (7.54) and rearranging, yields

$$\frac{t_1}{t_2} = e^{\frac{E_a}{k}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)},$$

By transferring this *deterministic* model to the means $MTTF_1$ and $MTTF_2$ of the (random) failure-free operating times of a given item at temperatures T_1 and T_2, it is possible to define an *acceleration factor* A

$$A = \frac{MTTF_1}{MTTF_2} \qquad \text{or} \qquad A = \frac{MTBF_1}{MTBF_2} = \frac{\lambda_2}{\lambda_1} \qquad (7.55)$$

expressed by

$$A = e^{\frac{E_a}{k}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)}. \qquad (7.56)$$

The right hand sides of Eq. (7.55) applies to the case of a constant (time independent but stress dependent) failure rate $\lambda(t) = \lambda$. Eq. (7.56) can be reversed to give an estimate $\hat{E}_a$ for the activation energy E_a based on $M\hat{T}TF_1$ and $M\hat{B}TF_2$ obtained empirically from two life tests at temperatures T_1 and T_2. Practical applications generally assume at *least three tests* at T_1, T_2, and T_3 in order also to check the *validity of the model*. Activation energy is highly dependent upon the particular *failure mechanism* involved (Table 3.6). High E_a values lead to high activation factors, due to the assumed relationships $\nu_1 t_1 = \nu_2 t_2$ and $\nu \sim 1/e^{E_a/kT}$. For ICs, global values of E_a lie between 0.3 and 0.7 eV, value which can basically be obtained empirically from the curves of the failure rate as a function of the junction temperature. It must be noted that the Arrhenius model does not hold for all electronic devices (Fig. 2.5) and for each temperature range. Figure 7.10 shows the acceleration factor A from Eq. (7.56) as a function of θ_2 in °C, for $\theta_1 = 35$ and 55°C and with E_a as parameter ($\theta_i = T_i - 273$).

In the case of a *constant failure rate* λ, the acceleration factor $A = MTBF_1 / MTBF_2 = \lambda_2 / \lambda_1$ can be used as a *multiplicative factor* in the conversion of the *cumulative operating time* from stress T_2 to stress T_1 (Example 7.13).

Example 7.13

Four failures have occurred during 10^7 cumulative operating hours of a digital CMOS IC at a chip temperature of 130°C. Assuming $\theta_1 = 35$°C, a constant failure rate λ, and an activation energy $E_a = 0.4$ eV, determine the interval estimation of λ for $\gamma = 0.8$.

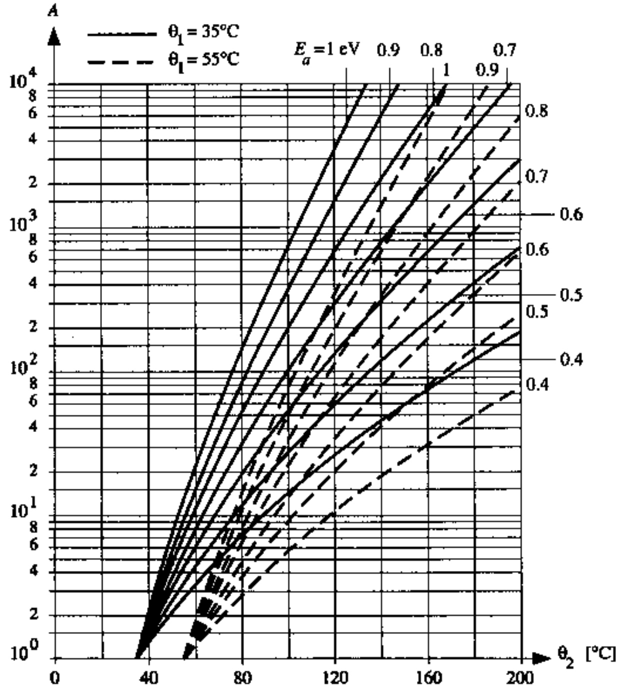


Figure 7.10 Acceleration factor A according to the Arrhenius model (Eq. (7.56)) as a function of θ_2 for $\theta_1 = 35$ and 55°C, and with E_a in eV as parameter ($\theta_i = T_i - 273$)

Solution

For $\theta_1 = 35°\text{C}$, $\theta_2 = 130°\text{C}$, and $E_a = 0.4\,\text{eV}$ it follows from Fig. 7.10 or Eq. (7.56) that $A \approx 35$. The cumulative operating time at $35°\text{C}$ is thus $T = 0.35 \cdot 10^9\,\text{h}$ and the point estimate for λ is $\hat{\lambda} = k/T \approx 11.4 \cdot 10^{-9}\,\text{h}^{-1}$. With $k = 4$ and $\gamma = 0.8$, it follows from Fig. 7.6 that $\hat{\lambda}_l / \hat{\lambda} \approx 0.43$ and $\hat{\lambda}_u / \hat{\lambda} \approx 2$; the confidence interval of λ is therefore $[4.9, 22.8] \cdot 10^{-9}\,\text{h}^{-1}$.

If the item under consideration exhibits more than one *dominant failure mechanism* or consists of elements $E_1, \dots, E_n$ having different failure mechanisms, the series reliability model (Eqs. (A6.80) and (2.16)) can be often used to calculate the *compound failure rate* $\lambda_S(T_2)$ at temperature (stress) T_2 by considering the failure rates of the individual elements $\lambda_i(T_1)$ and the corresponding acceleration factors A_i

$$\lambda_S(T_2) = \sum_{i=1}^{n} A_i \, \lambda_i(T_1). \tag{7.57}$$

Equation (7.57) also applies to the case of time-dependent failure rates $\lambda_i(T_1) = \lambda_i(t, T_1)$.

Example 7.14

A PCB contains 10 metal film resistors with stress factor $S = 0.1$ and $\lambda(25°\text{C}) = 0.2 \cdot 10^{-9}\,\text{h}^{-1}$, 5 ceramic capacitors (class 1) with $S = 0.4$ and $\lambda(25°\text{C}) = 0.8 \cdot 10^{-9}\,\text{h}^{-1}$, 2 electrolytic capacitors (Al wet) with $S = 0.6$ and $\lambda(25°\text{C}) = 6 \cdot 10^{-9}\,\text{h}^{-1}$, and 4 ceramic-packaged linear ICs with $\Delta\theta_{JA} = 10°\text{C}$ and $\lambda(35°\text{C}) = 20 \cdot 10^{-9}\,\text{h}^{-1}$. Neglecting the contribution of the printed wiring and of the solder joints, determine the failure rate of the PCB at a burn-in temperature θ_A of 80°C on the basis of failure rate relationships as given in Fig. 2.4.

Solution

The resistor and capacitor acceleration factors can be obtained from Fig. 2.4 as

resistor: $A \approx 2.5 / 0.7 \approx 3.6$
ceramic capacitor (class 1): $A \approx 4.2 / 0.5 \approx 8.4$
electrolytic capacitor (Al wet): $A \approx 13.6 / 0.35 \approx 38.9$.

Using Eq. (2.4) for the ICs, it follows that $\lambda \sim \Pi_T$. With $\theta_J = 35°\text{C}$ and $90°\text{C}$, the acceleration factor for the linear ICs can then be obtained from Fig. 2.5 as $A \approx 7.5 / 0.8 \approx 9.4$. From Eq. (7.57), the failure rate of the PCB is then

$$\lambda(25°\text{C}) = (10 \cdot 0.2 + 5 \cdot 0.8 + 2 \cdot 6 + 4 \cdot 20) 10^{-9}\text{h}^{-1} \approx 100 \cdot 10^{-9}\text{h}^{-1}$$

$$\lambda(85°\text{C}) = (10 \cdot 0.2 \cdot 3.6 + 5 \cdot 0.8 \cdot 8.4 + 2 \cdot 6 \cdot 38.9 + 4 \cdot 20 \cdot 9.4) 10^{-9}\text{h}^{-1} \approx 1{,}260 \cdot 10^{-9}\text{h}^{-1}.$$

A further model for investigating the time scale reduction (time compression) resulting from an increase in temperature has been proposed by H. Eyring [3.42, 7.16, 7.17]. The *Eyring model* defines the acceleration factor as

$$A = \frac{T_2}{T_1} e^{\frac{B}{k}(\frac{1}{T_1} - \frac{1}{T_2})}. \tag{7.58}$$

B is a constant, not necessarily an activation energy. Eyring also suggests the following model, which considers the influences of temperature T and of a further stress parameter X

$$A = \frac{T_2}{T_1} e^{\frac{B}{k}(\frac{1}{T_1} - \frac{1}{T_2})} e^{[X_1 (C + \frac{D}{k\,T_1}) - X_2 (C + \frac{D}{k\,T_2})]}. \tag{7.59}$$

Equation (7.59) is known as the *generalized Eyring model*. In this model, a function of the normalized variable $x = X / X_0$ can also be used instead of the quantity X itself (for example x^n, $1/x^n$, $\ln x^n$, $\ln(1/x^n)$). B is not necessarily an activation energy, C and D are constants. Modifications of the generalized Eyring model lead to more recent models, for example

$$A = (\frac{j_2}{j_1})^n e^{\frac{E_a}{k}(\frac{1}{T_1} - \frac{1}{T_2})} \tag{7.60}$$

for electromigration (j = current density) and

$$A = (\frac{RH_2}{RH_1})^n e^{\frac{E_a}{k}(\frac{1}{T_1} - \frac{1}{T_2})} \tag{7.61}$$

for corrosion (RH = relative humidity, see also Eqs. (3.2) to (3.6)).

Application/refinement of the above models to ULSI ICs is in progress with emphasis on:

1. Investigation of effects in oxide (inversion, time-dependent dielectric breakdown, hot carriers, trapping), as well as of package and externally induced failure mechanisms.
2. Identification and analysis of the causes for early failures or for a "premature" wearout.
3. Development of *physical models* for *failure mechanisms* and of *simplified models* for *reliability predictions* in practical applications.

Such efforts will allow a better *physical understanding* of the failure rate of complex components.

In addition to the accelerated tests introduced above, a rough estimate of component life times can be obtained through *short-term tests* under extreme stresses. Examples are humidity testing of plastic-packaged ICs at high pressure and nearly 100% *RH*, or tests of ceramic-packaged ICs at temperatures up to 350°C. Such tests can activate failure mechanisms which would not occur during normal operation, so care must be taken in extrapolating the results to situations exhibiting lower stresses. Experience shows that under high stress, life times are often lognormally distributed, thus with a *strong time dependence* of the failure rate (see Table A6.1 for an example).

7.5 Goodness-of-fit Tests

Goodness-of-fit tests deal with the testing (checking) of a hypothesis $H_0: \mathrm{F}(t) = \mathrm{F}_0(t)$ for a given type I error α against a general alternative hypothesis $H_1: \mathrm{F}(t) \neq \mathrm{F}_0(t)$, on the basis of n independent observations $t_1, \ldots, t_n$ of a random variable τ distributed according to $\mathrm{F}(t) = \Pr\{\tau \le t\}$. Many possibilities exist to test such a hypothesis, this section discusses the Kolmogorov-Smirnov test and the chi-square test (Appendices A8.3.2 and A8.3.3).

7.5.1 Kolmogorov-Smirnov Test

The *Kolmogorov-Smirnov test* is based on the convergence for $n \to \infty$ of the empirical distribution function (Eq. (A8.1))

$$\hat{\mathrm{F}}_n(t) = \begin{cases} 0 & \text{for } t < t_{(1)} \\ \dfrac{i}{n} & \text{for } t_{(i)} \le t < t_{(i+1)} \\ 1 & \text{for } t \ge t_{(n)} \end{cases} \tag{7.62}$$

to the true distribution function, and compares the experimentally obtained $\hat{\mathrm{F}}_n(t)$ with the given (postulated) $\mathrm{F}_0(t)$. $\mathrm{F}_0(t)$ is assumed here to be known and continuous. The procedure is as follows:

1. Determine the largest deviation D_n between $\hat{\mathrm{F}}_n(t)$ and $\mathrm{F}_0(t)$

$$D_n = \sup_{-\infty<t<\infty} |\hat{\mathrm{F}}_n(t) - \mathrm{F}_0(t)|. \tag{7.63}$$

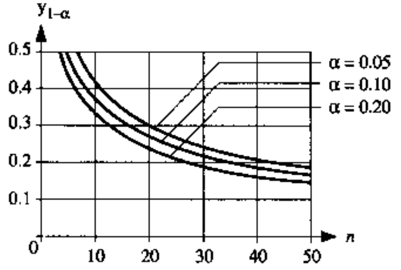


Figure 7.11 Largest deviation $y_{1-\alpha}$ between a postulated distribution function $\mathrm{F}_0(t)$ and the corresponding empirical distribution function $\hat{\mathrm{F}}_n(t)$ at the level $1-\alpha$ $(\Pr\{D_n \le y_{1-\alpha} \mid \mathrm{F}_0(t) \text{ true}\} = 1-\alpha)$

2. From the given type I error α and the sample size n, use Table A9.5 or Fig. 7.11 to determine the critical value $y_{1-\alpha}$.
3. Reject H_0: $\mathrm{F}(t) = \mathrm{F}_0(t)$ if $D_n > y_{1-\alpha}$; otherwise accept H_0.

This procedure can be easily combined with a graphical evaluation of the data. To do this, $\hat{\mathrm{F}}_n(t)$ and the band $\mathrm{F}_0(t) \pm y_{1-\alpha}$ are drawn using a *probability chart* on which $\mathrm{F}_0(t)$ can be represented by a straight line. If $\hat{\mathrm{F}}_n(t)$ leaves the band $\mathrm{F}_0(t) \pm y_{1-\alpha}$, the hypothesis $H_0: \mathrm{F}(t) = \mathrm{F}_0(t)$ is to be rejected (note that the band width is not constant when using a probability chart). Probability charts are discussed in Appendix A.8.1.3, examples are in Appendix A9.8 and Figs. 7.12 to 7.14. Example 7.15 (Fig. 7.12) shows a graphical evaluation of data for the case of a Weibull distribution, Example 7.16 (Fig. 7.13) investigates the distribution function of a population with *early failures* and a constant failure rate using a *Weibull probability chart*, and Example 7.17 (Fig. 7.14) uses the Kolmogorov-Smirnov test to check agreement with a lognormal distribution.

Example 7.15

Accelerated life testing of a wet Al electrolytic capacitor leads to the following 13 ordered observations of the lifetime: 59, 71, 153, 235, 347, 589, 837, 913, 1185, 1273, 1399, 1713, and 2567 h. (i) Draw the empirical distribution function of the data on a Weibull probability chart. (ii) Assuming that the underlying distribution function is Weibull-type, determine the parameters $\hat{\lambda}$ and $\hat{\beta}$ graphically. (iii) The maximum likelihood estimation of λ and β yields $\hat{\beta} = 1.12$, calculate $\hat{\lambda}$ and compare the results of (iii) with those of (ii).

Solution

(i) Figure 7.12 presents the empirical distribution function $\hat{\mathrm{F}}_n(t)$ on Weibull probability paper. (ii) The graphical determination of λ and β leads to (straight line (ii)) $\hat{\lambda} \approx 1/840\,\mathrm{h}$ and $\hat{\beta} \approx 1.05$. (iii) With $\hat{\beta} \approx 1.12$, Eq. (A8.31) yields $\hat{\lambda} \approx 1/908\,\mathrm{h}$, see the straight line (iii) in Fig. 7.12.

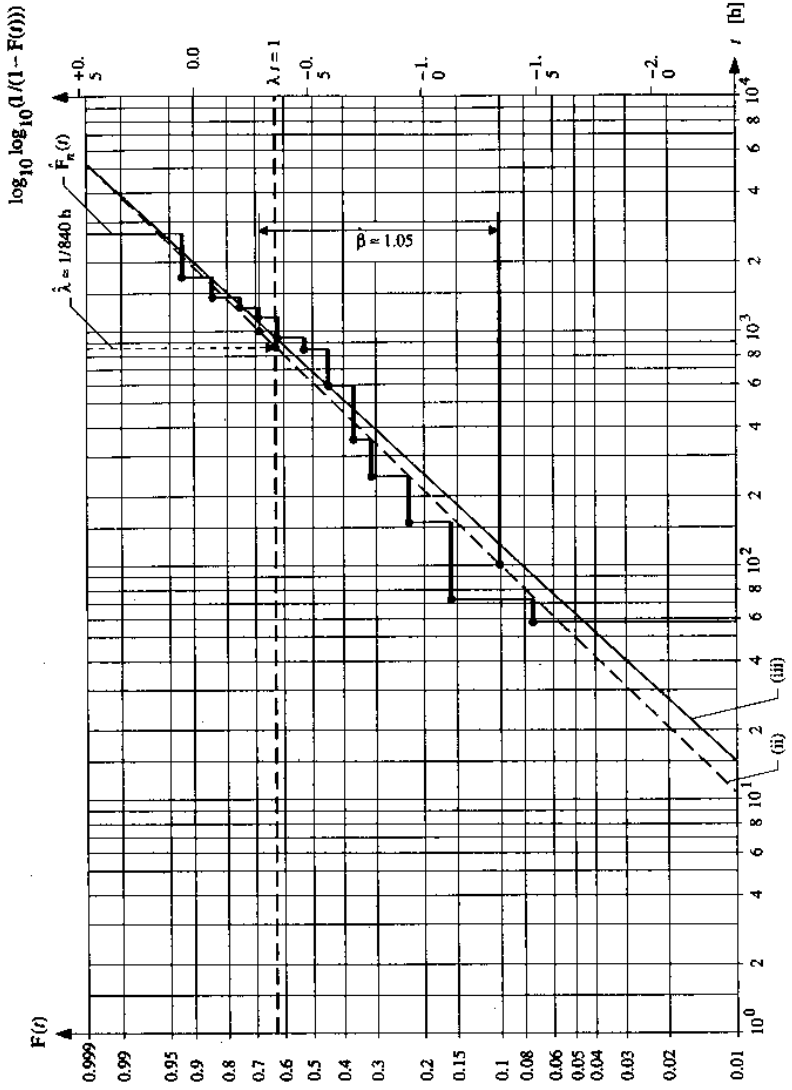


Figure 7.12 Empirical distribution function $\hat{F}_n(t)$ and estimated Weibull distribution functions ((ii) and (iii)) as in Example 7.15

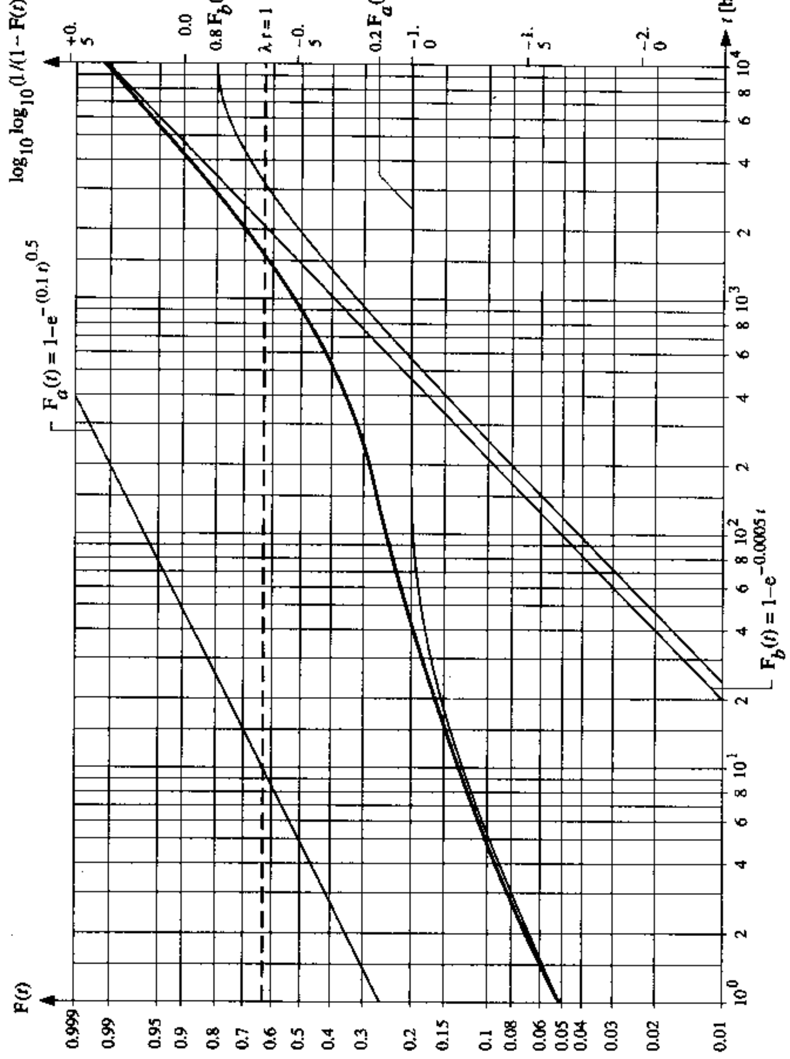


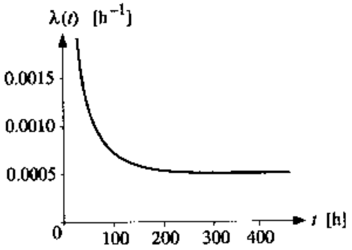
Figure 7.13 Shape of a weighted sum of a Weibull distribution $F_a(t)$ and an exponential distribution $F_b(t)$ as in Example 7.16

Example 7.16

Investigate the *mixed distribution function* $\mathrm{F}(t) = 0.2[1 - e^{-(0.1t)^{0.5}}] + 0.8[1 - e^{-0.0005t}]$ on a Weibull probability chart.

Solution

The weighted sum of a Weibull distribution ($\beta = 0.5$, $\lambda = 0.1\,\mathrm{h}^{-1}$, and $MTTF = 20\,\mathrm{h}$) with an exponential distribution ($\lambda = 0.0005\,\mathrm{h}^{-1}$ and $MTTF = MTBF = 1/\lambda = 2000\,\mathrm{h}$) represents the distribution function of a population of items with failure rate $\lambda(t) = [0.01(0.1t)^{-0.5}_{0.5} e^{-(0.1t)^{0.5}} + 0.0004\,e^{-0.0005t}] / [0.2\,e^{-(0.1t)^{0.5}} + 0.8\,e^{-0.0005t}]$, i..e. with early failures up to about $t \approx 200\,\mathrm{h}$, see graph ($\lambda(t)$ is practically constant at $0.0005\,\mathrm{h}^{-1}$ for t between $300\,\mathrm{h}$ and $400{,}000\,\mathrm{h}$, so that for $t > 300\,\mathrm{h}$ a constant failure rate can be assumed for practical purposes). Figure 7.13 gives the function F(t) on a Weibull probability chart, showing the typical *s-shape*.



Example 7.17

Use the Kolmogorov-Smirnov test to verify with a type I error $\alpha = 0.2$, whether the repair times defined by the observations $t_1, \ldots, t_{10}$ of Example 7.11 are distributed according to a lognormal distribution function with parameters $\lambda = 0.5\,\mathrm{h}^{-1}$ and $\sigma = 0.4$ (hypothesis H_0).

Solution

The lognormal distribution (Eq. (7.41)) with $\lambda = 0.5\,\mathrm{h}^{-1}$ and $\sigma = 0.4$ is represented by a straight line on Fig. 7.14 ($\mathrm{F}_0(t)$). With $\alpha = 0.2$ and $n = 10$, Table A9.5 or Fig. 7.11yields $y_{1-\alpha} = 0.323$ and thus the band $\mathrm{F}_0(t) \pm 0.323$. Since the empirical distribution function $\hat{\mathrm{F}}_n(t)$ does not leave the band $\mathrm{F}_0(t) \pm y_{1-\alpha}$, the hypothesis H_0 can be accepted.

7.5.2 Chi-square Test

The *chi-square test* (χ^2 *test*) can be used for both *continuous* or *noncontinuous* distribution functions $\mathrm{F}_0(t)$ of τ. Furthermore, $\mathrm{F}_0(t)$ need not be completely known.

For $\mathrm{F}_0(t)$ *completely known*, the procedure is as follows:

1. Partition the definition range of the random variable τ into k intervals (classes) $(a_1, a_2], (a_2, a_3], \ldots, (a_k, a_{k+1}]$, the choice of the classes must be made independently of the observations $t_1, \ldots, t_n$ (rule: $n\,p_i \ge 5$, with p_i as in point 3).
2. Determine the number of observations k_i in each class $(a_i, a_{i+1}]$, $i = 1, \ldots, k$ (k_i = number of t_j with $a_i < t_j \le a_{i+1}$.
3. Assuming the hypothesis H_0, compute the expected number of observations for each class $(a_i, a_{i+1}]$

$$n\,p_i = n(\mathrm{F}_0(a_{i+1}) - \mathrm{F}_0(a_i)), \qquad i = 1, \ldots, k. \tag{7.64}$$

4. Compute the statistics

$$X_n^2 = \sum_{i=1}^{k} \frac{(k_i - n\,p_i)^2}{n\,p_i} = \sum_{i=1}^{k} \frac{k_i^2}{n\,p_i} - n. \tag{7.65}$$

5. For a given type I error α, use Table A9.2 or Fig. 7.15 to determine the $(1-\alpha)$ quantile of the chi-square distribution with $k-1$ degrees of freedom $\chi^2_{k-1,1-\alpha}$.
6. Reject $H_0 : \mathrm{F}(t) = \mathrm{F}_0(t)$ if $X_n^2 > \chi^2_{k-1,1-\alpha}$, otherwise accept H_0.

If $\mathrm{F}_0(t)$ is *not completely known* ($\mathrm{F}_0(t) = \mathrm{F}_0(t, \theta_1, \ldots, \theta_r)$, where $\theta_1, \ldots, \theta_r$ are unknown parameters), modify the above procedure after step 2 as follows:

3'. On the basis of the observations k_i in each class $(a_i, a_{i+1}]$, $i = 1, \ldots, k$ determine the maximum likelihood estimates for the parameters $\theta_1, \ldots, \theta_r$ from the following system of r algebraic equations

$$\sum_{i=1}^{k} \frac{k_i}{p_i(\theta_1, \ldots, \theta_r)} \left.\frac{\partial p_i(\theta_1, \ldots, \theta_r)}{\partial \theta_j}\right|_{\theta_j = \hat{\theta}_j} = 0, \qquad j = 1, \ldots, r \tag{7.66}$$

with $p_i = \mathrm{F}_0(a_{i+1}, \theta_1, \ldots, \theta_r) - \mathrm{F}_0(a_i, \theta_1, \ldots, \theta_r) > 0$, $\quad p_0 + \ldots + p_k = 1$ and $k_1 + \ldots + k_k = n$; for each class $(a_i, a_{i+1}]$, compute then the expected number of observations, i.e.

$$n\,\hat{p}_i = n[\mathrm{F}_0(a_{i+1}, \hat{\theta}_1, \ldots, \hat{\theta}_r) - \mathrm{F}_0(a_i, \hat{\theta}_1, \ldots, \hat{\theta}_r)], \qquad i = 1, \ldots, k. \tag{7.67}$$

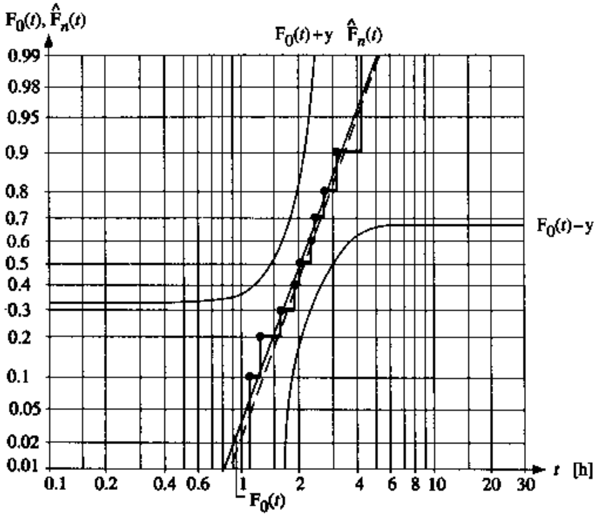


Figure 7.14 Kolmogorov-Smirnov test to check the repair times distribution as in Example 7.17 (the distribution function with $\hat{\lambda}$ and $\hat{\sigma}$ from Example 7.11 is shown dashed for information only)

4'. Calculate the statistics

$$\hat{X}_n^2 = \sum_{i=1}^{k} \frac{(k_i - n\,\hat{p}_i)^2}{n\,\hat{p}_i} = \sum_{i=1}^{k} \frac{k_i^2}{n\,\hat{p}_i} - n. \tag{7.68}$$

5'. For given type I error α, use Table A9.2 or Fig. 7.15 to determine the $(1-\alpha)$ quantile of the chi-square distribution with $k-1-r$ degrees of freedom.

6'. Reject $H_0 : \mathrm{F}(t) = \mathrm{F}_0(t)$ if $\hat{X}_n^2 > \chi^2_{k-1-r,1-\alpha}$, otherwise accept H_0.

Comparing the above two procedures, it can be noted that the number of degrees of freedom has been reduced from $k-1$ to $k-1-r$, where r is the number of parameters of $\mathrm{F}_0(t)$ which have been estimated from the observations $t_1, \ldots, t_n$ using the *multinomial distribution* (Example A8.8).

Example 7.18

Let 160, 380, 620, 650, 680, 730, 750, 920, 1000, 1100, 1400, 1450, 1700, 2000, 2200, 2800, 3000, 4600, 4700, and 5000 h be 20 independent observations (realizations) of the failure-free operating time τ for a given assembly. Using the chi-square test for $\alpha = 0.1$ and the 4 classes $(0, 500]$, $(500, 1000]$, $(1000, 2000]$, $(2000, \infty)$, determine whether or not τ is exponentially distributed (hypothesis $H_0 : \mathrm{F}(t) = 1 - e^{-\lambda t}$, λ unknown).

Solution

The given classes yield number of observations of $k_1 = 2$, $k_2 = 7$, $k_3 = 5$, and $k_4 = 6$. The point estimate of λ is then given by Eq. (7.66) with $p_i = e^{-\lambda a_i} - e^{-\lambda a_{i+1}}$, yielding for $\hat{\lambda}$ the numerical solution $\hat{\lambda} \approx 0.562 \cdot 10^{-3}\,\mathrm{h}^{-1}$. Thus, the numbers of expected observations in each of the 4 classes are according to Eq. (7.67) $n\,\hat{p}_1 = 4.899$, $n\,\hat{p}_2 = 3.699$, $n\,\hat{p}_3 = 4.90$, and $n\,\hat{p}_4 = 6.499$. From Eq. (7.68) it follows that $\hat{X}_{20}^2 = 4.70$ and from Table A9.2, $X_{2,0.9}^2 = 4.605$. The hypothesis $H_0 : \mathrm{F}(t) = 1 - e^{-\lambda t}$ must be rejected since $\hat{X}_n^2 > \chi^2_{k-1-r,1-\alpha}$.

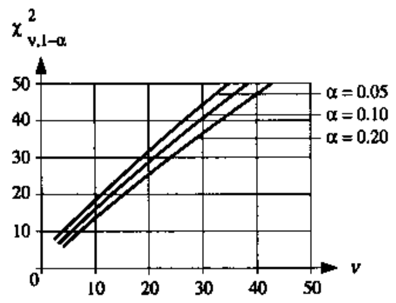


Figure 7.15 $\chi^2_{k-1,1-\alpha}$ quantile of the chi-square distribution function with ν degrees of freedom

8 Quality and Reliability Assurance During the Production Phase

Reliability assurance has to be continued during the *production phase,* hand in hand with *quality assurance activities*, in particular concerning the monitoring and control of *production processes* and *item's configuration*, the performance of *in-process* and *final tests*, the *screening* of critical components and assemblies, and the *systematic collection, analysis, and correction* of *defects* and *failures*. The last measure is basic for a *learning process*, whose aim is to *optimize* the *quality of manufacture*, taking into account cost and time schedule limitations. This chapter introduces the general aspects of quality and reliability assurance during production, discusses test and *screening procedures* for electronic components and assemblies, introduces the concept of a *test and screening strategy*, and discusses models for *reliability growth* during production. For specific poblems related to the qualification and monitoring of production processes, one should refer to the literature, e.g. [8.1 to 8.14].

8.1 Basic Activities

The quality and reliability built into an item in the design phase should be retained during production. For this purpose, the following basic activities are necessary:

1. Management of the *item's configuration* (review and release of the production documentation, control and accounting of changes and modifications).
2. Selection and *qualification of production facilities and processes* (assembling, soldering, testing, etc.).
3. Monitoring and control of the *production procedures* (handling, testing, transportation, storage, etc.).
4. *Protection against damage* during production (electrostatic discharge (ESD), mechanical, thermal or electrical stresses).
5. Systematic collection, analysis, and correction of *defects* and *failures* occurring during the item's production or testing (to avoid repetition of similar problems).

6. Quality and reliability assurance during *procurement* (procurement documentation, incoming inspection, supplier audits).
7. *Calibration* of measurement and testing equipment.
8. Performance of *in-process and final tests* (functional and environmental).
9. *Screening* of critical components and assemblies.
10. Realization of a *test and screening strategy* (optimization of the cost and time schedule for testing and screening).

Configuration management, planning and control of *corrective actions*, as well as some important aspects of *statistical quality control* and *reliability tests* have been considered in Section 1.5, Chapter 7, and Appendices A3 and A5. The following sections present test and screening procedures for electronic components and assemblies, introduce the concept of a test and screening strategy, and discuss the modeling of reliability growth during production.

8.2 Testing and Screening of Electronic Components

8.2.1 Testing of Electronic Components

Most electronic components are tested today by the end user on a *sampling basis*. To be cost effective, sampling plans should take into consideration the *quality assurance effort of the component manufacturer*, in particular the confidence which can be given to the data furnished by him. In critical cases, the sample should be large enough to allow acceptance of more than 2 defective components (see e.g. Tables 7.1 to 7.3 and Sections 7.1.2 to 7.1.3). *100% incoming inspection* has to be performed for components used in high reliability and/or safety equipment and systems, for new components, or for some critical devices like power semiconductors, mixed-signal ICs, and complex logic ICs used at the limits of their *dynamic parameters*. Advantages of a 100% incoming inspection of electronic components are:

1. Quick detection of all relevant defects.
2. Reduction of the number of defective populated printed circuit boards (PCBs).
3. Simplification of the tests at PCB level.
4. Replacement of the defective components by the supplier.
5. Protection against *quality changes from lot to lot*, or within the same lot.

Despite such advantages, different kinds of *damage* (overstress during testing, assembling, or soldering) can cause problems at the PCB level. *Defective probabilities p* range today from ppm (part per million) for established components up to some few % for critical components. In defining a *test strategy*, the possible *change* of p from lot to lot or within the same lot should also be considered. An example of a test procedure for electronic components is given in Section 3.2.1 for VLSI ICs. Test strategies with cost consideration are developed in Section 8.4.

8.2.2 Screening of Electronic Components

Electronic components which are *new on the market*, produced in *small series*, subjected to an important *redesign*, or manufactured with *insufficiently stable process parameters* can exibit *early failures*, i.e. failures during the first operating hours (generally up to some few thousand hours). Because of high replacement costs at equipment level or in the field, components exhibiting early failures should be eliminated before they are mounted on printed circuit boards. Defining a cost-effective *screening sequence* is difficult for at least following two reasons:

1. It may activate *failure mechanisms* that would not appear in field operation.
2. It could introduce *damage* (ESD, transients) which may be the cause of further *early failures*.

Ideally, screening should be performed by skilled personnel, be focused on the *failure mechanisms* which have to be activated, and should not cause *damage* or *alteration* to the components involved. Experience on a large number of components shows that for *established technologies* and *stable process parameters*, *thermal cycles* for discrete (in particular power) devices and *burn-in* for ICs are the most effective steps to precipitate *early failures*. Table 8.1 gives screening sequences of electronic components used in *high reliability or safety* equipment and systems.

The following is a description of some important steps in a screening sequence for *ICs in hermetic packages for high reliability or safety applications*:

1. *High-temperature storage*: The purpose of high temperature storage is the stabilization of the thermodynamic equilibrium and thus of the IC electrical parameters. Failure mechanisms related to surface problems (contamination, oxidation, contacts) are particularly activated. The ICs are placed on a metal tray (pins on the tray to avoid thermal voltage stresses) in an oven at 150°C for 24h. Should solderability be a problem, a protective atmosphere (N_2) can be used.
2. *Thermal cycles*: The purpose of thermal cycles is to test the ICs ability to endure rapid temperature changes, this activates failure mechanisms related to mechanical stresses caused by mismatch in the *expansion coefficients* of the materials used. Thermal cycles are generally performed air to air in a two-

Table 8.1 Test and screening procedures for electronic components used in *high reliability or safety* equipment and systems (not applicable to surface-mounted devices)

Component	Sequence
Resistors	Visual inspection, 20 thermal cycles ($-40/+125°C$) for resistor networks,* 48 h steady-state burn-in at 100°C and $0.6\,P_N$,* el. test at 25°C*
Capacitors	
• Film	Visual inspection, 48 h steady-state burn-in at $0.9\theta_{\max}$ and U_N,* el. test at 25°C (C, $\tan\delta$, R_{is}),* measurement of R_{is} at 70°C*
• Ceramic	Visual inspection, 20 thermal cycles (θ_{extr}),* 48 h steady-state burn-in at U_N and $0.9\theta_{\max}$,* el. test at 25°C (C, $\tan\delta$, R_{is}),* measurement of R_{is} at 70°C*
• Tantalum (solid)	Visual inspection, 10 thermal cycles (θ_{extr}),* 48 h steady-state burn-in at U_N and $0.9\theta_{\max}$ (low Z_Q),* el. test at 25°C (C, $\tan\delta$, I_r),* meas. of I_r at 70°C*
• Aluminum (wet)	Visual inspection, forming (as necessary), 48 h steady-state burn-in at U_N and $0.9\theta_{\max}$,* el. test at 25°C (C, $\tan\delta$, I_r),* measurement of I_r at 70°C*
Diodes (Si)	Visual inspection, 30 thermal cycles ($-40/+125°C$),* 48 h reverse bias burn-in at 125°C,* el. test at 25°C (I_r, U_F, $U_{R\min}$),* seal test (fine/gross leak)*+
Transistors (Si)	Visual inspection, 20 thermal cycles ($-40/+125°C$),* 50 power cycles (25/125°C, ca. 1 min on / 2 min off) for power elements,* el. test at 25°C (β, I_{CEO}, $U_{CEO\min}$),* seal test (fine/gross leak)*+
Optoelectronics	
• LED, IRED	Visual inspection, 72 h high temp. storage at 100°C,* 20 thermal cycles ($-20/+80°C$),* el. test at 25°C (U_F, $U_{R\min}$),* seal test (fine/gross leak)*+
• Optocoupler	Visual inspection, 20 thermal cycles ($-25/100°C$), 72 h reverse bias burn-in (HTRB) at 85°C,* el. test at 25°C (I_C/I_F, U_F, $U_{R\min}$, U_{CEsat}, I_{CEO}), seal test (fine/gross leak)*+
Digital ICs	
• BiCMOS	Visual inspection, reduced el. test at 25°C, 48 h dyn. burn-in at 125°C, el. test at 70°C, seal test (fine/gross leak)*+
• MOS (VLSI)	Visual inspection, reduced el. test at 25°C (rough functional test, I_{DD}), 72 h dyn. burn-in at 125°C,* el. test at 70°C,* seal test (fine/gross leak)*+,
• CMOS (VLSI)	Visual inspection, reduced el. test at 25°C (rough functional test, I_{DD}), 48 h dyn. burn-in at 125°C,* el. test at 70°C,* seal test (fine/gross leak)*+
• EPROM, EEPROM (>1M)	Visual inspection, programming (CHB), high temp. storage (48 h / 125°C), erase, programming (inv. CHB), high temp. storage 48 h / 125°C, erase, el. test at 70°C, seal test (fine/gross leak)*+
Linear ICs	Visual inspection, reduced el. test at 25°C (rough functional test, I_{CC}, offsets), 20 thermal cycles ($-40/+125°C$),* 96 h reverse bias burn-in (HTRB) at 125°C with red. el. test at 25°C,* el. test at 70°C,* seal test (fine/gross leak)*+
Hybrid ICs	Visual inspection, high temp. storage (24 h / 125°C), 20 thermal cycles ($-40/+125°C$), constant acceleration (2,000 to 20,000 g_n / 60 s), red. el. test at 25°C, 96 h dynamic burn-in at 85 to 125°C, el. test at 25°C, seal test (fine/gross leak)*+

*sampling, + hermetic packages, el. = electrical, red. = reduced, N = rated value, CHB = checkerboard

chamber oven (transfer from low to high temperature chamber and vice versa using a lift). The ICs are placed on a metal tray (pins on the tray to avoid thermal voltage stresses) and subjected to at least 10 thermal cycles from –65 to +150°C. Dwell time at the temperature extremes should be ≥ 10 min (after the thermal equilibrium of the IC has been reached within ±5°C), transition time less than 1 min. Should solderability be a problem, a protective atmosphere (N_2) can be used.

3. *Constant acceleration*: The purpose of the constant acceleration is to check the mechanical stability of die-attach, bonding, and package. This step is only performed for ICs in hermetic packages, when used in critical applications. The ICs are placed in a centrifuge and subjected to an acceleration of $30{,}000\,g_n$ ($300{,}000\,\mathrm{m/s^2}$) for 60 seconds (generally z-axis only).
4. *Burn-in*: Burn-in is a relatively expensive, but efficient screening step that provokes approx. 80% of the chip-related and 30% of the package-related *early failures*. The ICs are placed in an oven at 125°C for 160 hours and are operated statically or dynamically at this temperature (cooling under power at the end of burn-in is often required). Ideally, ICs should operate with the same electrical signals as in the field. The consequence of the high burn-in temperature is a time *acceleration* factor *A* often given by the Arrhenius model (Eq. (7.56))

$$A = \frac{\lambda_2}{\lambda_1} = \frac{MTBF_1}{MTBF_2} \approx e^{\frac{E_a}{k}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)},$$

where E_a is the *activation energy*, *k* the Boltzmann's constant ($8.6 \cdot 10^{-5}$ eV/K), and λ_1 and λ_2) are the failure rates at chip temperatures T_1 and T_2 (in K), respectively, see Fig. 7.10 for a graphical representation. The *activation energy* E_a varies according to the *failure mechanisms* involved. *Global average values* for ICs lie between 0.4 and 0.7 eV. Using Eq. (7.56), the burn-in duration can be computed for a given application. For instance, if the period of early failures is 3,000 h, $\theta_1 = 55°C$, and $\theta_2 = 130°C$ (junction temperature in °C), the effective burn-in duration would be of about 50 h for $E_a \approx 0.65$ eV and 200 h for $E_a \approx 0.4$ eV. It is often difficult to decide whether a *static* or a *dynamic burn-in* is more effective. Should surface, oxide, and metallization problems be dominant, a static burn-in is better. On the other hand, a dynamic burn-in activates practically all failure mechanisms. It is therefore important to make such a choice on the basis of practical experience.
5. *Seal*: A seal test is performed to check the seal integrity of the cavity around the chip in hermetically-packaged ICs. It begins with the *fine leak test*: ICs are placed in a vacuum (1 h at 0.5 mmHg) and then stored in a helium atmosphere under pressure (ca. 4 h at 5 atm); after a waiting period in open air (30 min), helium leakage is measured with the help of a specially-calibrated mass spectrometer (required sensitivity approx. 10^{-8} atm cm³/s, depending on the

cavity volume). After the fine leak test, ICs are tested for *gross leak*: ICs are placed in a vacuum (1h at 5mmHg) and then stored under pressure (2h at 5atm) in fluorocarbon FC-72; after a short waiting period in open air (2min), the ICs are immersed in a fluorocarbon indicator bath (FC-40) at 125°C; a continuous stream of small bubbles or two large bubbles from the same place within 30 s indicates a defect.

8.3 Testing and Screening of Electronic Assemblies

Electrical testing of electronic assemblies, for instance populated printed circuit boards (PCBs), can be basically performed in one of the following ways:

1. Functional test within the assembly or unit in which the PCB is used.
2. Functional test with the help of functional test equipment.
3. In-circuit test followed by a functional test with the assembly or unit in which the PCB is used.

The first method is useful for small series production. It assumes that components have been tested (or are of sufficient quality) and that automatic or semi-automatic *localization* of defects on the PCB is possible. The second method is suitable for large series production, in particular from the point of view of protection against *damage* (ESD, backdriving, mechanical stresses), but is generally expensive. The third and most commonly used method assumes the availability of an appropriate *in-circuit test equipment*. With such equipment, each component is electrically isolated and tested statically or quasi-statically. This can be sufficient for passive components and discrete semiconductors, as well as for SSI and MSI ICs, but it cannot replace an electrical test at the incoming inspection for LSI and VLSI ICs (functional tests on in-circuit test equipment are limited to some few 100kHz and dynamic tests are not possible). Thus, even if in-circuit testing is used, incoming inspection of critical components should not be omitted. A further disadvantage of in-circuit testing is that the outputs of an IC can be forced to a LOW or a HIGH state. This stress (*backdriving*) is generally short (50ns), but may be sufficient to cause damage to the IC in question. In spite of this, and of some other problems (polarity of electrolytic capacitors, paralleled components, tolerance of analog devices), in-circuit testing is today the most effective means to test populated printed circuit boards (PCBs), on account also of its good *defect localization* capability.

Because of the large number of components and solder joints involved, the *defective probability* of a PCB can be relatively high in stable production conditions too. Experience shows that for a PCB with about 500 components and 3,000 solder joints, the following *indicative values* can be expected (see e.g. Table 1.3 for a fault report form):

- 0.5 to 2% defective PCBs (often, for 3/4 assembling and 1/4 components),
- 1.5 defects per defective PCB (mean value).

Considering such figures, it is important to remember that defective PCBs are often *reworked* and that generally a repair or rework has a negative influence on the quality and reliability of a PCB.

Screening populated printed circuit boards (PCBs) or assemblies with a higher integration level is in most cases a difficult task, because of the many different technologies involved. Experience on a large number of PCBs [3.76] has led to the following *screening sequence* which can be recommended for PCBs of standard technology used in *high reliability applications*:

1. Visual inspection and reduced electrical test.
2. 100 thermal cycles between 0°C and +80°C, with temperature gradient < 5°C/min (within the components), dwell time $\geq 10\,\text{min}$, and power off during cooling (gradient > 20°C/min only if this also occurs in the field and is compatible with the PCB technology).
3. 15min random vibration at 2 g_{rms}, 20 - 500Hz (to be performed if significant vibrations occur in the field).
4. 48h *run-in* at normal ambient temperature, with periodic power on/off switching.
5. Final electrical and functional test.

Careful investigation on *SMT assemblies* down to pitch 0.3mm [3.79, 3.80, 3.88] have shown that basically two different deformation mechanisms can be present in tin based solder joints, *grain boundary sliding* at rather low temperature (or thermal) gradients and *dislocation climbing* at higher temperature gradients (Section 3.4). For this reason, screening of populated PCBs in SMT *should be avoided* if the temperature gradient occuring in the field is not known. *Preventive actions* have to be preferred here, to build in quality and reliability during manufacturing.

The above procedure can be considered as an *environmental stress screening* (ESS), often performed on a 100% basis in a series production of *PCBs* used *in high reliability or safety applications* to provoke *early failures*. It can serve as basis for screening at higher integration levels.

Thermal cycles can be combined with power on/off switching or vibration to increase effectiveness. However, in general a *screening strategy* for PCBs (or at higher integration level) should be established on a case-by-case basis, and be periodically reconsidered (reduced or even cancelled if the percentage of early failures drops below a given value, of say 1%).

8.4 Test and Screening Strategies, Economic Aspects

8.4.1 Basic Considerations

Testing and screening of complex electronic components and populated printed circuit boards (PCBs) often accounts for over 20% of the total production costs. In view of the optimization of costs associated with testing and screening during production, every manufacturer of high-performance equipment and systems is confronted with the following question:

> *What is the most cost-effective approach to eliminate all defects, systematic failures, and early failures prior to shipment to the customer* ?

The answer to this question depends essentially on the level of quality, reliability, and safety required for the item considered, the consequence of a defect or a failure, the effectiveness of each test or screening step, as well as on the direct and *deferred costs* involved, warranty cost for instance. A *test and screening strategy* should thus be *tailored* to the item considered, in particular to its complexity, technology, and production procedures, but also to the facilities and skill of the manufacturer. In setting up such a strategy, the following aspects must be considered:

1. *Cost equations* should include *warranty costs* and *costs for loss of image*.
2. Testing and screening should begin at the lowest level of integration and be *selective*, i.e. consider the effectiveness of each test or screening step.
3. *Qualification tests* on prototypes are important to eliminate *defects* and *systematic failures*, they should include performance, environmental, and reliability tests.
4. Testing and screening should be carefully planned to ensure *high interpretability* of the results, and be supported by a *quality data reporting system* (Fig. 1.8).
5. Problems of testing and screening should be discussed early in the design phase, during *preliminary design reviews*.

Figure 8.1 can be used as a starting point for the development of a *test and screening strategy* at assembly level. The basic relationship between *test strategy and costs* is illustrated in the example of Fig. 8.2, in which two different strategies

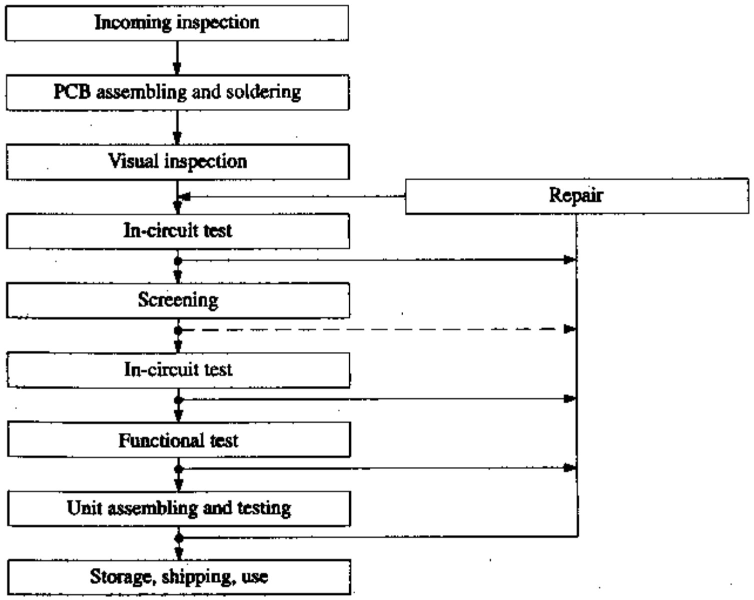


Figure 8.1 Flow chart as a basis for a test and screening strategy for electronic assemblies (PCBs)

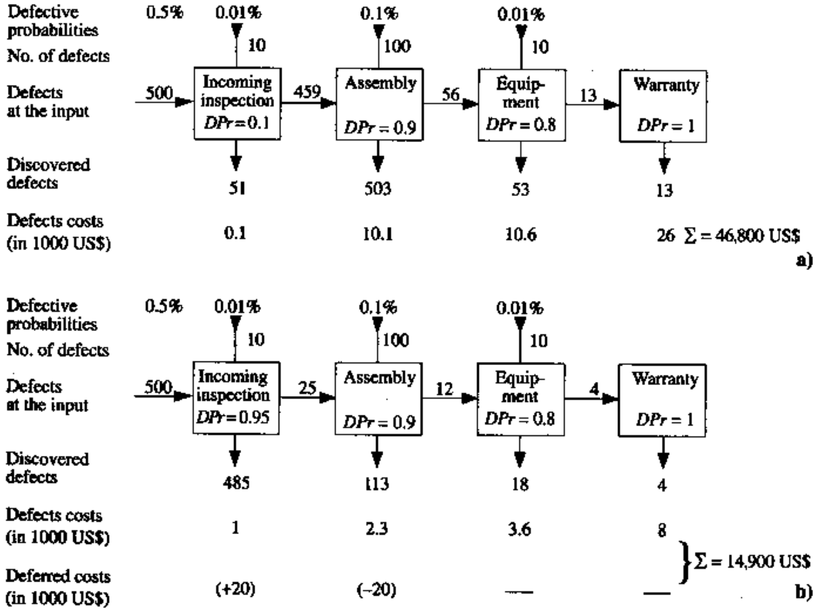


Figure 8.2 Comparison between two test strategies (figures for defects and costs have to be considered as expected values, the arithmetic mean on the basis of 100,000 ICs at the input is used for convenience only): **a)** Emphasis on assembly test; **b)** Emphasis on incoming inspection (*DPr* = detection probability, i.e. probability of finding and eliminating a defective IC)

are compared.

Both cases in Fig. 8.2 deal with the production of a stated quantity of equipment or systems for which a total of 100,000 ICs of a given type are necessary. The ICs are delivered with a *defective probability* $p = 0.5\%$. During the production, additional defects occur as a result of incorrect handling, mounting, etc., with probabilities of 0.01% at incoming inspection, 0.1% at assembly level, and 0.01% at equipment level. The cost of eliminating a defective IC is assumed to be \$2 (US\$) at incoming inspection, \$20 at assembly level, \$200 at equipment level, and \$2,000 during warranty. The two test strategies differ in the probability (DPr) of discovering and eliminating a defect. This probability is for the four levels 0.1, 0.9, 0.8, 1.0 in the first strategy and 0.95, 0.9, 0.8, 1.0 in the second strategy. It is assumed, in this example, that the additional costs to improve the detection probability at incoming inspection (\$0.20 per IC) are compensated by the savings in the test at the assembly level (giving −\$20,000). As Fig. 8.2 shows, total costs of the second test strategy are (for this example) lower than those of the first one.

Number of defects and costs are in *all* this kind of considerations *expected values* (means of random variables). The use of arithmetic means in the example of Fig. 8.2, on the basis of 100,000 ICs at the input, is for convenience only.

Models like that of Fig. 8.2 can be used to identify *weak points* in the production process (e.g. with respect to the defective probabilities at the different production steps) or to evaluate the effectiveness of additional measures introduced to decrease quality costs.

8.4.2 Quality Cost Optimization at Incoming Inspection Level

In this section, optimization of quality costs in the context of a *testing and screening strategy* is solved for the case of the choice whether a 100% *incoming inspection* or an incoming inspection on a sampling basis is more cost effective. Two cases will be distinguished, incoming inspection without screening (test only, see Figs. 8.3 and 8.4) and incoming inspection with screening (test and screening, see Figs. 8.5 and 8.6). The following notation is used:

A_t = probability of acceptance at the sampling test (i.e. probability of having no more than c defective components in a sample of size n (function of p_d, given by Eq. (A6.121) with $p = p_d$ and k=c, see also Fig. 7.2 or Fig. A8.7)
A_s = same as A_t, but for screening (screening with test)
c_d = deferred cost per defective component
c_f = deferred cost per component with early failure
c_r = replacement cost per component at the incoming inspection
c_t = testing cost per component (test only)
c_s = screening cost per component (screening with test)

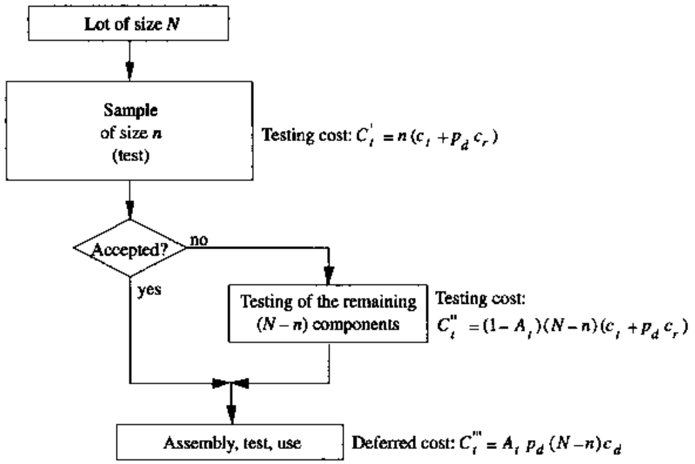


Figure 8.3 Model for quality cost optimization (direct and deferred cost) at the incoming inspection without screening of a lot of N components (all costs are expected values)

C_t = expected value (mean) of the total costs (direct and deferred) for incoming inspection without screening (test only) of a lot of N components
C_s = expected value (mean) of the total costs (direct and deferred) for incoming inspection with screening (screening with test) of a lot of N components
n = sample size
N = lot size
p_d = defective probability (defects are detected at the test)
p_f = probability for an early failure (early failures are precipitated by the screening)

Consider first the *incoming inspection without screening* (test only). The corresponding model is shown in Fig. 8.3. From Fig. 8.3, the following *cost equation* can be established for the expected value (mean) of the total costs C_t

$$\begin{aligned} C_t &= C_t' + C_t'' + C_t''' \\ &= n(c_t + p_d c_r) + (N-n)(1-A_t)(c_t + p_d c_r) + (N-n) A_t\, p_d\, c_d \\ &= N(c_t + p_d c_r) + (N-n) A_t [p_d c_d - (c_t + p_d c_r)]. \end{aligned} \tag{8.1}$$

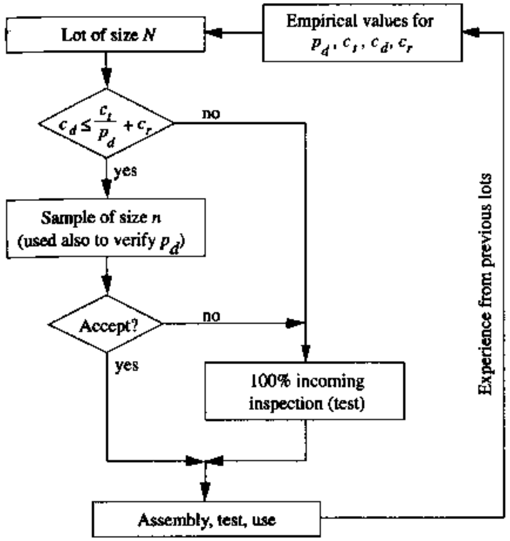


Figure 8.4 Practical realization of the procedure described by the model of Fig. 8.3

Investigating Eq. (8.1) leads to the following cases:

1. For $p_d = 0$, $A_t = 1$ and thus

$$C_t = n c_t. \tag{8.2}$$

2. For a 100% incoming inspection, $n = N$ and thus

$$C_t = N(c_t + p_d c_r). \tag{8.3}$$

3. For

$$c_d < c_r + \frac{c_t}{p_d} \tag{8.4}$$

it follows

$$C_t < N(c_t + p_d c_r)$$

and thus *a sampling test is more cost effective*.

4. For

$$c_d > c_r + \frac{c_t}{p_d} \tag{8.5}$$

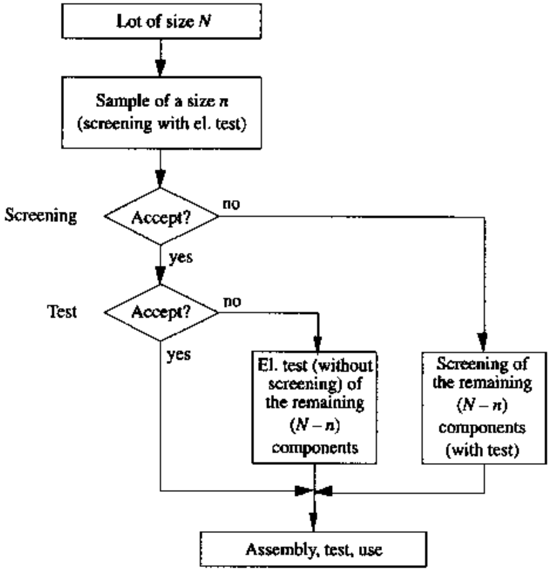


Figure 8.5 Model for quality costs optimization (direct and deferred cost) at the incoming inspection with screening of a lot of N components (all costs are expected values)

it follows

$$C_t > N(c_t + p_d c_r)$$

and thus a 100% incoming inspection is more cost effective.

The practical realization of the procedure according to the model of Fig. 8.3 is given in Fig. 8.4. The sample of size n to be tested instead of the 100% incoming inspection if the inequality (8.4) is fulfilled, is also used to verify the value of p_d, which for the actual lot can differ from the assumed one. A table of *AQL-values* (Table 7.1) can be used to determine values for n and c of the sampling plan, $AQL = p_d$ in uncritical cases and $AQL < p_d$ if a reduction for the risk of *deferred costs* is desired.

As a second case, let us consider the situation of an *incoming inspection with screening* (Section 8.2). Figure 8.5 gives the corresponding model and leads to the following *cost equation*

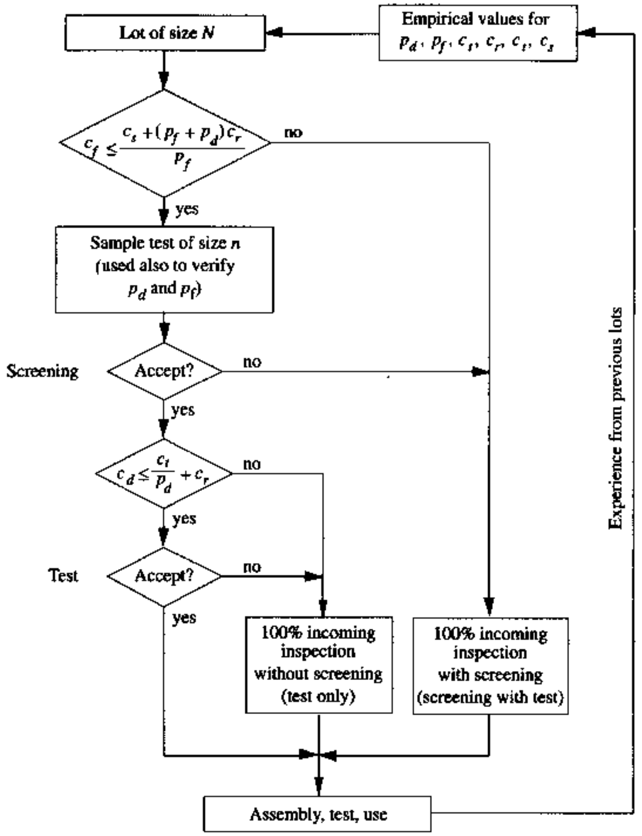


Figure 8.6 Practical realization of the procedure described by the model of Fig. 8.5

$$
\begin{aligned}
C_s &= n[c_s + (p_f + p_d)c_r] + (N-n)A_s[p_f c_f + A_t p_d c_d + (1-A_t)(c_t + p_d c_r)] \\
&\quad + (N-n)(1-A_s)[c_s + (p_f + p_d)c_r] \\
&= N[c_s + (p_f + p_d)c_r] + (N-n)A_s[p_f c_f + A_t p_d c_d + (1-A_t)(c_t + p_d c_r) \\
&\quad - (c_s + (p_f + p_d)c_r)].
\end{aligned} \tag{8.6}
$$

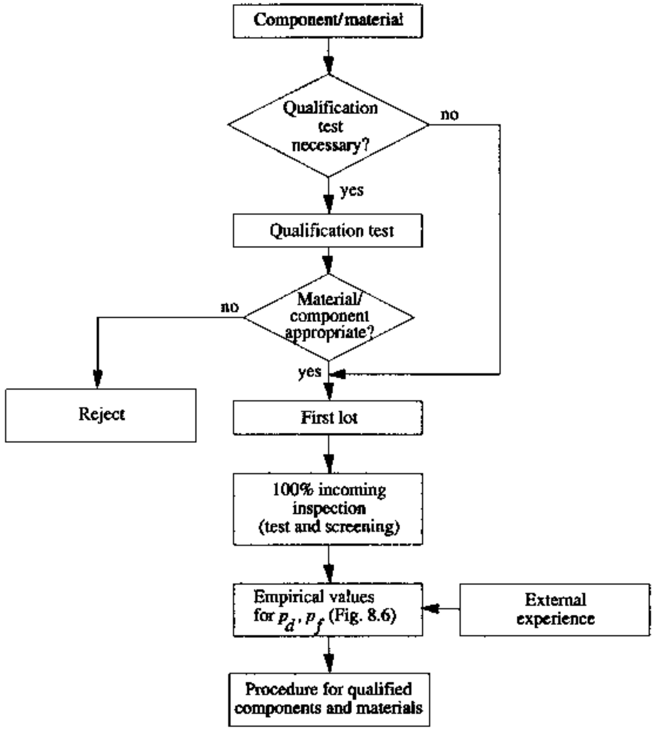


Figure 8.7 Selection procedure for nonqualified components and materials

The same considerations as with Eqs. (8.2) – (8.5) lead to the conclusion that if

$$
p_f c_f + A_t p_d c_d + (1-A_t)(c_t + p_d c_r) < c_s + (p_f + p_d)c_r \tag{8.7}
$$

holds, then a sampling screening (with test) is more *cost effective* than a 100% screening. The practical realization of the procedure according to the model of Fig. 8.5 is given in Fig. 8.6. As in Fig. 8.4, the sample of size n to be screened instead of the 100% screening if the inequality (8.7) is fulfilled, is also used to verify the values of p_f and p_d, which for the actual lot can differ from the assumed ones. The lower part on the left-hand side of Fig. 8.6 is identical to Fig. 8.4. The first inequality in Fig. 8.6 follows from inequality (8.7) with the assumption $p_f c_f >> A_t p_d c_d + (1-A_t)(c_t + p_d c_r)$, valid for $A_t \to 1$. The second inequality in Fig. 8.6 refers to the cost for incoming inspection without screening (inequality (8.4)).

8.4.3 Procedure to handle first deliveries

Components, materials, and externally manufactured subassemblies or assemblies should be submitted at the *first delivery* to an appropriate selection procedure. Part of this procedure can be performed in cooperation with the manufacturer to avoid duplication of efforts. Figure 8.7 gives the basic structure of such a procedure, see Sections 3.2 and 3.4 for qualification tests on components and assemblies.

8.5 Reliability Growth

In prototype qualification, the reliability of complex equipment or systems often proves to be less than expected. Discounting any imprecision of the applied model or uncertainty in data used in computing the predicted reliability (Section 2.2), such discrepancies are in general the consequence of *weaknesses* (errors or flaws) during design or production (for instance, dimensioning or cooling problems, use of components or materials with internal weaknesses, interface problems, transient phenomena, interference between hardware and software, assembly or soldering problems, damage during handling or testing, etc.). Superimposed on the *defects* and *systematic failures* caused by the above errors or flaws are the failures with *constant failure rate* (wearout should not be present at this stage). The distinction between deterministic (defect and systematic failures) and random events (early failures and failures with constant failure rate) is only possible with a *cause analysis*, and is important because of the different actions necessary to eliminate the fault observed: *change* or *redesign* for defects and systematic failures, *screening* for *early failures*, and *repair* for failures with constant failure rate (defects can also be randomly distributed, e.g. those caused by a time-limited flaw in the production, but still differing from failures because they are present before the item is switched-on for test or operation).

The aim of a *reliability growth program* is the *cost-effective* improvement of an item's reliability through successful correction of design or production weaknesses, see Fig. 8.8. As flaws found during reliability growth are often *deterministic* (defects and systematic failures) and thus present in every item of a given lot, reliability growth is often performed during *pilot production*, seldom for series-produced items. Similarly to *environmental stress screening* (ESS), stresses during reliability growth are in general higher than those expected in the field. A large number of models have been proposed to describe reliability growth, see e.g. [5.49, 5.51, 8.41 to 8.55, A2.5 (61014/61164)], some of them on a purely mathematical basis. A realistic model, proposed (as deterministic model) by J.T. Duane [8.46] and refined (as a statistical model) by L.H. Crow [8.45 (1975)], known also as the *AMSAA model*, assumes that the *flow of events* (defects and systematic failures) constitutes a *nonhomogenous Poisson process* with *intensity* (Eq. (A7.44))

$$\mathrm{m}(t) = \frac{d\,\mathrm{M}(t)}{dt} = \alpha\beta t^{\beta-1}, \qquad 0 < \beta < 1. \tag{8.8}$$

m(t) has the same analytical form as the failure rate $\lambda(t)$ in the case of a Weibull distribution. However, the two quantities are *fundamentally different*. Assuming that the underlying model is described by Eq. (8.8), the parameters α and β can be estimated from data. If in the cumulative operating time T, n events have occurred at the times $t_1^* < t_2^* < \ldots < t_n^*$ then, noting that for a nonhomogenous Poisson process

$$\Pr\{k \text{ events in } (a, b]\} = \frac{(\mathrm{M}(b) - \mathrm{M}(a))^k}{k!} e^{-(\mathrm{M}(b)-\mathrm{M}(a))} \tag{8.9}$$

holds for any $k = 0, 1, \ldots$, independently of the number and distribution of events outside $(a, b]$, the following *likelihood function* (Eq. (A8.24)) can be found for the estimation of the parameters α and β

$$\begin{aligned} \mathrm{L} &= \mathrm{m}(t_1^*)e^{-\mathrm{M}(t_1^*)}\,\mathrm{m}(t_2^*)e^{-(\mathrm{M}(t_2^*)-\mathrm{M}(t_1^*))} \ldots \mathrm{m}(t_n^*)e^{-(\mathrm{M}(t_n^*)-\mathrm{M}(t_{n-1}^*))}\,e^{-(\mathrm{M}(T)-\mathrm{M}(t_n^*))} \\ &= \prod_{i=1}^{n} \mathrm{m}(t_i^*)e^{-\mathrm{M}(T)} = \alpha^n \beta^n e^{-\alpha T^\beta} \prod_{i=1}^{n} t_i^{*\beta-1}, \end{aligned} \tag{8.10}$$

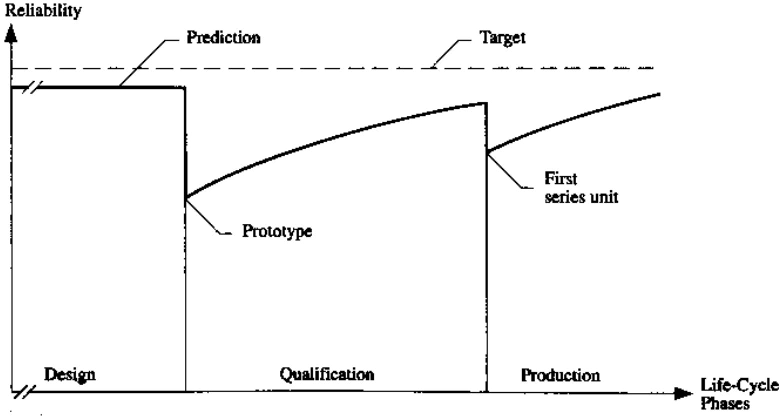


Figure 8.8 Reliability Growth (the difference between Prediction and Target is not relevant here)

or

$$\ln \mathrm{L} = n \ln(\alpha\beta) - \alpha T^{\beta} + (\beta - 1)\sum_{i=1}^{n} \ln(t_i^*). \tag{8.11}$$

The *maximum likelihood estimates* $\hat{\alpha}$ and $\hat{\beta}$ of the parameter α and β are then obtained from

$$\left.\frac{\partial \ln \mathrm{L}}{\partial \alpha}\right|_{\alpha=\hat{\alpha}} = 0 \qquad \text{and} \qquad \left.\frac{\partial \ln \mathrm{L}}{\partial \beta}\right|_{\beta=\hat{\beta}} = 0 \tag{8.12}$$

yielding

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ln \frac{T}{t_i^*}} \qquad \text{and} \qquad \hat{\alpha} = \frac{n}{T^{\hat{\beta}}}. \tag{8.13}$$

As estimate for the *intensity* of the underlying nonhomogeneous Poisson process follows as

$$\hat{\mathrm{m}}(t) = \hat{\alpha}\hat{\beta}t^{\hat{\beta}-1}. \tag{8.14}$$

With known values for $\hat{\alpha}$ and $\hat{\beta}$, Eq. (8.14) can be used to extrapolate the *attainable* intensity if the reliability growth process would be continued after T (say for a further time interval T_1) with the same statistical properties (Example 8.1).

Example 8.1

During the reliability growth program of some complex equipment, the following data was gathered: $T = 1200\,\mathrm{h}$, $n = 8$ and $\sum \ln(T / t_i^*) = 20$. Assuming that the underlying process can be described by a Duane model, estimate the density $\mathrm{m}(t)$ for $t = 1200\,\mathrm{h}$ and the value attainable for $t = 3000\,\mathrm{h}$.

Solution

With $T = 1200\,\mathrm{h}$, $n = 8$ and $\sum \ln(T / t_i^*) = 20$, it follows from Eq. (8.13) that $\hat{\beta} = 0.4$ and $\hat{\alpha} \approx 0.47$, and from Eq. (8.14) the estimate for the intensity $\mathrm{m}(t)$ for $t = 1200\,\mathrm{h}$ is then $\hat{\mathrm{m}}(1200) \approx 2.67 \cdot 10^{-3}\,\mathrm{h}^{-1}$. The attainable intensity after an extension of the program for reliability growth by $1800\,\mathrm{h}$ of operation is given (assuming that the statistical properties of the underlying stochastic process remain unaltered) by Eq. (8.14) with $\hat{\alpha} \approx 0.47$, $\hat{\beta} \approx 0.4$ and $t = 3000\,\mathrm{h}$ h as $\hat{\mathrm{m}}(3000) \approx 1.54 \cdot 10^{-3}\,\mathrm{h}^{-1}$.

For the mean value $\mathrm{T}(t)$ of the random time from an arbitrary time point t to the next defect or systematic failure on the time axis, the following holds (because of the assumed nonhomogeneous Poisson process)

$$\mathrm{T}(t) = \int_0^{\infty} \Pr\{\text{no event in } (t, t + x]\}\,dx = \int_0^{\infty} e^{-(\mathrm{M}(t+x) - \mathrm{M}(t))}dx = e^{\mathrm{M}(t)} \int_t^{\infty} e^{-\mathrm{M}(y)}dy, \tag{8.15}$$

giving e.g. $\mathrm{T}(t) = 1/\lambda$ for $\mathrm{M}(x) = \mathrm{M}(t) + (x - t)\lambda$, or $\mathrm{m}(x) = \lambda$, for $x > t$.

The statistical methods used to investigate *reliability growth models* are in general *basically different* from those used in Section 7.2.2 for the *homogenous Poisson process*. This is because nonhomogenous Poisson processes are *not renewal processes*, thus the use in Eqs. (8.10) to (8.13) of $t_1^*, \ldots, t_n^*$ to distinguish them from $t_1, \ldots, t_n$ occurring as n independent observations of a random variable τ.

The accuracy of the *Duane model* is often sufficiently good for electronic, electromechanical, and mechanical equipment and systems. It can also be used to describe the occurrence of *software defects* (*dynamic defects*), often appearing as failures only because of the software complexity. However, other models have been discussed in the literature especially for software (Section 5.3.4). Among these, the *logarithmic Poisson model*, which assumes a *nonhomogenous Poisson process* with intensity

$$\mathrm{m}(t) = \frac{1}{\delta + \gamma t} \qquad \text{or} \qquad \mathrm{m}(t) = \frac{\alpha + 1}{\beta + t}, \qquad 0 < \alpha, \beta, \delta, \gamma < \infty. \tag{8.16}$$

For the logarithmic Poisson model it holds that $d\,\mathrm{m}(t)/dt < 0$, with $\mathrm{m}(\infty) = 0$ and $\mathrm{m}(0) < \infty$. The last can be an advantage with respect to the Duane model, for which $\mathrm{m}(0) = \infty$. Assuming $\mathrm{M}(0) = 0$, it follows

$$\mathrm{M}(t) = \frac{\ln(1 + \gamma t / \delta)}{\gamma} \qquad \text{or} \qquad \mathrm{M}(t) = \ln\left(\frac{\beta + t}{\beta}\right)^{\alpha+1} \tag{8.17}$$

Promising for hardware and/or software are two models investigated recently [8.43], which assume

$$\mathrm{M}(t) = a \ln(1 + \frac{t}{b}) \cdot (1 - e^{-\frac{t}{b}}), \qquad 0 < a, b < \infty \tag{8.18}$$

and

$$\mathrm{M}(t) = \alpha t^{\beta}[1 - (1 + \frac{t}{\gamma})e^{-\frac{t}{\gamma}}], \qquad 0 < \alpha, \gamma < \infty,\ 0 < \beta < 1\ , \tag{8.19}$$

respectively, i.e. which combine in a multiplicative way two possible $\mathrm{M}(t)$. The corresponding *intensities* are

$$m(t) = \frac{a}{b+t} + [\frac{a}{b}\ln(1+\frac{t}{b}) - \frac{a}{b+t}]e^{-\frac{t}{b}} \tag{8.20}$$

and

$$m(t) = \alpha\beta t^{\beta-1}[1-(1+\frac{t}{\gamma}+\frac{t^2}{\beta\gamma^2})e^{-\frac{t}{\gamma}}], \tag{8.21}$$

respectively. In both cases it holds that $m(0) = 0$, $m(t)$ grows then to a maximum, from which it goes to 0 with a shape similar to that of previous models. The models described by Eqs. (8.18) and (8.20) fit well with most of the data sets known in the literature [8.43]. However, in general it is not possible to fix a priori the "best" model to be used in a given situation. A *physical explanation* of the model used could help in such a choice.

A1 Terms and Definitions

This appendix provides definitions and comments on the terms most commonly used in reliability engineering, see Fig. A1.1 for a classification of these terms. Table 5.4 extends this appendix to the terms used in software quality. Relevant standards [A1.1 to A1.6] and recent trends have been considered.

System, Systems Engineering, Concurrent Engineering, Cost Effectiveness, Quality
- Capability
- Availability, Dependability
  - Reliability
    - Item
    - Required Function, Mission Profile
    - Reliability Block Diagram, Redundancy
    - MTTF, MTBF
    - Failure, Failure Rate, Derating
    - FMEA, FMECA, FTA
    - Reliability Growth, Environmental Stress Screening, Burn-in
  - Maintainability
    - Preventive Maintenance, MTTPM, MTBUR
    - Corrective Maintenance, MTTR
  - Logistical Support
- Fault
  - Defect, Nonconformity
  - Systematic Failure
  - Failure
- Safety
- Total Quality Management (TQM)
  - Quality Assurance
    - Configuration Management, Design Review
    - Quality Test
    - Quality Control during Production
    - Quality Data Reporting System
- Life Time, Useful Life
- Life-Cycle Cost, Value Engineering, Value Analysis
- Product Assurance, Product Liability

Figure A1.1 Terms commonly used in reliability engineering

Availability / Point Availability (PA(t))

> Probability that an item will perform its required function under given conditions at a stated instant of time.

According to the above definition, point availability is a *characteristic* of an item, generally designated by PA(t). A *qualitative* definition is: *Ability of an item to perform a required function under given conditions at a stated instant of time.* The term item stands for an entity of arbitrary complexity (from component to system). Computation of the point availability generally assumes continuous operation (item down only for repair), complete renewal (repaired element in the reliability block diagram as-good-as-new after repair, valid for the whole item in the case of constant failure rates), and ideal human factors and logistical support. The *steady-state value* of the point availability is then given by $PA = MTTF/(MTTF + MTTR)$. PA is also equal to the steady-state value of the *average availability* (often simply designated as availability). Other kinds of availability (mission availability, work-mission availability, joint availability etc.) can be defined (point availability was formerly also known as instantaneous or pointwise availability).

Burn-in (nonrepairable items)

> Type of screening test while an item is in operation.

For electronic devices, stresses during burn-in are often constant higher ambient temperature (e.g. 125°C for ICs) and constant higher supply voltage. Burn-in can be considered as a part of a *screening procedure*, performed on a 100% basis to *provoke early failures* and to stabilize the characteristics of an item. Often it can be used as an *accelerated reliability test* to investigate the item's failure rate. Stresses are then higher than would be expected in field operation, but not so high as to stimulate failure mechanisms which would not occur in normal use.

Burn-in (repairable items)

> Operation of an item in a prescribed environment with successive corrective maintenance at every failure during the early failure period.

For large equipment or systems, the term *run-in* is often used instead of *burn-in*. The stress conditions have to be chosen as near as possible to those expected in *field operation*. Flaws detected during burn-in can be deterministic (defects or systematic failures) during the pilot production (reliability growth), but should be attributable only to *early failures* (randomly distributed) during the series production.

Capability (Performance)

> Ability of an item to meet a service demand of stated quantitative characteristics under given conditions.

Performance (technical performance) is often used instead of capability.

Concurrent Engineering

> Systematic approach to reduce the time to develop and market an item, a.o. by integrating production activities into the design and development phase.

Concurrent engineering is achieved through intensive *teamwork* between all engineers involved in the design, production, and marketing of an item. It supports TQM and quality assurance, and also has a positive influence on the optimization of *life-cycle cost*.

Configuration Management

> Procedure to specify, describe, audit, and release the configuration of an item, as well as to control it during modifications or changes.

Configuration includes all of an item's *functional* and *physical* characteristics as given in the documentation (to specify, build, test, accept, operate, maintain, and logistically support the item) and as present in the hardware and/or software. Configuration management is subdivided into *configuration identification, auditing, control*, and *accounting*. Configuration management is a part of *quality assurance* and *TQM*, particularly important during the design and evaluation phase.

Corrective Maintenance

> Maintenance carried out after recognition of a fault, intended to put an item back into a state in which it can again perform its required function.

Corrective maintenance is also known as *repair*. It can include any or all of the following steps: *localization, isolation, disassembly, exchange, reassembly, alignment*, and *checkout*. To simplify computations it is generally assumed that the repaired element in the reliability block diagram is *as-good-as-new* after each repair (also including a possible environmental stress screening of the spare parts). This assumption applies to the *whole item* (equipment or system) if all components of the item (which have not been repaired/renewed) have *constant failure rate*.

Cost Effectiveness

> Measure of the ability of an item to meet a service demand of stated quantitative characteristics, with the best possible usefulness to life-cycle cost ratio.

The term *system effectiveness* is often used instead of *cost effectiveness*.

Defect

> Nonfulfillment of an intended usage requirement or reasonable expectation, essentially present at $t = 0$, when the item is put in operation.

From a technical point of view, a *defect* is similar to a *nonconformity*, however not necessarily from a legal point of view. Defects do not need to influence the item's functionality. They are caused by *flaws* (errors, mistakes) during design, development, production, or installation. The term defect should be preferred to that of *error*, which is a *cause*. Unlike *failures*, which always appear *in time* (generally randomly distributed), *defects* are present at $t = 0$. However, some defects can only be detected when the item is operating and are referred to as *dynamic defects* (in software). Similar to defects, with regard to causes, are *systematic failures*, however, they are generally not present at $t=0$.

Dependability

> Collective term used to describe the availability performance and its influencing factors, such as reliability performance, maintainability performance, and logistical support performance.

Dependability is used in a qualitative sense only.

Derating

> Nonutilization of the full load capability of an item with the intent to reduce the failure rate.

The *stress factor S* expresses the ratio of actual to rated load under normal operating conditions (generally at 25°C ambient temperature).

Design Review

> A formal documented, comprehensive, and systematic examination of a design to evaluate the capability of the design to meet all requirements, to identify problems, and propose solutions.

Design reviews are an important tool of *quality assurance* and *TQM* during the design and development of hardware and software (Tables A3.2 and 5.3). The term *design* is used here in a broad sense (design reviews cover all phases from the definition to the pilot production).

Environmental Stress Screening (ESS)

> Test or set of tests intended to remove defective items, or those likely to exhibit early failures.

ESS is a screening procedure often performed at assembly (PCB) or equipment level on a 100% basis to find *defects* and *systematic failures* during the pilot production (reliability growth), or to provoke *early failures* in a series production. It consists mainly of temperature cycles and/or random vibrations. Stresses are in general *higher* than those expected in field operation. Experience shows that to be cost effective, ESS has to be *tailored* to the item and production processes considered. At component level, the term *screening* is used instead of ESS.

Failure

> Termination of the ability of an item to perform a required function.

Failures should be considered (classified) with respect to the *mode*, *cause*, *effect*, and *mechanism*. The *cause* of a failure can be *intrinsic* (early failure, failure with constant failure rate, and wearout failure) or *extrinsic* (systematic failure, i. e. failure resulting from errors or mistakes in design, production, or operation which are *deterministic* and has to be considered as a defect). The *effect* (consequence) of a failure is generally different if considered on the directly affected item or on a higher level. A failure is an *event* appearing in time (randomly distributed), in contrast to a *fault* which can be a *state*.

Failure Rate ($\lambda(t)$)

> Limit for $\delta t \to 0$ of the probability that an item will fail in the time interval $(t, t+\delta t]$, given that the item was new at $t = 0$ and did not fail in the interval $(0, t]$, divided by δt.

The failure rate is generally designated by $\lambda(t)$. If τ is the item *failure-free operating time*, then

$$\lambda(t) = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{t < \tau \le t + \delta t \mid \tau > t\} = \frac{f(t)}{1 - F(t)} = -\frac{d\,R(t)/dt}{R(t)}.$$

The existence of $f(t)$ is tacitly assumed. For $R(0) = 1$, it follows that $R(t) = e^{-\int_0^t \lambda(x)\,dx}$. Thus, for $\lambda(t) = \lambda$, it holds that $R(t) = e^{-\lambda t}$. *Only in this case* one can estimate the failure rate λ by $\hat{\lambda} = k/T$, where T is the given, *fixed* cumulative operating time (cumulated over an arbitrary number of statistically identical items) and k the total number of failures during T. In general, the failure rate of a large population of *statistically identical* (independent) items exhibits the typical form of a bathtub curve in which the phases of *early failures*, *failures with nearly constant failure rate* λ, and *wearout failures* can be distinguished. The term random failures for the period with constant failure rate should be avoided, as only systematic failures have a deterministic character; also the terms hazard rate and instantaneous failure rate should be omitted, to avoid confusion.

Fault

> State of an item, characterized by the inability to perform a required function, excluding inabilities due to preventive maintenance, other planned actions, or lack of external resources.

A fault can be a *defect* or a *failure*, having thus as possible cause an *error* (for defects or systematic failures) or a *failure mechanism* (for failures).

Fault Modes and Effects Analysis (FMEA)

> Qualitative method of reliability analysis which involves for each element of an item the investigation of all possible fault modes, and of the corresponding effects on other elements as well as on the required function of the item.

See FMECA.

Fault Modes, Effects, and Criticality Analysis (FMECA)

> Qualitative/quantitative method of reliability analysis which includes the analysis of fault modes and effects (FMEA) while considering for each fault mode the probability of occurrence and the ranking of its severity.

Goal of an FMEA/FMECA is to determine all potential hazards and to analyze the possibilities of reducing their effect, or their probability of occurrence. *All possible* failure (and defect) modes and causes have to be considered *bottom-up* from the lowest to the highest integration level of the item considered. FMECA was formerly used for *failure modes, effects, and criticality analysis.* Often one distinguishes between a design and a production FMEA or FMECA.

Fault Tree Analysis (FTA)

> Analysis to determine which fault modes of the elements of an item and/or which external events may result in a stated fault mode of the item, presented in the form of a fault tree.

FTA is a *top-down* approach, which allows the inclusion of external causes more easily than an FMEA/ FMECA. A graphical description of cause-to-effect relationships, which combines and can in some cases extend FMEA/FMECA and FTA procedures, is the *cause-and-effect diagram,* also known as a *fishbone* or *Ishikawa diagram.*

Item

> Any component, device, assembly, equipment, subsystem, or system that can be considered individually.

An item is a functional or structural *unit,* which is considered as an *entity* for investigations. It may consist of hardware and/or software and also include human resources (to emphasize this fact, the term *system* has been defined separately).

Life Cycle Cost (LCC)

> Sum of the costs for acquisition, operation, maintenance, and disposal or recycling of an item.

Life-cycle cost optimization is undertaken within the framework of *cost effectiveness* or *systems engineering* and can be positively influenced by *concurrent engineering.* International regulations will take more and more into account the effects to the environment of the production, use, and ecologically acceptable disposal or recycling of an item, when considering life-cycle costs.

Life Time

> Time span between initial operation and failure of a nonrepairable item.

Logistical Support

> All activities undertaken to provide effective and economical use of an item during its operating phase.

Logistical support is no longer reserved to the defense field. An emerging aspect related to maintenance and logistical support is that of *obsolescence management,* i.e. how to assure operation over for instance 20 years *when technology is rapidly evolving* and components need for maintenance are no longer manufactured (say 5 years after the time the equipment has been put into operation).

Maintainability

> Probability that preventive maintenance or repair of an item will be performed within a stated time interval for given procedures and resources.

According to the above definition, maintainability is a *characteristic* of an item. A *qualitative* definition is: *Ability of an item to be retained in, or restored to a specified state in a given time interval under stated procedures and resources.* Maintainability is subdivided into *serviceability* (preventive maintenance) and *reparability* (corrective maintenance or repair). In specifying or evaluating maintainability, it is important to consider the logistical support, i.e. procedures, personnel (number, skill level), and other resources (spare parts, test facilities), available for maintenance.

Mission Profile

> Specific task which must be fulfilled by an item during a stated time under given conditions.

The mission profile defines the *required function* and the *environmental conditions* as a function of time. A representative mission profile and the corresponding reliability targets have to be defined in the item specifications.

MTBF

> $MTBF = 1/\lambda$.

MTBF stands for mean operating time between failures (formerly, mean time between failures). *MTBF* should be *reserved* for items with *constant* failure rate λ. In this case $R(t) = e^{-\lambda t}$, and $MTBF = 1/\lambda$ is the expected value (mean) of the item's failure-free operating time (see MTTF). The definition given here agrees with the statistical methods often used to estimate or demonstrate an *MTBF*; in particular $\hat{MTBF} = T/k$, where T is the given, *fixed* cumulative operating time (cumulated over an arbitrary number of statistically identical items) and k the total number of failures during T. The use of *MTBF* as the mean time between consecutive failures of a repairable item (i.e. the sum of a failure-free operating time and a repair time) should be avoided.

MTTF

> Expected value (mean) of an item's failure-free operating time.

MTTF stands for *mean time to failure. MTTF* is obtained from the reliability function R(t) as $MTTF = \int_0^\infty R(t)\,dt$, with T_L as the upper limit of the integral, if the life time is limited to T_L ($R(t) = 0$ for $t \geq T_L$). *MTTF* applies both to *nonrepairable* and to *repairable* items if one assumes that after a repair the item is *as-good-as-new*; if this is not the case, a new *MTTF* can be considered ($MTTF_{Si}$ in Table 6.2). An unbiased (empirical) estimate for *MTTF* is $\hat{MTTF} = (t_1 + \ldots + t_n)/n$, where $t_1, \ldots, t_n$ are observed failure-free operating times of statistically identical items.

MTTPM

> Expected value (mean) of an item's preventive maintenance time.

MTTPM stands for *mean time to preventive maintenance*, see *MTTR* for further remarks.

MTBUR

> Expected value (mean) of the time between unscheduled removals.

MTBUR stands for *mean time between unscheduled removals*. It is used for instance in avionic, where an estimate of *MTBUR* is obtained by dividing the total number of flight hours logged by all units of a given airplane, over a certain period of time, by the number of unscheduled removals during the same period of time.

MTTR

> Expected value (mean) of an item's repair time.

MTTR stands for *mean time to repair. MTTR* is obtained from the distribution function G(t) of the repair time as $MTTR = \int_0^\infty (1 - G(t))\,dt$. In specifying or evaluating *MTTR*, it is necessary to consider the logistical support, i.e. procedures, personnel (number, skill level), and other resources (spare parts, test facilities) available for repair. Repair times are often *lognormally distributed*. However, for reliability or availability computations of repairable equipment and systems, a *constant repair rate* μ (i.e. exponentially distributed repair times with $\mu = 1/MTTR$) can generally be used (for $MTTR \ll MTTF$) to get approximate results. An unbiased (empirical) estimate of *MTTR* is $\hat{MTTR} = (t_1 + \ldots + t_n)/n$, where $t_1, \ldots, t_n$ are observed repair times of statistically identical items.

Nonconformity

> Nonfulfillment of a specified requirement.

From a technical point of view, the term *nonconformity* is close to that of *defect*, however not necessarily from a legal point of view. Nonconformity was formerly defined as *deviation of an item's characteristic from the specified value*.

Preventive Maintenance

> Maintenance carried out at predetermined intervals and according to prescribed procedures to reduce the probability of failure or the degradation of the functionality of an item.

The aim of preventive maintenance must also be to detect and remove *hidden failures*, i.e. failures in redundant elements. To simplify computations it is generally assumed that the item (element in the reliability block diagram for which a preventive maintenance has been performed) is *as-good-as-new* after each preventive maintenance. This assumption applies to the *whole item* (equipment or system) if all components of the item (which have not been renewed) have *constant failure rate*.

Product Assurance

> All planned and systematic activities necessary to reach specified targets for the reliability, maintainability, availability, and safety of an item, as well as to provide adequate confidence that the item will meet given requirements for quality.

The concept of product assurance is currently used in aerospace programs. It includes besides quality assurance, also reliability, maintainability, availability, safety, and logistical support engineering.

Product Liability

Onus on a manufacturer or others to compensate for losses related to injury to persons, material damage, or other unacceptable consequences caused by a product.

The manufacturer (producer) usually specifies a safe operational mode for the product (item). Basically, *strict liability* is applied. In the case of a claim, the manufacturer has then to demonstrate that the product was free from defects when it left the production plant. This holds in particular in the USA, but partially also in Europe [1.13, 1.19]. However, in Europe the causality between damage and defect or failure has still to be demonstrated by the user and the *limitation period* is short (generally 3 years after the identification of the damage, defect, and manufacturer or 10 years after the appearance of the product on the market). Liability is today mainly limited to hardware. Product liability forces producers to place greater emphasis on quality assurance.

Quality

Totality of features and characteristics of an item (product or service) that bear on its ability to satisfy stated or implied needs.

This definition is general, accounting for all objective and subjective attributes or characteristics of a product or service. Quality was in the ninteen-sixties defined as *fitness for use* and then as the *degree to which an item conforms to applicable specifications*, a definition often used for software quality.

Quality Assurance (Hardware and Software)

All planned and systematic activities necessary to provide adequate confidence that an item (product or service) will satisfy given requirements for quality.

Quality assurance is used in this book in the sense also of *quality management* as per TQM. It includes *configuration management*, *quality tests*, *quality control during production*, and *quality data reporting systems*, (Fig. 1.3). For complex equipment and systems, quality assurance activities are coordinated by a *quality assurance program* (Appendix A3). An important target of quality assurance is to achieve the quality requirements with a minimum of *cost* and *time*. *Concurrent engineering* activities also strive to short the time to develop and market a product. A coordinated extension of the quality assurance activities to *everyone* involved in the conception, development, production, distribution, and use of a product or service leads to the concept of *total quality management* (TQM).

Quality Control During Production

Control of the production processes and procedures to reach a stated quality of manufacturing.

Quality control during production is a part of *quality assurance* and TQM.

Quality Data Reporting System

System to collect, analyze, and correct all defects and failures occurring during production and testing of an item, as well as to evaluate and feedback the corresponding quality and reliability data.

A quality data reporting system is generally computer aided. Analysis of defects and failures must be traced to the *cause* in order to determine the best *corrective action* necessary to avoid repetition of the same problem. The quality data reporting system should also remain active during the *operating phase*. It is a part of *quality assurance* and TQM. A quality data reporting system is important to monitor *reliability growth* during the production of hardware and can also be used for software.

Quality Test

Test to verify whether an item conforms to specified requirements.

Quality tests include incoming inspections, qualification tests, production tests, and acceptance tests. They also cover the *reliability*, *maintainability*, and *safety* aspects. To be cost effective, quality tests must be coordinated and integrated in a *test (and screening) strategy*. Quality tests are a part of *quality assurance* and TQM.

Redundancy

Existence of more than one means for performing a required function in an item.

For hardware, distinction is made between *active* (hot, parallel), *warm* (lightly loaded), and *standby* (cold) redundancy. Redundancy does not necessarily imply a *duplication* of hardware, it can also be implemented for example by coding or by software. To avoid *common mode failures*, redundant elements should be realized (designed and produced) *independently* from each other. Should the redundant elements fulfill only a part of the required function, a *pseudo redundancy* is present.

Reliability (R, $R(t)$)

Probability that an item will perform its required function under given conditions for a stated time interval.

According to the above definition, reliability is a *characteristic* of an item, generally designated by R. A *qualitative* definition is: *Ability of an item to remain functional under given conditions for a stated time interval.* Reliability specifies the probability that no *operational interruption* will occur during a stated mission, say of duration T. This does not mean that redundant parts may not fail, such parts can fail and be repaired. The concept of reliability applies thus to *nonrepairable* as well as to *repairable* items. Should T be considered as a variable t, the *reliability function* is given by $R(t)$. If τ is the failure-free operating time, with distribution function $F(t)$, then $R(t) = \Pr\{\tau > t\} = 1 - F(t)$.

Reliability Block Diagram

> Block diagram showing how failures of elements, represented by the blocks, result in the failure of an item.

The reliability block diagram is an *event diagram*. It answers the following question: *Which elements of an item are necessary to fulfill the required function and which ones can fail without affecting it?* The elements which must operate are connected *in series* (the ordering of these elements is not relevant for reliability computations) and the elements which can fail (redundant elements) are connected *in parallel*. Elements which are not relevant (used) for the required function are removed from the reliability block diagram (and put into a reference list), after having verified (FMEA) that their failure does not affect elements involved in the required function. In a reliability block diagram, redundant elements still appear *in parallel*, irrespective of the failure mode. However, only *one failure mode* (e.g. short or open) and *two states* (good or failed) can be considered for each element.

Reliability Growth

> A condition characterized by a progressive improvement of the reliability of an item with time, through successful correction of design or production weaknesses.

Flaws (errors, mistakes) detected during a reliability growth program are in general *deterministic* (defects or systematic failures) and thus present in every item of a given lot. Reliability growth is often performed during the *pilot production*, seldom for series-produced items. Similarly to *environmental stress screening*, stresses during reliability growth often *exceed* those expected in field operation. Models for reliability growth can also often be used to investigate the occurrence of *defects in software*. Although software defects often appear in time (dynamic defects), the term *software reliability* should be avoided.

Required Function

> Function or combination of functions of an item which is considered necessary to provide a given service.

The definition of the required function is the starting point for *any* reliability analysis, as it defines failures. However, difficulties can appear with complex items. For practical purposes, parameters should be specified with tolerances.

Safety

> Ability of an item to cause neither injury to persons, nor significant material damage or other unacceptable consequences.

Safety is subdivided into *accident prevention* (the item is safe working while it is operating correctly) and *technical safety* (the item has to remain safe even if a failure occurs). *Technical safety* can be defined as the *probability that an item will not cause injury to persons, significant material damage or other unacceptable consequences above a given level for a stated time interval, when operating under given conditions*. Methods and procedures used to investigate technical safety are similar to those used for reliability analyses, however with emphasis on fault/failure effects.

System

> Combination of components, assemblies, and subsystems, as well as skills and techniques, capable of performing and/or supporting autonomously an operational role.

A system generally includes hardware, software, services, and personnel (for operation and support) to the degree that it can be considered self-sufficient in its intended operational environment. For calculations, ideal conditions for *human factors* and *logistical support* are often assumed, leading to a *technical system* (the term *system* is often used instead of *technical system*, for simplicity).

Systematic Failure

> Failure whose cause is a flaw (error, mistake) in the design, production, or use of an item.

Systematic failures are also known as *dynamic defects*, for instance in software quality assurance, and have a deterministic character. However, because of the item complexity they can appear as if they were randomly distributed in time.

Systems Engineering

> Application of the mathematical and physical sciences to develop systems that utilize ressources *economically* for the benefit of society.

TQM and *quality assurance* help to optimize systems engineering.

Total Quality Management (TQM)

> Management approach of an organization centered on quality, based on the participation of all its members, and aiming at long-term success through customer satisfaction and benefits to all members of the organization and to society.

With TQM, everyone involved in a product (directly during development, production, installation, servicing, or indirectly with management or staff activity) is responsible for the quality of that product.

Useful Life

Total operating time of an item, ending for a nonrepairable item when the failure probability becomes too high or the item's functionality is obsolete, and for a repairable item when the intensity of failures becomes unacceptable or when a fault occurs and the item is considered to be no longer repairable.

Typical values for useful life are 3 to 6 years for commercial applications, 5 to 15 years for military installations, and 10 to 30 years for distribution or power systems.

Value Analysis

Optimization of the configuration of an item as well as of the production processes and procedures to provide the required item characteristics at the lowest possible cost without loss of capability, reliability, maintainability, or safety.

Value Engineering

Application of value analysis methods during the design phase to optimize the life-cycle cost of an item.

A2 Quality and Reliability Standards

Besides *quantitative reliability requirements* (figures), such as *MTBF, MTTR*, or availability, as given in system specifications, customers often require a *quality assurance/management system*, and for complex items also the realization of a *quality and reliability assurance program*. Such general requirements are covered by *standards*, the most important of which are discussed in this appendix. A basic procedure for setting up and realizing quality and reliability requirements as well as the structure and content of a quality and reliability assurance program for complex equipment and systems are considered in Appendix A3.

A2.1 Introduction

Customer requirements for quality and reliability can be quantitative or qualitative. As with performance parameters, *quantitative* reliability requirements are given in system specifications or contracts. They fix targets for reliability, maintainability, and availability, along with associated information concerning the required function, operating conditions, logistical support, and criteria for acceptance tests (Appendix A3). *Qualitative* requirements are in standards and generally deal with a quality and reliability *assurance/management system*. Depending upon the field of application (aerospace, defense, nuclear, industrial), these requirements may be more or less stringent. The main objectives of such standards are:

1. Standardization of configuration, operating conditions, test procedures, selection and qualification of components/materials and production processes, logistical support, etc.
2. Harmonization of quality and reliability assurance/management systems.
3. Agreement on terms and definitions.

Important standards for quality and reliability assurance/management of equipment and systems are given in Table A2.1, see [A2.1 to A2.11] as well as [A2.21 to A2.26] for a comprehensive list. Some of the standards in Table A2.1 are discussed in the following sections.

A2.2 Requirements in the Industrial Field

The ISO 9000 family of standards [A2.8] is well established, with world wide a great number of certified manufacturers of components, equipment, and systems. In its present form it requires from a producer a *quality assurance/management system* able to cover the development, production, installation, and servicing (ISO 9001), the production and installation (ISO 9002), or the final inspection and test (ISO 9003) of a product. ISO 9001-2000 has been announced which will combine ISO 9001 to 9003. ISO 9001-2000 will differ basically in structure from ISO 9001 and focus on four main chapters (Management Responsibility, Resource Management, Product and/or Service Realization, and Measurement). A quality assurance/management system must ensure that *everyone* involved with a product (whether in its conception, development, production, installation, or servicing) shares responsibility for the quality of that product, in accordance with *Total Quality Management* (TQM) aims. At the same time, the system must be cost effective and contribute to a reduction in the time to market. Such a system must cover all aspects of a quality assurance/management system, including *organization*, *planning*, *configuration management*, *quality tests*, *quality control*, and *corrective actions*. The customer expects that only items with agreed quality and reliability will be delivered. The ISO 9000 family deals with a broad class of products and services (technical and non-technical), its content is thus lacking in details, compared with application specific standards used for instance in aerospace, the nuclear industry, and defense (Appendix A2.3). It has been accepted as a national standard in many countries. A mutual international recognition of the corresponding certification has been partly achieved today.

Dependability aspects, including reliability, maintainability, and logistical support of systems are considered in IEC standards, in particular IEC 60300, 60605, and 60706 for global requirements and IEC 60812, 60863, 61025, 61078, and 61709 for specific procedures [A2.5]. IEC 60300 deals with *dependability programs* (management, task descriptions, and application guides). Reliability tests for a *constant failure rate* λ (or $MTBF = 1/\lambda$) are considered in IEC 60605. *Maintainability* aspects are considered in IEC 60706.

For electronic equipment and systems, IEEE Std 1332-1998 [A2.6] has been issued as a guide to a reliability program for the development and production phases. This document gives in a short form the basic requirements, putting a clear accent on an *active cooperation* between supplier (as manufacturer) and customer.

Software aspects are covered by IEEE Software Engineering Standards [A2.7]. Requirements for *product liability* are given in national and international directives, for instance in the EU Directive 85/374 [1.19].

Table A2.1 Important standards for quality and reliability assurance/manag. of equipment or systems

Industrial			
1994	Int.	ISO 9000	Quality management and quality assurance standards – Guidelines for selection and use (-1 to -4)
		ISO 9001	Quality systems – Model for quality assurance in design, development, production, installation and servicing
		ISO 9002	Quality systems – Model for quality assurance in production, installation and servicing
		ISO 9003	Quality systems – Model for quality ass. in final inspection & test
		ISO 9004	Quality management and quality system elements–Guidelines 1-4
1991–97	Int.	IEC 60300	Dependability management (-1: Program management, -2: Program element tasks, -3: Application guides)
1982–97	Int.	IEC 60605	Equipment reliability testing (-1: Gen. req., -2: Test cycles, -3: Test conditions, -4: Point and interval estimates, -6: Test for constant failure rate)
1982–94	Int.	IEC 60706	Guide on maintainability of equipment (-1: Maint. program, -2: Analysis, -3: Data evaluation, -4: Support planning, -5: Diagnostic, -6: Statistical methods)
1969–98	Int.	IEC	60068, 60319, 60410, 60721, 60749, 60812, 60863, 61000, 61014, 61025, 61070, 61078, 61123, 61124, 61160, 61163, 61164, 61165, 61649, 61650, 61709
1998	Int.	IEEE Std1332	IEEE Standard Reliability Program for the Development and Production of Electronic Systems and Equipment
1985	EU	85/374	Product Liability
Software Quality			
1987-1998	USA	IEEE/ANSI Std.	IEEE Software Eng. Standards Vol. 1 to 4, 1999 (in particular 610.12, 730, 828, 829, 830, 982.1/.2, 1008, 1012, 1028, 1042, 1045, 1058, 1059, 1061, 1062, 1063, 1074, 1219, 1465); see also ISO/IEC 12207 (Software Life Cycle Processes) 1995
Defense			
1959	USA	MIL-Q-9858	Quality Program Requirements (ed. A, 1963)
1965	USA	MIL-STD-785	Rel. Program for Systems and Eq. Devel. and Prod. (ed. B, 1980)
1965	USA	MIL-STD-781	Rel. Testing for Eng. Devel., Qualif. and Prod. (ed. D, 1986)
1966	USA	MIL-STD-470	Maintainability Program for Systems and Equip. (ed. A, 1983)
1968	NATO	AQAP-1	NATO Req. for an Industrial Quality Control System (ed.3, 1984)
Aerospace			
1974	USA (NASA)	NHB-5300.4	Safety, Reliability, Maintainability, and Quality Provisions for the Space Shuttle Program (1D-1)
1996	Europe (ESA)	ECSS	European Corporation for Space Standardization
		ECSS-E	Engineering (-00, -10)
		ECSS-M	Project Management (-00, -10, -20, -30, -40, -50, -60,-70)
		ECSS-Q	Product Assurance (-00, -20, -30, -40, -60, -70, -80)
1998	Europe	pr EN 9000-1	Aerospace Industry Quality System (Part 1: Req. for Suppliers)

A2.3 Requirements in the Aerospace, Defense, and Nuclear Fields

Requirements in the *space field* often combine the aspects of quality, reliability, maintainability, safety, and software quality in a *Product Assurance* document, well conceived in its structure and content [A2.4, A2.10]. In the *avionic field*, a standard, prEN 9000-1:1998 [A2.3], has been issued by reinforcing locally the requirements of ISO 9001 and ISO 9002. It can be expected that the space and avionic fields will unify standards in an *Aerospace Series* [A2.3].

The MIL-Standards are applicable in the *defense field*. Despite their national character (USA), they have also been often used during the last 30 years in international projects, in particular MIL-Q-9858 and MIL-STD-470, -471, -781 & -785 for equipment and systems [A2.9]. MIL-Q-9858 (first ed. 1959) was the basis for almost all quality assurance standards. However, as it does not cover specific aspects of reliability, maintainability, and safety, MIL-STD-785, -470, and -882 were issued (1965–1970). MIL-STD-785 (ed.B,1980) requires the realization of a *reliability program*. Such a program must be *tailored* according to the complexity of the equipment or system considered, tasks are described in detail with regard to planning, analysis, selection/qualification of components and materials, design reviews, FMEA/FMECA, prototype qualification tests, collection and analysis of failures, corrective actions, reliability growth, and acceptance tests. *MTBF* $(=1/\lambda)$ *acceptance test procedures* are given by MIL-STD-781 and MIL-HDBK-781. MIL-STD-470 requires the realization of a *maintainability program*, with particular emphasis on design rules, design reviews, FMEA/FMECA, and acceptance tests (the program is to be developed in accordance with a *maintenance concept* which also includes *logistical support*). *Maintainability demonstration* is covered by MIL-STD-471 (new ed. of MIL-HDBK-470 and -471 is announced). MIL-STD-882 requires the realization of a *safety program*, in particular the *analysis* of all *potential hazards*, a review of the configuration with respect to safety, and safety tests. For *NATO* countries, *AQAP Requirements* were issued starting 1968. These requirements deal essentially with quality assurance (AQAP-1 corresponds to MIL-Q-9858, differing only in structure). The above MIL-STDs apply to the development, production, and operation of equipment or systems and can be useful in developing standards/procedures for industrial applications.

The *nuclear field* has its own specific and well established standards, with emphasis on safety, design reviews, configuration accounting, components/materials and production processes qualification, quality control, and safety tests.

A3 Definition and Realization of Quality and Reliability Requirements

In defining *quality and reliability requirements*, market needs, life cycle cost aspects, time to market, as well as development and production risks (for instance when using new technologies) have to be considered with care. For complex equipment and systems the *realization* of quality and reliability requirements is best achieved with a *quality and reliability assurance program*. Such a *program* defines the *project-specific* activities for quality and reliability assurance and *assigns responsibilities* for their realization in agreement with *TQM* aims. To be effective, a quality and reliability assurance program must be *integrated* in the *project activities*. This appendix deals with important aspects in defining quality and reliability requirements, and considers the structure and the content of a quality and reliability assurance program for *complex equipment and systems with high quality and reliability requirements*. For less stringent requirements, *tailoring* is necessary to meet real needs and to be cost effective.

A3.1 Definition of Quality and Reliability Requirements

In defining *quantitative*, project specific *quality and reliability requirements* attention has to be paid to the possibilities for their *concrete realization* during design and production as well as for their *demonstration* at an acceptance test. These requirements are derived from customer or market *needs*, taking into account technical, cost, and ecological limitations *(sustainable development)*. This section deals with some important aspects for setting *MTBF*, *MTTR*, and steady-state availability $PA = MTBF/(MTBF + MTTR)$ requirements.

Tentative targets for *MTBF*, *MTTR* and *PA* are set by considering

- operational requirements relating to reliability, maintainability, and availability,
- allowed logistical support,
- required function and expected environmental conditions,

- experience with similar equipment or systems,
- possibility for redundancy at higher integration level,
- requirements for life-cycle cost, power consumption, etc.,
- ecological consequences (sustainability).

Typical *figures for the failure rates* $\lambda = 1/MTBF$ of complex electronic assemblies or equipment are between 300 and $5{,}000 \cdot 10^{-9}\,\mathrm{h}^{-1}$ at ambient temperature θ_A of 40°C and with a duty cycle d of 0.3, see Table A3.1 for some examples. The *duty cycle* $(0 < d \le 1)$ gives the mean of the ratio between operational time and calendar time for the item considered. Assuming a *constant* failure rate λ and *no reliability degradation* caused by power on/off, an *equivalent failure rate*

$$\lambda_d = d\lambda \tag{A3.1}$$

can be used for practical purposes. Often it can be useful to operate with the *mean expected number of failures per year* and 100 items

$$m = \lambda_d \cdot 8{,}600\ \mathrm{h} \cdot 100\% \approx \lambda_d \cdot 10^6\,\mathrm{h}. \tag{A3.2}$$

$m < 1\%$ is a good target for complex equipment and can influence aquisition cost.

Tentative targets are refined successively by performing rough analysis and comparative studies (definition of goals down to assembly level can be necessary at this time). For *acceptance testing* (demonstration) of an *MTBF*, the following data are important:

1. Quantitative requirements: $MTBF_0$ = specified *MTBF* and/or $MTBF_1$ = minimum acceptable *MTBF* (Sections 7.2.2.2 and 7.2.2.3).
2. Required function (mission profile).
3. Environmental conditions (thermal, mechanical, climatic).

Table A3.1 Indicative values of failure rates λ and mean expected number m of failures per year and 100 items for a duty cycle $d = 30\%$ and $d = 100\%$ $(\theta_A = 40°\mathrm{C})$

	$d = 30\%$		$d = 100\%$	
	$\lambda\ [10^{-9}\,\mathrm{h}^{-1}]$	m [%]	$\lambda\ [10^{-9}\,\mathrm{h}^{-1}]$	m [%]
Telephone switchboard	2,000	2	6,000	6
Telephone receiver (multi-function)	200	0.2	600	0.6
Photocopier incl. mechanical parts	200,000	200	600,000	600
Personal computer	3,000	3	9,000	9
Radar equipment (ground, mobile)	300,000	300	900,000	900
Control card for autom. process control	1,000	1	3,000	3
Mainframe computer system	—	—	100,000	100

4. Allowed producer's and/or consumer's risks (α and/or β).
5. Acceptance conditions: cumulative operating time T and number c of allowed failures during T (Sections 7.2.2.2 and 7.2.2.3).
6. Number of systems under test ($T / MTBF_0$ as a rule of thumb).
7. Parameters which should be tested and frequency of measurement.
8. Failures which should be ignored for the *MTBF* acceptance test.
9. Maintenance and screening before the acceptance test.
10. Maintenance procedures during the acceptance test.
11. Form and content of test protocols and reports.
12. Actions in the case of a negative test result.

For acceptance testing (demonstration) of an *MTTR*, the following data are important (Section 7.3.2):

1. Quantitative requirements (*MTTR*, variance, quantile).
2. Test conditions (environment, personnel, tools, external support, spare parts).
3. Number and extent of repairs to be undertaken (simulated failures).
4. Allocation of the repair time (diagnostic, repair, functional test, logistical time).
5. Acceptance conditions (Section 7.3.2).
6. Form and content of test protocols and reports.
7. Actions in the case of a negative test result.

Availability usually follows indirectly from the relationship $PA = MTBF / (MTBF + MTTR)$. However, for an acceptance test, procedures for an unknown probability $p = 1 - PA$ can also be used (Sections 7.1.2, 7.1.3, and 7.2.1).

A3.2 Realization of Quality and Reliabilty Requirements

For complex items, in particular at the equipment and system level, *quality and reliability targets* can best be achieved by setting up and realizing a *quality and reliability assurance program*. In such a program, project specific tasks and activities are clearly described and assigned. Table A3.2 can be used as a *checklist to define the content* of a quality and reliability assurance program for *complex* equipment and systems. It is a refinement of Table 1.2 and shows a possible *task assignment* in a company organized as in Fig. 1.7, conforming to TQM. Depending on the item technology and complexity, Table A3.2 is to be shortened or extended. The task assignments, *R* for primary responsibility, *C* for secondary responsibility (cooperates actively), and *I* for information, can be modified to reflect the company's personnel situation.

Table A3.2 Tasks for quality and reliability assurance of *complex equipment and systems with high quality and reliability requirements,* as well as possible task assignment conforming to TQM (Fig. 1.7)

Example of task assignment for the quality and reliability assurance of *complex* equipment and systems *with high quality and reliability requirements,* conforming to TQM (checklist for the preparation of a quality and reliability assurance program) R stands for responsibility, C for cooperation, and I for information	Marketing	Development	Production	Q&R Assurance
1 Customer and market requirements				
1 Evaluation of delivered equipment and systems	R	I	I	C
2 Determination of market and customer demands and real needs	R	I	I	C
3 Customer support	R			C
2 Preliminary analyses				
1 Definition of tentative quantitative targets for reliability, maintainability, availability, safety, and quality level	C	C	C	R
2 Rough analyses and identification of potential problems	I	C		R
3 Comparative investigations	I	C		R
3 Quality and reliability aspects in specifications, quotations, contracts, etc.				
1 Definition of the required function	I	R		C
2 Determination of external environmental stresses	C	R		C
3 Definition of realistic quantitative targets for reliability, maintainability, availability, safety, and quality level	C	C	C	R
4 Specification of test and acceptance criteria	C	C	C	R
5 Identification of the possibility to obtain field data	R			C
6 Cost estimate for quality and reliability assurance activities	C	C	C	R
4 Quality and reliability assurance program				
1 Preparation	C	C	C	R
2 Realization				
– design and evaluation	I	R	I	C
– production	I	I	R	C
5 Reliability and maintainability analyses				
1 Specification of the required function for each element		R		C
2 Determination of environmental, functional, and time-dependent stresses (detailed operating conditions)		R		C
3 Assessment of derating factors		C		R
4 Reliability and maintainability allocation		C		R
5 Preparation of reliability block diagrams				
– assembly level		R		C
– system level		C		R
6 Identification and analysis of reliability weaknesses (FMEA/FMECA, FTA, worst-case, drift, stress-strength-analyses)				
– assembly level		R		C
– system level		C		R

Table A3.2 (cont.)

7 Carrying out comparative studies				
– assembly level		R		C
– system level		C		R
8 Reliability improvement through redundancy				
– assembly level		R		C
– system level		C		R
9 Identification of components with limited life time	I	R		C
10 Elaboration of the maintenance concept	I	R	I	C
11 Elaboration of a test and screening strategy	C	C	C	R
12 Analysis of maintainability		R		C
13 Elaboration of mathematical models		C		R
14 Computation of the predicted reliability and maintainability				
– assembly level	I	R		C
– system level	I	C		R
15 Reliability and availability computation at system level	I	I		R
6. Safety and human factor analyses				
1 Analysis of safety (avoidance of liability problems)				
– accident prevention	C	R	C	C
– technical safety				
• identification and analysis of critical failures and of risk situations (FMEA/FMECA, FTA, etc.)				
– assembly level		R		C
– system level	I	C		R
• theoretical investigations		C		R
2 Analysis of human factors (man-machine interface)	C	R	C	C
7. Selection and qualification of components and materials				
1 Updating of the list of preferred components and materials	I	C	I	R
2 Selection of non-preferred components and materials		R	C	C
3 Qualification of non-preferred components and materials				
– planning		C	I	R
– realization			C	R
– analysis of test results		I	I	R
4 Screening of components and materials		I	C	R
8. Supplier selection and qualification				
1 Supplier selection				
– purchased components and materials		R	C	C
– external production		C	R	C
2 Supplier qualification (quality and reliability)				
– purchased components and materials		I	I	R
– external production		I	I	R
3 Incoming inspections				
– planning		C	C	R
– realization			C	R
– analysis of test results			C	R
– decision on corrective actions				
• purchased components and materials		C	C	R
• external production		R	C	C

Table A3.2 (cont.)

9. *Project-dependent procedures and work instructions*				
1 Reliability guidelines		C		R
2 Maintainability guidelines	C	C	I	R
3 Safety guidelines	I	C	I	R
4 Other procedures, rules, and work instructions				
• for development		R	I	C
• for production		I	R	C
5 Compliance monitoring	C	C	C	R
10. *Configuration management*				
1 Planning and monitoring	C	C	C	R
2 Realization				
– configuration identification				
• during design		R		C
• during production		I	R	C
• during use (warranty period)	R	I	I	C
– configuration auditing (design reviews)	C	R	C	C
– configuration control (evaluation, coordination, and release or rejection of changes and modifications)				
• during design	C	R	C	C
• during production	C	C	R	C
• during use (warranty period)	R	C	C	C
– configuration accounting		R	C	C
11. *Prototype qualification tests*				
1 Planning	I	R	I	C
2 Realization	C	R	C	C
3 Analysis of test results	I	R	I	C
4 Special tests for reliability, maintainability, and safety	I	C	C	R
12. *Quality control during production*				
1 Selection and qualification of processes and procedures		R	C	C
2 Production planning		C	R	C
3 Monitoring of production processes		I	R	C
13. *In-process tests*				
1 Planning		C	R	C
2 Realization		I	R	I
14. *Final and acceptance tests*				
1 Environmental tests and/or screening of series-produced items				
– planning	I	C	C	R
– realization	I	I	C	R
– analysis of test results	I	C	C	R
2 Final and acceptance tests				
– planning	C	C	C	R
– realization	I	I	C	R
– analysis of test results	C	C	C	R
3 Procurement, maintenance, and calibration of test equipment	I	C	C	R

Table A3.2 (cont.)

15. *Quality data reporting system*				
1 Data collection	C	C	C	R
2 Decision on corrective actions				
– during prototype qualification		R	I	C
– during in-process tests		C	R	C
– during final and acceptance tests	C	C	C	R
– during use (warranty period)	R	C	C	C
3 Realization of corrective actions on hardware or software (repair, rework, waiver, scrap)	I	C	C	R
4 Implementation of the changes in the documentation (technical, production, customer)	C	C	C	R
5 Data compression, processing, storage, and feedback	I	I	I	R
6 Monitoring of the quality data reporting system	I	I	I	R
16. *Logistical support*				
1 Supply of special tools and test equipment for maintenance	C	R	I	C
2 Preparation of customer documentation	R	C	I	I
3 Training of operating and maintenance personnel	R	I	I	I
4 Determination of the required number of spare parts, maintenance personnel, etc.	R	C		C
5 After-sales support	R	I	I	C
17. *Coordination and monitoring*				
1 Project-specific	C	C	C	R
2 Project-independent	I	I	I	R
3 Planning and realization of quality audits				
– project-specific	C	C	C	R
– project-independent	I	I	I	R
4 Information feedback	I	I	I	R
18. *Quality costs*				
1 Collection of quality cost	C	C	C	R
2 Cost analysis and initiation of appropriate actions	C	C	C	R
3 Preparation of periodic and special reports	C	C	C	R
4 Evaluation of the effectiveness of quality and rel. assurance	I	I	I	R
19. *Concepts, methods, and general procedures (quality and reliability)*				
1 Development of concepts	C	C	C	R
2 Investigation of methods	I	I	I	R
3 Preparation and updating of the quality and reliability assurance handbook	C	C	C	R
4 Development of software packages	I	I	I	R
5 Collection, evaluation, and distribution of data, experience and know-how	I	I	I	R
20. *Motivation and training*				
1 Planning	C	C	C	R
2 Preparation of courses and documentation	C	C	C	R
3 Realization of the motivation and training program	C	C	C	R

A3.3 Elements of a Quality and Reliability Assurance Program

The basic elements of a quality and reliability assurance program are:

1. Project organization, planning, and scheduling
2. Quality and reliability requirements
3. Reliability and safety analysis
4. Selection and qualification of components, materials, and processes
5. Configuration management
6. Quality tests
7. Quality data reporting system

These elements are discussed in this section for the case of a *quality and reliability assurance program* for *complex* equipment and systems. In addition, Appendix A4 gives a catalog of questions to generate *checklists for design reviews* and Appendix A5 specifies the requirements for a *quality data reporting system*. As suggested in task 4 of Table A3.2, the *preparation* of a quality and reliability assurance program should be the responsibility of the *quality and reliability manager*. However, its *realization* falls within the competence of the *project manager*.

It is often useful to start with a quality and reliability programm for the development phase, covering items 1 to 5 of the above list, and continue then with the production phase for points 5 to 7. Both programs must be correlated, especially if *concurrent engineering* is applied.

A3.3.1 Project Organization, Planning, and Scheduling

A clearly defined project organization and planning is necessary for the realization of a quality and reliability assurance program. Organization and planning must also satisfy modern needs of *concurrent engineering*.

The *system specification* is the basic document for all considerations at project level. The following is a typical outline for system specifications:

1. State of the art, need for a new product
2. Target to be achieved
3. Costs, time schedule
4. Market potential (turnover, price, competition)
5. Technical performance
6. Environmental conditions
7. Operational capabilities (reliability, maintainability, availability, logistical support)
8. Quality and reliability assurance

9. Special aspects (new technologies, patents, value engineering, etc.)
10. Appendices

The organization of a project begins with the definition of the main task groups. The following groups are usual for a complex system: Project Management, System Engineering, Life-Cycle Cost, Quality and Reliability Assurance, Assembly Design, Prototype Qualification Tests, Production, Assembly, and Final Testing. Project organization, task lists, task assignment, and milestones can be derived from the task groups, allowing the quantification of the personnel, material, and financial resources needed for the project. The quality and reliability assurance program must require that the project is clearly and suitably organized and planned.

A3.3.2 Quality and Reliability Requirements

The most important steps in defining quality and reliability targets for complex equipment and systems have been discussed in Appendix A.3.1.

A3.3.3 Reliability and Safety Analyses

Reliability and safety analyses include *failure rate* analysis, *failure mode* analysis (FMEA/FMECA, FTA), *sneak circuit analysis* (to identify latent paths which cause unwanted functions or inhibit desired functions, while all components are functioning properly), evaluation of *concrete* possibilities to improve reliability and safety (derating, screening, redundancy), as well as *comparative studies*; see Chapters 2 to 6 for methods and tools.

The quality and reliability assurance program must show what is *actually being done* for the project considered, in particular, it should be able to supply answers to the following questions:

1. Which derating rules are considered?
2. Which failure rate data will be used? How were the factors determined?
3. How were the component-level operating conditions determined?
4. Which tool is used for failure mode analysis? To which elements does it apply?
5. Which kind of comparative studies will be performed?
6. Which design guidelines for reliability, maintainability safety and software quality are used?
7. How will it be verified that the design guidelines have been followed?

Additionally, interfaces to the selection and qualification of components and materials, design reviews, test and screening strategies, reliability tests, quality data reporting system, and subcontractor activities must be shown. The data used for component failure rate calculation should be critically evaluated (source, current relevance, assumed environmental and quality factors).

A3.3.4 Selection and Qualification of Components, Materials, and Manufacturing Processes

Components (parts), materials, and production processes have a great impact on the product quality and reliability, they must be therefore carefully selected and qualified. Qualification tests are given in Chapter 3 for electronic components and for assemblies. For production processes one may refer to the literature, e.g. [8.1 to 8.14].

The quality and reliability assurance program should provide a clear indication as to how components, materials, and processes are selected and qualified. In particular the following questions should be answered:

1. Does a *list of preferred components and materials* exist? How was it established? Do the assumptions of the list apply to the project under consideration (operating conditions for example)? Have second sources been considered? Will critical components be available on the market-place at least for the required production and warranty time?
2. How are new components selected? What is the qualification procedure?
3. Under what conditions can a designer use non qualified components/materials?
4. How will obsolescence problems be solved?
5. How have the standard manufacturing processes been qualified?
6. How are *special manufacturing* processes qualified?

Special manufacturing processes are those which quality cannot be tested directly on the product, have high requirements with respect to reproducibility, or may have negative effect on the product quality or reliability.

A3.3.5 Configuration Management

Configuration management is an important tool for quality assurance and TQM during design. Within a project, it is subdivided into configuration identification, auditing, control, and accounting.

The *identification* of an item is recorded in its documentation. A possible documentation outline for *complex* equipment and systems is given in Fig. A3.1.

Configuration auditing is done via *design reviews*, the aim of which is to assure that the system will meet all requirements. In a design review, all aspects of design (selection and use of components and materials, dimensioning, interfaces, construction problems), production (manufacturability, testability, reproducibility), reliability, maintainability, safety, patent regulations, value engineering, and value analysis are critically examined with the help of *checklists* (Appendix A4). The most important design reviews are described in Table A3.3. For complex systems a review of the first production unit (FCA/PCA) also usually takes place. Design reviews are chaired by the *project manager* and co-chaired by the project quality and reliability assurance manager. For complex equipment or systems, the review team may vary according to the following list:

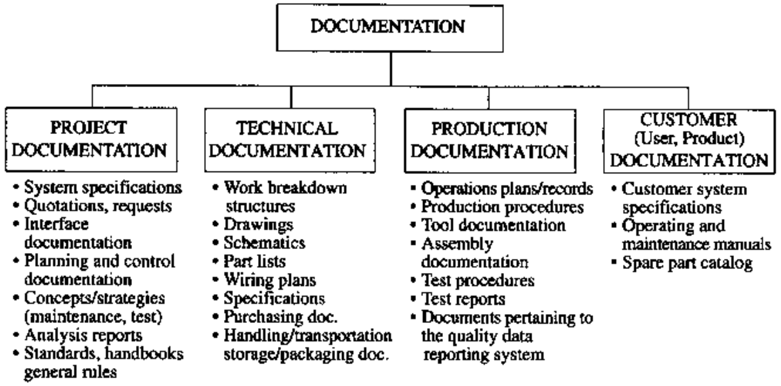


Fig. A3.1 Documentation outline for complex equipment and systems

- project manager,
- project engineer for quality and reliability assurance,
- design engineers,
- representatives from production and marketing,
- independent design engineer or external consultant,
- customer representatives (if appropriate).

Some weeks before the design review, participants should present project-specific checklists (Appendix A4, Table 2.7, Table 4.3).

Configuration control includes evaluation, coordination, and release or rejection of all proposed changes and modifications. *Changes* occur as a result of defects or failures, *modifications* are triggered by a revision of the system specifications.

Configuration accounting ensures that all approved changes and modifications have been implemented and recorded. This calls for a defined procedure, as changes/modifications must be realized in hardware, software, and documentation.

A one-to-one correspondence between hardware or software and documentation is important during all life-cycle phases of a product. Complete records over all life-cycle phases become necessary if *traceability* is required, as in the aerospace or nuclear field for example. Partial traceability can also be required for products which are critical with respect to *safety*, or because of *product liability*.

Table A3.3 Design reviews during definition, design, and dev. of complex equipment and systems

	System Design Review (SDR)	Preliminary Design Reviews (PDR)	Critical Design Review (CDR)
To be performed	At the end of the definition phase	During the design phase, each time an assembly has been developed	At the end of prototype qualification tests
Goals	• Critical review of the system specifications on the basis of results from market research, rough analysis, comparative studies, patent situation, etc.	• Critical review of all documents belonging to the assembly under consideration (computations, schematics, parts lists, test specifications, etc.) • Comparison of the target achieved with the system specifications requirements • Checking interfaces to other assemblies	• Critical comparison of prototype qualification test results with system requirements • Formal review of the correspondence between technical documentation and prototype • Verification of producibility, testability, and reproducibility
Input	• Item list • System specifications (draft) • Documentation (analyses, reports, etc.) • Checklists (one for each participant)*	• Item list • Documentation (analyses, schematics, drawings, parts lists, test specifications, work breakdown structure, interface specifications, etc.) • Reports of earlier design reviews • Checklists (one for each participant)*	• Item list • Technical documentation • Testing plan and procedures for prototype qualification tests • Results of prototype qualification tests • List of deviations from the system requirements • Maintenance concept • Checklists (one for each participant)*
Output	• System specifications • Proposal for the design phase • Interface definitions • Rough maintenance and logistical support concept • Report	• Reference configuration (baseline) of the assembly considered • List of deviations from the system specifications • Report	• List of the final deviations from the system specs. • Qualified and released prototypes • Frozen technical documentation • Revised maintenance concept • Production proposal • Report

* see Appendix A4 for a catalog of questions to generate project-specific checklists

Referring to configuration management, the quality and reliability assurance program should answer the following questions:

1. Which documents will be produced by whom, when, and with what content?
2. Are document contents in accordance with quality and reliability requirements?
3. Is the release procedure for all documents compatible with quality requirements?
4. Are the procedures for changes clearly defined?
5. How is compatibility (upward and/or downward) assured?
6. How is configuration accounting assured during production?
7. Which items are subject to traceability requirements?

A3.3.6 Quality Tests

Quality tests are necessary to verify whether an item conforms to specified requirements. Such tests cover performance, reliability, maintainability, and safety aspects, and include incoming inspections, qualification tests, production tests, and acceptance tests. To optimize costs and time schedules, tests should be integrated in a *test* (and screening) *strategy* at system level. Methods for statistical quality control and reliability tests are presented in Chapter 7. Qualification tests and screening procedures are discussed in Sections 3.2 and 8.2 to 8.4. Basic considerations for test and screening strategies are given in Section 8.5.

The quality and reliability assurance program should answer the following questions:

1. What are the test and screening strategies at system level? How are they released?
2. How were subcontractors selected, qualified and monitored?
3. What is specified in the procurement documentation? How is it released?
4. How is the incoming inspection performed? With which test procedures? How are these documents released?
5. Which components and materials are 100% tested? Which are 100% screened? What are the procedures for screening?
6. How are prototypes qualified? Who has released the relevant procedures? Who decides on test results?
7. How are production tests performed? Who has released the relevant procedures? Who decides on test results?
8. How is environmental stress screening determined for series-produced items? Who has released the relevant procedures? Who decides on the test results?
9. Which procedures are applied to defective or failed items? Who has released them?
10. What are the instructions for handling, transportation, storage, and shipping? Who has released the relevant procedures?

A3.3.7 Quality Data Reporting System

Starting at the prototype qualification tests, all defects and failures should be systematically collected, analyzed and corrected. Analysis should go back to the *cause of the fault*, in order to find those actions most appropriate for avoiding *repetition of the same problem*. The concept of a *quality data reporting system* is illustrated in Fig. 1.8, detailed requirements are given in Appendix A5.

The quality and reliability assurance program should answer the following questions:

1. How is the collection of defect and failure data carried out? At which project phase is the quality data reporting system introduced?
2. How are defects and failures analyzed? What are the procedures?
3. Who carries out corrective actions? Who monitors their realization? Who checks the final configuration?
4. How is evaluation and feedback of quality and reliability data organized?
5. Who is responsible for the quality data reporting system? Does production have their own limited version of such a system? How do these systems interface?

A4 Checklists for Design Reviews

The following catalog of questions can be used to generate *project-specific checklists for design reviews* (Tables A3.2 (Point 10), A3.3, and 5.5) during the design and development of *complex* equipment and systems.

A4.1 System Design Review

1. What experience exists with similar equipment and systems?
2. What are the goals for performance (capability), reliability, maintainability, availability and safety? How have they been defined? Which mission profile (required function and environmental conditions) is applicable?
3. What tentative allocation of the performance parameters, reliability, and maintainability down to the assembly level was undertaken?
4. What are the critical elements? Are potential problems to be expected (new technologies, interfaces)?
5. Have comparative studies been done? What are their results?
6. Are interference problems (external or internal EMC) to be expected?
7. Are there potential safety problems?
8. Are the requirements realistic? Do they correspond to a market need?
9. Is there a maintenance concept? Do special ergonomic requirements exist?
10. Are there special software requirements?
11. Has the patent situation been verified? Are licenses necessary?
12. Are there estimates of life-cycle cost? Have these been optimized with respect to reliability and maintainability requirements?
13. Is there a feasibility study? Where does the competition stand? Has development risk been assessed?
14. Is the project time schedule realistic? Can the system be marketed at the right time?
15. Can supply problems be expected during production ramp-up?

A4.2 Preliminary Design Reviews (Assembly Level)

a) General

1. Is the assembly under consideration a new development or only a change/modification?
2. Can existing elements be used?
3. Will the assembly be used in other systems?
4. Is there experience with similar assemblies? What were the problems?
5. Is there a block diagram or functional schematic?
6. Have specification been fulfilled? How were these checked? Where do deviations exist? Can individual requirements be reduced?
7. Are there patent problems? Do licenses have to be purchased?
8. Have expected costs and deadlines been met?
9. Were value engineering methods used?
10. Have customer and market demands changed since the beginning of development?

b) Performance Parameters

1. How have been defined the main performance parameters of the assembly under consideration?
2. How was the fulfillment of performance parameters verified (assumptions, computations, simulation, tests)?
3. Have worst case situations been considered in computations?
4. Have interference problems (EMC) been solved?
5. Have currently applicable standards been observed during dimensioning?
6. Have interface problems with other assemblies been solved?
7. Have prototypes been adequately tested in the laboratory?

c) Environmental Conditions

1. Have environmental conditions been defined? Also as a function of time?
2. Were these combined with the effects of internal load to determine component operating conditions?
3. Has external interference (EMC) been determined (experimentally or by computation) or just been assumed? Has this influence been taken into account in worst case calculations?

d) Components and Materials

1. Which components (parts) and materials do not appear in the preferred lists? For what reasons? How were these parts and materials qualified?
2. Are incoming inspections necessary? For which components and materials? How/Who will they be performed?
3. Which components and materials were screened? How/Who will screening be performed?
4. Do some components require special screening? Why? Can these components be delivered on time?
5. Are suppliers guaranteed for series production? Is there at least one second source for each component and material? Have requirements for quality, reliability, and safety been met?
6. Are obsolescence problems to be expected? How will they be solved?

e) Reliability

See Table 2.8.

f) Maintainability

See Table 4.3.

g) Safety

1. Have design rules concerning accident prevention been observed?
2. Has safety been considered with regard to external causes (natural catastrophe, sabotage)?
3. Has a FMEA/FMECA been performed? Have all single-point failures been identified? Can these be avoided?
4. Has a fail-safe analysis been performed? What were the results?
5. What safety tests are planned? Are they sufficient?
6. Have safety aspects been dealt with adequately in the documentation?

h) Human Factors, Ergonomy

1. Have operating and maintenance sequences been defined with regard to the training level of operators and maintenance personnel?
2. Have ergonomic factors been taken into account by defining operating sequences?
3. Has the man-machine interface been considered?

i) Standardization

1. Have standard components (parts) and materials been used wherever possible?
2. Has part exchangeability been considered during design and construction?
3. Are symbols used in accordance with appropriate standards?

j) Configuration

1. Is the technical documentation (schematics, drawings, etc.) complete, error-free, and does it reflect the current state of the project?
2. Have all interface problems with other assemblies been solved?
3. Can the technical documentation be frozen and considered as reference documentation (baseline)?
4. Can the same be done with the configuration as a whole?
5. How is compatibility (upward and/or downward) assured?

k) Production and Testing

1. Which qualification tests are foreseen for the prototype? Have reliability, maintainability, and safety aspects been considered sufficiently in these tests? Can these tests be combined with qualification testing of other assemblies (or of the whole system)?
2. Have all questions been answered regarding producibility, testability, and reproducibility?
3. Are special production processes necessary? Were they qualified? What were the results?
4. Are special transport, packaging, or storage problems to be expected?

A4.3 Critical Design Review (System Level)

a) Technical Aspects

1. Does the technical documentation allow an exhaustive and correct interpretation of test procedures and results? Has the technical documentation been frozen? Has conformance with hardware and software been checked?
2. Are test specifications and procedures complete? In particular are conditions for functional, environmental, reliability, and safety tests clearly defined?

3. Have fault criteria been defined for critical parameters? Is an indirect measurement planned for those parameters which cannot be measured accurately enough during tests?
4. Has a representative mission profile, with a corresponding required function, been sufficiently clearly defined for reliability tests?
5. Have test criteria for maintainability been defined? Which failures were simulated? How have personnel and material conditions been fixed?
6. Have test criteria for safety been defined (accident prevention and technical safety)?
7. Have ergonomic aspects been checked? How?
8. Have packaging, transport and storage caused any problem?
9. Have defects and failures been systematically analyzed (mode, cause, effect)? Has the usefulness of corrective actions been verified? How? Also with respect to cost?
10. Have all deviations been recorded? Can they be accepted?
11. Does the system still satisfy customer/market needs?
12. Are manufacturability and reproducibility guaranteed within the framework of a production environment? Can pilot production items be released?

b) Formal Aspects

1. Is the technical documentation complete?
2. Has the technical documentation been checked for correctness? How?
3. Have all parts of the technical documentation been checked for coherency? How?
4. Is uniqueness in numbering guaranteed? Even in the case of changes?
5. Is hardware labeling appropriate? Does it satisfy production and maintenance requirements?
6. Has conformance between prototype and documentation been checked?
7. Is the maintenance concept mature? Are spare parts having a different change status fully interchangeable?
8. Are production tests sufficient from today's point of view? Can testability or manufacturability problems be expected?

A5 Requirements for Quality Data Reporting Systems

A *quality data reporting system* is a system to collect, analyze, and correct all *defects* and *failures* occurring during production and testing of an item, as well as to evaluate and feedback the corresponding quality and reliability data (Fig. 1.8). The system is generally computer aided. Analysis of failures and defects must go back to the *root cause* in order to determine the *most appropriate action* necessary to avoid *repetition of the same problem*. The quality data reporting system should also remain active during the *operating phase*. This appendix summarizes the requirements for a computer aided quality data reporting system for complex equipment and systems.

a) General Requirements

1. Up-to-dateness, completeness, and utility of the delivered information must be the primary concern (best compromise).
2. A high level of *usability* (user friendliness) and minimal manual intervention should be a goal.
3. Procedures and responsibilities should be clearly defined (several levels depending upon the consequence of defects or failures).
4. The system should be flexible and easily adaptable to new needs.

b) Requirements Relevant to Data Collection

1. All data concerning *defects* and *failures* (relevant to quality, reliability, maintainability, and safety assurance) have to be collected, from the begin of prototype qualification tests to (at least) the end of the warranty period.
2. Data collection forms should
    - be 8" × 11" or A4 format
    - be project-independent and easy to fill in
    - ensure that only the relevant information is entered and answers the questions: what, where, when, why, and how?
    - have a separate field (20-30%) for free-format input for comments (requests for analysis, logistical information, etc.), these comments do not need to be processed and should be easily separable from the fixed portion of the form.

3. Description of the *symptom* (mode), *analysis* (cause, effect), and *corrective action* undertaken should be recorded in clear text and coded at data entry by trained personnel.
4. Data collection can be carried out in different ways
    - at a single reporting location (adequate for simple problems which can be solved directly at the reporting location)
    - from different reporting locations which report the *fault* (defect or failure), *analysis result*, and *corrective action* separately.

    Current reliability, maintainability, or logistical data can also be reported.
5. Data collection forms should be entered into the computer *daily* or even on line, so that corrective actions can be initiated as quickly as possible (for field data, a weekly or monthly entry can be sufficient for many purposes).

c) Requirements for Analysis

1. The *cause* should be found for each defect or failure
    - at the reporting location, in the case of simple problems
    - by a fault review board, in critical cases.
2. Failures (and defects) should be classified according to
    - mode
    - – sudden failure (short, open, fracture, etc.)
    - – gradual failure (drift, wearout, etc.)
    - – intermittent failures
    - – others if needed
    - cause
    - – intrinsic (inherent weaknesses, wearout, or some other intrinsic cause)
    - – extrinsic (systematic failure, i.e. misuse, mishandling, design, or manuf. failure)
    - – secondary failure
    - effect
    - – irrelevant
    - – partial failure
    - – complete failure
    - – critical failure (safety problem).
3. Analysis results (*repair, rework, change*, or *scraping*) must be reported.

d) Requirements for Corrective Actions

1. Every record is considered *pending* until the necessary corrective action has been successfully completed and certified.
2. The quality data reporting system must monitor *all* corrective actions.
3. Procedures and responsibilities pertaining to *corrective action* are to be defined (simple cases are usually solved by the reporting location).
4. The reporting location must be informed about a *completed* corrective action.

e) Requirements Related to Data Processing, Feedback , and Storage

1. Adequate coding must allow *data compression* and simplify data processing.
2. Up-to-date information should be available *on-line.*
3. Problem-dependent and periodic *evaluations* must be possible.
4. At the end of a project, relevant information should be stored for comparison.

f) Requirements Related to Compatibility with other Software Packages

1. Compatibility with firm-specific configuration management and data banks should be assured.
2. Data transfer with the following external software packages should be assured
   - important reliability data banks
   - quality data reporting systems of subsidiary companies
   - quality data reporting systems of large contractors.

The effort required for implementing a quality data reporting system as described above can take 5 to 10 man-years for a medium-sized company. Competence for operation and maintenance of the quality data reporting system should be with the company's quality and reliability assurance department. The priority for the realization of *corrective actions* is project-specific and should be fixed by the project manager. Major problems (defects and failures) should be discussed periodically by a *fault review board* chaired by the company's quality and reliability assurance manager, which can take go/nogo decisions.

A6 Basic Probability Theory

In many practical situations, experiments have a *random outcome*, i.e., the results cannot be predicted exactly, although the *same experiment* is repeated under *identical conditions.* Examples are inspection of a given item during production, failure-free operating time of a given system, repair time of an equipment, etc. Experience shows that as the number of repetitions of the same experiment increases, certain *regularities* appear regarding the occurrence of the event considered. *Probability theory* is a mathematical discipline which investigates the laws describing such regularities. The assumption of *unlimited repeatability* of the same experiment is basic to probability theory. This assumption permits the introducion of the concept of *probability* for an event starting from the properties of the *relative frequency* of its occurrence in a long series of trials. The *axiomatic theory of probability*, introduced 1933 by A.N. Kolmogorov [A6.10], brought probability theory to a *mathematical discipline*. In *reliability analysis*, probability theory allows the investigation of the probability that a given item will operate failure-free for a stated period of time under given conditions, i.e. the calculation of the item's *reliability* on the basis of a *mathematical model.* The rules necessary for such calculations are presented in Sections A6.1 to A6.4. The following sections are devoted to the concept of *random variables*, necessary to investigate reliability as a function of time and as a basis for *stochastic processes* (Appendix A7) and *mathematical statistics* (Appendix A8). This appendix is a compendium of probability theory, consistent from a mathematical point of view but still with engineering applications in mind.

A6.1 Field of Events

As introduced by A.N. Kolmogorov [A6.10], the mathematical model of an experiment with random outcome is a triplet $[\Omega, \mathcal{F}, \Pr]$. Ω is the *sample space*, $\mathcal{F}$ the *event field*, and Pr the *probability* of each element of $\mathcal{F}$. Ω is a set containing as elements all possible outcomes of the experiment considered. Hence $\Omega = \{1, 2, 3, 4, 5, 6\}$ if the experiment consists of a single throw of a die, and $\Omega = [0, \infty)$ in the case of failure-free operating times of an item. The elements of Ω

are called *elementary events* and are represented by ω. If the logical statement "the outcome of the experiment is a subset A of Ω" is identified with the subset A itself, combinations of statements become equivalent to operations with subsets of Ω. If the sample space Ω is finite or countable, a probability can be assigned to every subset of Ω. In this case, the event field $\mathcal{F}$ contains *all* subsets of Ω and all combinations of them. If Ω is continuous, restrictions are necessary. The *event field* $\mathcal{F}$ is thus a system of subsets of Ω to each of which a probability has been assigned according to the situation considered. Such a field is called a *Borel field* (σ-field) and has the following properties:

1. Ω is an element of $\mathcal{F}$.
2. If A is an element of $\mathcal{F}$, its complement $\bar{A}$ is also an element of $\mathcal{F}$.
3. If $A_1, A_2, \ldots$ are elements of $\mathcal{F}$, the countable union $A_1 \cup A_2 \cup \ldots$ is also an element of $\mathcal{F}$.

From the first two properties it follows that the *empty set* $\emptyset$ belongs to $\mathcal{F}$. From the last two properties and *De Morgan's law* one recognizes that the countable intersection $A_1 \cap A_2 \cap \ldots$ also belongs to $\mathcal{F}$. In probability theory, the elements of $\mathcal{F}$ are called (random) *events*. The most important *operations on events* are the union, the intersection, and the complement:

1. The *union* of a finite or countable sequence $A_1, A_2, \ldots$ of events is an event which occurs if *at least one* of the events $A_1, A_2, \ldots$ occurs; it will be denoted by $A_1 \cup A_2 \cup \ldots$ or by $\bigcup_i A_i$.
2. The *intersection* of a finite or countable sequence $A_1, A_2, \ldots$ of events is an event which occurs if *each one* of the events $A_1, A_2, \ldots$ occurs; it will be denoted by $A_1 \cap A_2 \cap \ldots$ or by $\bigcap_i A_i$.
3. The *complement* of an event A is an event which occurs if and only if A does not occur; it is denoted by $\bar{A}$, $\bar{A} = \{\omega : \omega \notin A\} = \Omega \setminus A$, $A \cup \bar{A} = \Omega$, and $A \cap \bar{A} = \emptyset$.

Important properties of *set operations* are:

- Commutative law : $A \cup B = B \cup A$; $A \cap B = B \cap A$
- Associative law : $A \cup (B \cup C) = (A \cup B) \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C$
- Distributive law : $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- Complement law : $A \cap \bar{A} = \emptyset$; $A \cup \bar{A} = \Omega$
- Idempotent law : $A \cup A = A$; $A \cap A = A$
- De Morgan's law : $\overline{A \cup B} = \bar{A} \cap \bar{B}$; $\overline{A \cap B} = \bar{A} \cup \bar{B}$
- Identity law : $\bar{\bar{A}} = A$; $A \cup (\bar{A} \cap B) = A \cup B$.

The sample space Ω is also called the *sure event* and $\emptyset$ is the *impossible event*. The events $A_1, A_2, \ldots$ are *mutually exclusive* if $A_i \cap A_j = \emptyset$ holds for any $i \neq j$. The events A and B are *equivalent* if either they occur together or neither of them occur, equivalent events have the same probability. In the following, events will be mainly enclosed in braces { }.

A6.2 Concept of Probability

Let us assume that 10 (random) *samples* of size $n = 100$ were taken from a large and homogeneous lot of populated printed circuit boards (PCBs), for incoming inspection. Examination yielded the following results:

Sample number:	1	2	3	4	5	6	7	8	9	10
No. of defective PCBs:	6	5	1	3	4	0	3	4	5	7

For 1000 repetitions of the "testing a PCB" experiment, the *relative frequency* of the occurrence of event {PCB defective} is

$$\frac{6+5+1+3+4+0+3+4+5+7}{1000} = \frac{38}{1000} = 3.8\%.$$

It is *intuitively appealing* to consider 0.038 as the *probability* of the event {PCB defective}. As shown below, 0.038 is a *reasonable estimation* of this probability (on the basis of the experimental observations made).

Relative frequencies of the occurrence of events have the property that if n is the number of trial repetitions and $n(A)$ the number of those trial repetitions in which the event A occurred, then

$$\hat{p}_n(A) = \frac{n(A)}{n} \qquad \text{(A6.1)}$$

is the *relative frequency* of the occurrence of A, and the following rules apply:

1. R1: $\hat{p}_n(A) \geq 0$.
2. R2: $\hat{p}_n(\Omega) = 1$.
3. R3: if the events $A_1, \ldots, A_m$ are *mutually exclusive*, then $n(A_1 \cup \ldots \cup A_m) = n(A_1) + \ldots + n(A_m)$ and $\hat{p}_n(A_1 \cup \ldots \cup A_m) = \hat{p}_n(A_1) + \ldots + \hat{p}_n(A_m)$.

Experience shows that for a second group of n trials, the relative frequency $\hat{p}_n(A)$ can be different from that of the first group. $\hat{p}_n(A)$ also depends on the number of trials n. On the other hand, experiments have confirmed that with increasing n, the value $\hat{p}_n(A)$ converges toward a fixed value $p(A)$, see Fig. A6.1. It therefore seems reasonable to designate the limiting value $p(A)$ as the *probability* $\Pr\{A\}$ of the event A, with $\hat{p}_n(A)$ as an *estimate* of $\Pr\{A\}$. Although intuitive, such a definition of probability would lead to problems in the case of continuous (non-denumerable) sample spaces.

Since Kolmogorov's work [A6.10], the probability $\Pr\{A\}$ has been defined as a function on the event field $\mathcal{F}$ of subsets of Ω. The following axioms hold for this function:

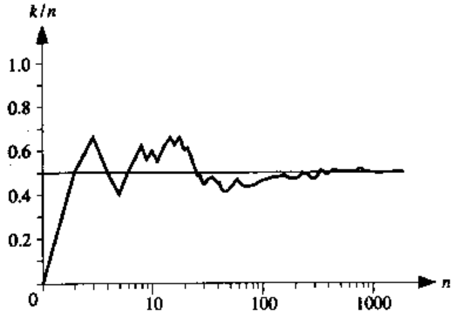


Figure A6.1 Relative frequency k/n of "heads" when tossing a symmetric coin n times

1. Axiom 1: For each $A \in \mathcal{F}$ is $\Pr\{A\} \geq 0$.
2. Axiom 2: $\Pr\{\Omega\} = 1$.
3. Axiom 3: If events $A_1, A_2, \ldots$ are mutually exclusive, then
$$\Pr\{\bigcup_{i=1}^{\infty} A_i\} = \sum_{i=1}^{\infty} \Pr\{A_i\}.$$

Axiom 3 is equivalent to the following statements taken together:

4. Axiom 3': For any *finite* collection of mutually exclusive events, $\Pr\{A_1 \cup \ldots \cup A_n\} = \Pr\{A_1\} + \ldots + \Pr\{A_n\}$.
5. Axiom 3": If events $A_1, A_2, \ldots$ are increasing, i.e. $A_n \subseteq A_{n+1}$, $n = 1, 2, \ldots$,
$$\text{then } \lim_{n \to \infty} \Pr\{A_n\} = \Pr\{\bigcup_{i=1}^{\infty} A_i\}.$$

The relationships between Axiom 1 and R1, and between Axiom 2 and R2 are obvious. Axiom 3 postulates the *total additivity* of the set function $\Pr\{A\}$. Axiom 3' corresponds to R3. Axiom 3" implies a *continuity property* of the set function $\Pr\{A\}$ which cannot be derived from the properties of $\hat{p}_n(A)$, but which is of great importance in probability theory. It should be noted that the interpretation of the probability of an event as the *limit of the relative frequency* of occurrence of this event in a long series of trial repetitions, appears as a theorem within the probability theory (law of large numbers, Eqs. (A6.144) and (A6.146)).

From axioms 1 to 3 it follows that:

$$\Pr\{\emptyset\} = 0,$$

$$\Pr\{A\} \leq \Pr\{B\} \text{ if } A \subseteq B,$$

$$\Pr\{\bar{A}\} = 1 - \Pr\{A\},$$

$$0 \leq \Pr\{A\} \leq 1.$$

When modeling an experiment with random outcome by means of the probability space $[\Omega, \mathcal{F}, \Pr]$, the difficulty is often in the determination of the probabilities $\Pr\{A\}$ for every $A \in \mathcal{F}$. The structure of the experiment can help here. Beside the *statistical probability*, defined as the limit for $n \to \infty$ of the relative frequency k/n, the following rules can be used if one assumes that all elementary events ω have the same chance of occurrence:

1. *Classical probability* (discrete uniform distribution): If Ω is a finite set and A a subset of Ω, then
$$\Pr\{A\} = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega}$$
or
$$\Pr\{A\} = \frac{\text{number of favorable outcomes}}{\text{number of possible outcomes}}. \qquad \text{(A6.2)}$$
2. *Geometric probability* (spatial uniform distribution): If Ω is a set in the plane $\mathcal{R}^2$ of area Ω and A a subset of Ω, then
$$\Pr\{A\} = \frac{\text{area of } A}{\text{area of } \Omega}. \qquad \text{(A6.3)}$$

It should be noted that the geometric probability can also be defined if Ω is a part of the Euclidean space having a finite area. Examples A6.1 and A6.2 illustrate the use of Eqs. (A6.2) and (A6.3).

Example A6.1

From a shipment containing 97 good and 3 defective ICs, one IC is randomly selected. What is the probability that it is defective?

Solution

From Eq. (A6.2),

$$\Pr\{\text{IC defective}\} = \frac{3}{100}.$$

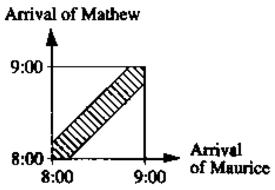
Example A6.2

Maurice and Mathew wish to meet between 8:00 and 9:00 a.m. according to the following rules: 1) They come independently of each other and each will wait 12 minutes. 2) The time of arrival is equally distributed between 8:00 and 9:00 a.m. What is the probability that they will meet?

Solution

Equation (A6.3) can be applied and leads to, see graph,

$$\Pr\{\text{Mathew meets Maurice}\} = \frac{1 - 2\frac{0.8 \cdot 0.8}{2}}{1} = 0.36.$$



Another way to determine probabilities is to calculate them from other probabilities which are known. This involves paying attention to the structure of the experiment and application of the rules of probability theory (Appendix A6.4). For example, the predicted reliability of a system can be calculated from the reliabilities of its elements and the system's structure. However, there is often no alternative to determining probabilities as the limits of relative frequencies, with the aid of statistical methods (Appendices A6.11 and A8).

A6.3 Conditional Probability, Independence

The concept of conditional probability is of great importance in practical applications. It is not difficult to accept that the information "event A has occurred in an experiment" can modify the probabilities of other events. These new probabilities are defined as *conditional probabilities* and denoted by $\Pr\{B \mid A\}$. If for example $A \subseteq B$, then $\Pr\{B \mid A\} = 1$, which is in general different from the original unconditional probability $\Pr\{B\}$. The concept of conditional probability $\Pr\{B \mid A\}$ of the event B under the condition "event A has occurred", is introduced here using the properties of relative frequency. Let n be the total number of trial repetitions and let $n(A)$, $n(B)$, and $n(A \cap B)$ be the number of occurrences of A, B and $A \cap B$, respectively, with $n(A) > 0$ assumed. When considering only the $n(A)$ trials (trials in which A occurs), then B occurs in these $n(A)$ trials exactly when it occurred together with A in the original trial series, i.e. $n(A \cap B)$ times. The relative frequency of B in the trials with the information "A has occurred" is therefore

$$\frac{n(A \cap B)}{n(A)} = \frac{\dfrac{n(A \cap B)}{n}}{\dfrac{n(A)}{n}} = \frac{\hat{p}_n(A \cap B)}{\hat{p}_n(A)}. \tag{A6.4}$$

Equation (A6.4) leads to the following *definition of the conditional probability* $\Pr\{B \mid A\}$ of an event B under the condition A, i.e. assuming that A has occurred,

$$\Pr\{B \mid A\} = \frac{\Pr\{A \cap B\}}{\Pr\{A\}}, \qquad \Pr\{A\} > 0. \tag{A6.5}$$

From Eq. (A6.5) it follows that

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B \mid A\} = \Pr\{B\}\Pr\{A \mid B\}. \tag{A6.6}$$

Using Eq. (A6.5), probabilities $\Pr\{B \mid A\}$ will be defined for all $B \in \mathcal{F}$. $\Pr\{B \mid A\}$ is

a function of B which satisfies Axioms 1 to 3 of Appendix A6.2, obviously with $\Pr\{A \mid A\} = 1$. The information "event A has occurred" thus leads to a new probability space $[A, \mathcal{F}_A, \Pr_A]$, where $\mathcal{F}_A$ consists of events of the form $A \cap B$, with $B \in \mathcal{F}$ and $\Pr_A\{B\} = \Pr\{B \mid A\}$, see Example A6.5.

It is reasonable to define the events A and B as *independent* if the information "event A has occurred" does not influence the *probability* of the occurrence of event B, i.e. if

$$\Pr\{B \mid A\} = \Pr\{B\}. \tag{A6.7}$$

However, when considering Eq. (A6.6), another definition, with symmetry in A and B is obtained, where $\Pr\{A\} > 0$ is not required. Two events A and B are *independent* if and only if

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}. \tag{A6.8}$$

The events $A_1, \dots, A_n$ are *totally independent* if for each $k\,(1 < k \le n)$ and any selection of distinct $i_1, \dots, i_k \in \{1, \dots, n\}$

$$\Pr\{A_{i_1} \cap \dots \cap A_{i_k}\} = \Pr\{A_{i_1}\} \dots \Pr\{A_{i_k}\} \tag{A6.9}$$

holds.

A6.4 Fundamental Rules of Probability Theory

The probability calculation of event combinations is based on the fundamental rules of probability theory introduced in this section.

A6.4.1 Addition Theorem for Mutually Exclusive Events

The events A and B are *mutually exclusive* if the *occurrence* of one event *excludes* the occurrence of the other, formally $A \cap B = \emptyset$. Considering a component which can fail due to a short or an open circuit, the events

failure occurs due to a short circuit

and

failure occurs due to an open circuit

are mutually exclusive. Application of Axiom 3 (Appendix A6.2) leads to

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}. \qquad \text{(A6.10)}$$

Equation (A6.10) is considered a theorem by tradition only; indeed, it is a particular case of Axiom A3 in Appendix A6.2.

Example A6.3

A shipment of 100 diodes contains 3 diodes with shorts and 2 diodes with opens. If one diode is randomly selected from the shipment, what is the probability that it is defective?

Solution

From Eqs. (A6.10) and (A6.2),

$$\Pr\{\text{diode defective}\} = \frac{3}{100} + \frac{2}{100} = \frac{5}{100}$$

If the events $A_1, A_2, \ldots$ are mutually exclusive ($A_i \cap A_j = \emptyset$ for all $i \neq j$, they are also *totally exclusive*, and according to Axiom 3 it follows that

$$\Pr\{A_1 \cup A_2 \cup \ldots\} = \sum_i \Pr\{A_i\}. \qquad \text{(A6.11)}$$

A6.4.2 Multiplication Theorem for Two Independent Events

The events A and B are *independent* if the information about occurrence (or nonoccurrence) of one event has no influence on the *probability* of occurrence of the other event. In this case Eq. (A6.8) applies

$$\Pr\{A \cap B\} = \Pr\{A\}\,\Pr\{B\}.$$

Example A6.4

A system consists of two elements E_1 and E_2 necessary to fulfill the required function. The failure of one element has *no influence* on the other. $R_1 = 0.8$ is the reliability of E_1 and $R_2 = 0.9$ is that of E_2. What is the reliability R_S of the system?

Solution

Considering the assumed independence between the elements E_1 and E_2 and the definition of R_1, R_2, and R_S as $R_1 = \Pr\{E_1 \text{ fulfills the required function}\}$, $R_2 = \Pr\{E_2 \text{ fulfills the required function}\}$, and $R_S = \Pr\{E_1 \text{ fulfills the required function} \cap E_2 \text{ fulfills the required function}\}$, one obtains from Eq. (A6.8)

$$R_S = R_1 R_2 = 0.72.$$

A6.4.3 Multiplication Theorem for Arbitrary Events

For *arbitrary events* A and B, with $\Pr\{A\} > 0$ and $\Pr\{B\} > 0$, Eq. (A6.6) applies

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B \mid A\} = \Pr\{B\}\Pr\{A \mid B\}.$$

Example A6.5

2 ICs are randomly selected from a shipment of 95 good and 5 defective ICs. What is the probability of having (i) no defective ICs, and (ii) exactly one defective IC?

Solution

(i) From Eqs. (A6.6) and (A6.2),

$$\Pr\{\text{first IC good} \cap \text{second IC good}\} = \frac{95}{100}\cdot\frac{94}{99} = 0.902.$$

(ii) $\Pr\{\text{exactly one defective IC}\} = \Pr\{(\text{first IC good} \cap \text{second IC defective}) \cup (\text{first IC defective} \cap \text{second IC good})\}$; from Eqs. (A6.6) and (A6.2),

$$\Pr\{\text{one IC defective}\} = \frac{95}{100}\cdot\frac{5}{99} + \frac{5}{100}\cdot\frac{95}{99} = 0.096.$$

Generalization of Eq. (A6.6) leads to the *multiplication theorem*

$$\Pr\{A_1 \cap \ldots \cap A_n\} = \Pr\{A_1\}\Pr\{A_2 \mid A_1\}\Pr\{A_3 \mid (A_1 \cap A_2)\} \\ \ldots \Pr\{A_n \mid (A_1 \cap \ldots \cap A_{n-1})\}. \qquad \text{(A6.12)}$$

Here, $\Pr\{A_1 \cap \ldots \cap A_{n-1}\} > 0$is assumed. An important special case arises when the events $A_1, \ldots, A_n$ are *totally independent*, in this case Eq. (A6.9) yields

$$\Pr\{A_1 \cap \ldots \cap A_n\} = \Pr\{A_1\} \ldots \Pr\{A_n\} = \prod_{i=1}^{n} \Pr\{A_i\}.$$

A6.4.4 Addition Theorem for Arbitrary Events

The probability of *occurrence of at least one* of the (possibly non-exclusive) events A and B is given by

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}. \qquad \text{(A6.13)}$$

To prove this theorem, consider Axiom 3 (Appendix A6.2) and the partitioning of the events $A \cup B$ and B into mutually exclusive events ($A \cup B = A \cup (\overline{A} \cap B)$ and $B = (A \cap B) \cup (\overline{A} \cap B)$).

Example A6.6

To increase the reliability of a system, 2 machines are used in active (parallel) redundancy. The reliability of each machine is 0.9 and each machine operates and fails *independently* of the other. What is the system's reliability?

Solution

From Eqs. (A6.13) and (A6.8), $\Pr\{$the first machine fulfills the required function $\cup$ the second machine fulfills the required function$\} = 0.9 + 0.9 - 0.9 \cdot 0.9 = 0.99$.

The *addition theorem* can be generalized to n arbitrary events. For $n = 3$ one obtains

$$\begin{aligned}\Pr\{A \cup B \cup C\} &= \Pr\{A \cup (B \cup C)\}\\ &= \Pr\{A\} + \Pr\{B \cup C\} - \Pr\{A \cap (B \cup C)\}\\ &= \Pr\{A\} + \Pr\{B\} + \Pr\{C\} - \Pr\{B \cap C\} - \Pr\{A \cap B\}\\ &\quad - \Pr\{A \cap C\} + \Pr\{A \cap B \cap C\}. \qquad \text{(A6.14)}\end{aligned}$$

In general, $\Pr\{A_1 \cup \ldots \cup A_n\}$ can be obtained by the so-called *inclusion/exclusion* method

$$\Pr\{A_1 \cup \ldots \cup A_n\} = \sum_{k=1}^{n} (-1)^{k+1} S_k \qquad \text{(A6.15)}$$

with

$$S_k = \sum_{1 \le i_1 < \ldots < i_k \le n} \Pr\{A_{i_1} \cap \ldots \cap A_{i_k}\}. \qquad \text{(A6.16).}$$

It can be shown that $S = \Pr\{A_1 \cup \ldots \cup A_n\} \le S_1$, $S \ge S_1 - S_2$, $S \le S_1 - S_2 + S_3$, etc. Although the upper bounds do not necessarily decrease and the lower bounds do not necessarily increase, a good approximation for S often results from only a few S_i. For a further investigation one can use the *Fréchet theorem* $S_{k+1} \le S_k\,(n-k)/(k+1)$, which follows from $S_{k+1} = S_k \binom{n}{k+1} / \binom{n}{k} = S_k (n-k)/(k+1) < S_k$ for $A_1 = A_2 = \cdots = A_n$.

A6.4.5 Theorem of Total Probability

Let $A_1, A_2, \ldots$ be *mutually exclusive* events ($A_i \cap A_j = \emptyset$ for all $i \ne j$), $\Omega = A_1 \cup A_2 \cup \ldots$, and $\Pr\{A_i\} > 0$, $i = 1, 2, \ldots$. For an arbitrary event B one has $B = B \cap \Omega = B \cap (A_1 \cup A_2 \cup \ldots) = (B \cap A_1) \cup (B \cap A_2) \cup \ldots$, where the events $B \cap A_1$, $B \cap A_2$, ... are also mutually exclusive. Use of Axiom 3 and Eq. (A6.6) yields

$$\Pr\{B\} = \sum_i \Pr\{B \cap A_i\} = \sum_i \Pr\{A_i\} \Pr\{B \mid A_i\}. \qquad \text{(A6.17)}$$

Equation (A6.17) expresses the *theorem* (or formula) *of total probability*.

Example A6.7

ICs are purchased from 3 suppliers (A_1, A_2, A_3) in quantities of 1000, 600, and 400 pieces, respectively. The probabilities for an IC to be defective are 0.006 for A_1, 0.02 for A_2, and 0.03 for A_3. The ICs are stored in a common container disregarding their source. What is the probability that one IC randomly selected from the stock is defective?

Solution

From Eqs. (A6.17) and (A6.2),

$$\Pr\{\text{the selected IC is defective}\} = \frac{1000}{2000}\,0.006 + \frac{600}{2000}\,0.02 + \frac{400}{2000}\,0.03 = 0.015.$$

Equations (A6.17) and (A6.6) lead to *Bayes theorem*, which allows calculation of the *a posteriori probability* $\Pr\{A_k \mid B\}$, $k = 1, 2, \ldots$ as a function of *a priori probabilities* $\Pr\{A_i\}$,

$$\Pr\{A_k \mid B\} = \frac{\Pr\{A_k \cap B\}}{\Pr\{B\}} = \frac{\Pr\{A_k\} \Pr\{B \mid A_k\}}{\sum_i \Pr\{A_i\} \Pr\{B \mid A_i\}}. \qquad \text{(A6.18)}$$

Example A6.8

Let the IC as selected in Example A6.7 be defective. What is the probability that it is from supplier A_1?

Solution

From Eq. (A6.18),

$$\Pr\{\text{IC from } A_1 \mid \text{IC defective}\} = \frac{\frac{1000}{2000}\,0.006}{0.015} = 0.2.$$

A6.5 Random Variables, Distribution Functions

If the result of an experiment with a random outcome is a (real) number, then the underlying quantity is a (real) *random variable*. For example, the number appearing when trowing a die is a random variable taking on values in $\{1, \ldots, 6\}$. Random variables are designated here with Greek letters τ, ξ, ζ, etc. The triplet $[\Omega, \mathcal{F}, \Pr]$ introduced in Appendix A6.2 becomes $[\mathcal{R}, \mathcal{B}, \Pr]$, where $\mathcal{R} = (-\infty, \infty)$ and $\mathcal{B}$ is the smallest event field containing all (semi) intervals $(a, b]$ with $a < b$. The probabilities $\Pr\{A\} = \Pr\{\tau \in A\}$, $A \in \mathcal{B}$, define the distribution law of the random variable τ. Among the many possibilities to characterize this *distribution law*, the most frequently used is to define

$$F(t) = \Pr\{\tau \le t\}. \qquad \text{(A6.19)}$$

$F(t)$ is called the *distribution function* of the *random variable* τ.*) For each t, $F(t)$ gives the *probability* that the random variable will assume a value smaller than or equal to t. Because for $s > t$ one has $\{\tau \le t\} \le \{\tau \le s\}$, $F(t)$ is a *nondecreasing function*. Moreover, $F(-\infty) = 0$ and $F(\infty) = 1$. If $\Pr\{\tau = t_0\} > 0$ holds, then $F(t)$ has a jump of height $\Pr\{\tau = t_0\}$ at t_0. It follows from the above definition and Axiom 3" (Appendix A6.2) that $F(t)$ is *continuous from the right*. Due to Axiom 2, $F(t)$ can have at most a countable number of jumps. The probability that the random variable τ takes on a value within the interval $(a, b]$ is given by

$$\Pr\{a < \tau \le b\} = F(b) - F(a). \tag{A6.20}$$

The following classes of random variables are of particular importance:

1. *Discrete random variables*: A random variable τ is discrete if it can only assume a finite or a countable number of values, i.e. if there is a sequence $t_1, t_2, \ldots$ such that

$$p_k = \Pr\{\tau = t_k\}, \qquad \text{with} \quad \sum_k p_k = 1. \tag{A6.21}$$

A discrete random variable is best described by a table

Values of τ	t_1	t_2	...
Probabilities	p_1	p_2	...

The distribution function $F(t)$ of a discrete random variable τ is a *step function*

$$F(t) = \sum_{k:\, t_k \le t} p_k .$$

If the sequence $t_1, t_2, \ldots$ is ordered so that $t_k < t_{k+1}$, then

$$F(t) = \sum_{j \le k} p_j, \qquad \text{for} \quad t_k \le t < t_{k+1}. \tag{A6.22}$$

If only the value $k = 1$ occurs in Eqs. (A6.21) and (A6.22), τ is a *constant* ($\tau = t_1$). A constant C can therefore be regarded as a random variable which takes the value C with probability 1. The distribution function of $\tau = C$ is

$$F(t) = \begin{cases} 0 & \text{for } t < C \\ 1 & \text{for } t \ge C. \end{cases}$$

---

*) In textbooks on probability theory, the random variable τ is defined as a *measurable mapping* of Ω onto the axis of real numbers $\mathcal{R} = (-\infty, \infty)$, i.e. a mapping such that for each real value x the set of ω for which $\{\tau = \tau(\omega) \le x\}$ belongs to $\mathcal{F}$, the distribution function of τ is then obtained by setting $F(t) = \Pr\{\tau \le t\} = \Pr\{\omega : \tau(\omega) \le t\}$.

An especially important case of discrete random variables is that of *arithmetic random variables*. The random variable τ is arithmetic if it can take the values $\ldots, -\Delta t, 0, \Delta t, \ldots$, with probabilities

$$p_k = \Pr\{\tau = k\,\Delta t\}, \qquad k = \ldots, -1, 0, 1, \ldots . \tag{A6.23}$$

2. *Continuous random variables*: The random variable τ is (absolutely) continuous if a function $F(t) \ge 0$ exists such that

$$F(t) = \Pr\{\tau \le t\} = \int_{-\infty}^{t} f(x)\,dx. \tag{A6.24}$$

$f(t)$ is called (probability) *density* of the random variable τ and satisfies the condition

$$\int_{-\infty}^{\infty} f(t)\,dt = 1.$$

The distribution function $F(t)$ and the density $f(t)$ are related (almost everywhere) by

$$f(t) = \frac{dF(t)}{dt}, \tag{A6.25}$$

see Fig. A6.2 for an example.

Mixed distribution functions exhibiting both jumps and continuous growth can occur in some applications. These distribution functions can generally be represented by a mixture (weighted sum) of discrete and continuous distribution functions (Eq. (A6.34)).

In reliability theory, τ represents the *failure-free operating time* of the item under consideration. τ is then a *nonnegative random variable*, i.e. $\tau \ge 0$ and $F(t) = 0$ for $t < 0$. Often τ is a *positive random variable*, i.e. $\tau > 0$ and $F(0) = 0$. The *reliability function* (survival function) $R(t)$ of an item gives the probability that the item will operate failure-free in the interval $(0, t]$

$$R(t) = \Pr\{\tau > t\} = 1 - F(t). \tag{A6.26}$$

The *failure rate* $\lambda(t)$ of an item exhibiting a *continuous* failure-free operating time τ is defined as

$$\lambda(t) = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{t < \tau \le t + \delta t \mid \tau > t\}.$$

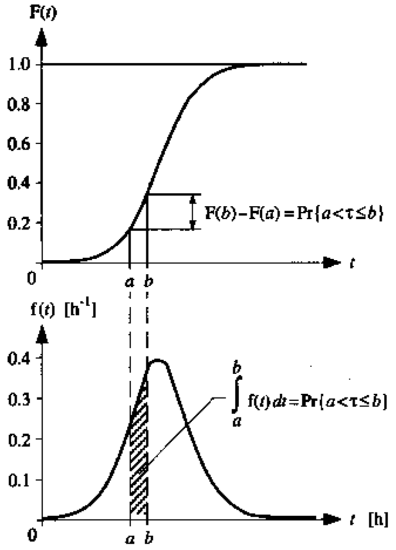


Figure A6.2 Relationship between the distribution function $F(t)$ and the density $f(t)$ for a continuous random variable $\tau > 0$

Computation leads to (see Eq. (A6.5) and Fig. A6.3)

$$\lambda(t) = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \frac{\Pr\{t < \tau \le t + \delta t \cap \tau > t\}}{\Pr\{\tau > t\}} = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \frac{\Pr\{t < \tau \le t + \delta t\}}{\Pr\{\tau > t\}},$$

and thus, assuming $f(t) = dF(t)/dt$ exists,

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = -\frac{dR(t)/dt}{R(t)}. \tag{A6.27}$$

It is important to distinguish between *failure rate* $\lambda(t)$ and *density* $f(t)$. For an item with a failure-free operating time τ, the value $f(t)\delta t$ is for small δt the *unconditional probability* for failure in $(t, t+\delta t]$. On the other hand, the quantity $\lambda(t)\delta t$ is for small δt the *conditional probability* that the item will fail in the interval $(t, t+\delta t]$ *given* that it has not failed in $(0, t]$. From Eq. (A6.27) it follows that, with $R(0) = 1$,

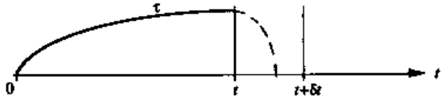


Figure A6.3 To compute of the failure rate $\lambda(t)$

$$\Pr\{t > \tau\} = R(t) = e^{-\int_0^t \lambda(x)dx} \tag{A6.28}$$

and thus

$$\Pr\{a < \tau \le b\} = e^{-\int_0^a \lambda(x)dx} - e^{-\int_0^b \lambda(x)dx},$$

or

$$\Pr\{\tau > t + x_0 \mid \tau > x_0\} = \frac{R(t + x_0)}{R(x_0)} = e^{-\int_{x_0}^{t+x_0} \lambda(x)dx}. \tag{A6.29}$$

Important conclusions as to the *aging behavior* of an item can be drawn from the shape of its failure rate $\lambda(t)$. Assuming $\lambda(t)$ is *non-decreasing*, it follows for $u < s$ and any $t > 0$ that

$$\Pr\{\tau > t + u \mid \tau > u\} \ge \Pr\{\tau > t + s \mid \tau > s\}. \tag{A6.30}$$

For an item with increasing failure rate, inequality (A6.30) shows that the probability of surviving a further period t decreases as a function of the achieved age, i.e. the *item ages*. The opposite holds for an item with decreasing failure rate. No aging exists in the case of a *constant failure rate*, i.e. for $R(t) = e^{-\lambda t}$

$$\Pr\{\tau > t + x_0 \mid \tau > x_0\} = \Pr\{\tau > t\} = e^{-\lambda t}.$$

The failure rate of an arithmetic random variable, is defined as

$$\lambda(k) = \Pr\{\tau = k\Delta t \mid \tau > (k-1)\Delta t\} = \frac{p_k}{\sum_{i \ge k} p_i}, \qquad k = 1, 2, \ldots. \tag{A6.31}$$

Two further concepts important to reliability theory are:

1. *Function of a random variable*: If $u(x)$ is a monotonically increasing function and τ a continuous random variable with the distribution function $F_\tau(t)$, then the random variable $\eta = u(\tau)$ has the distribution function

   $$F_\eta(t) = \Pr\{\eta = u(\tau) \le t\} = \Pr\{\tau \le u^{-1}(t)\} = F_\tau(u^{-1}(t)), \tag{A6.32}$$

   where u^{-1} is the *inverse function* of u (Example A6.17). If $u(x)$ is derivable, then

   $$f_\eta(t) = f_\tau(u^{-1}(t)) \frac{d\,u^{-1}(t)}{dt}. \tag{A6.33}$$

If $u(x)$ is a monotonically decreasing function, the absolute value must be used for $d\,u^{-1}(t)/dt$.

2. *Mixture of distributions*: If $F_1(t)$ and $F_2(t)$ are distribution functions and p is a constant ($0 < p < 1$), then the distribution function

$$F(t) = p F_1(t) + (1-p) F_2(t) \tag{A6.34}$$

is the mixture of $F_1(t)$and $F_2(t)$ with weights p and $1-p$ (Example 7.16). Mixture also applies for more than 2 distribution functions and for densities as well

$$f(t) = p f_1(t) + (1-p) f_2(t).$$

The main properties of distribution functions frequently used in reliability theory are discussed in Appendix A6.10 and summarized in Table A6.1. Of the following sections preceding Appendix A6.10, Appendices A6.7 and A6.8 *can be omitted at a first reading.*

A6.6 Numerical Parameters of Random Variables

For a rough characterization of a random variable τ, some typical values such as the expected value (mean), variance, and median can be used. This section introduces these quantities.

A6.6.1 Expected Value (Mean)

For a discrete random variable τ taking values

$$t_1, t_2, \ldots$$

with probabilities

$$p_1, p_2, \ldots$$

the *expected value* or *mean* $E[\tau]$ is given by

$$E[\tau] = \sum_k t_k\, p_k, \tag{A6.35}$$

provided the series converges absolutely. If τ only takes the values $t_1, \ldots, t_m$, the definition (A6.35) can be heuristically explained as follows. Consider n repetitions of a trial whose outcome is τ and assume that k_1 times the value $t_1, \ldots, k_m$ times the value t_m has been observed ($n = k_1 + \ldots + k_m$), the arithmetic mean of the observed values is

$$\frac{t_1 k_1 + \ldots + t_m k_m}{n} = t_1 \frac{k_1}{n} + \ldots + t_m \frac{k_m}{n}.$$

As $n \to \infty$, k_i / n converges to p_i (Eq. (A6.146)), and the arithmetic mean obtained above tends towards the *expected value* $E[\tau]$ given by Eq. (A6.35). For this reason, the terms *expected value* and *mean* are often used for the same quantity $E[\tau]$. From Eq. (A6.35), the *mean of a constant C* is the constant itself, i.e. $E[C] = C$.

The mean of a continuous random variable τ with density $f(t)$ is given by

$$E[\tau] = \int_{-\infty}^{\infty} t\, f(t)\, dt, \tag{A6.36}$$

provided the integral converges absolutely. For positive continuous random variables, Eq. (A6.36) reduces to

$$E[\tau] = \int_0^{\infty} t\, f(t)\, dt \tag{A6.37}$$

which, for $E[\tau] < \infty$ can be expressed (Example A6.9) as

$$E[\tau] = \int_0^{\infty} (1 - F(t))\, dt = \int_0^{\infty} R(t)\, dt. \tag{A6.38}$$

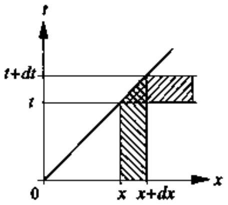
Example A6.9

Prove the equivalence of Eqs. (A6.37) and (A6.38).

Solution

From $R(t) = 1 - F(t) = \int_t^{\infty} f(x)\, dx$ it follows that

$$\int_0^{\infty} R(t)\, dt = \int_0^{\infty} \Big(\int_t^{\infty} f(x)\, dx \Big) dt.$$

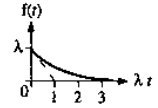
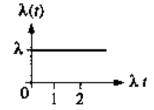
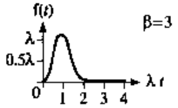
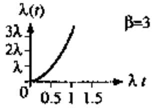
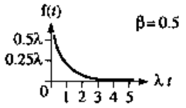
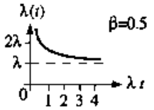
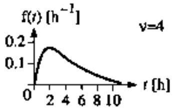
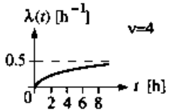
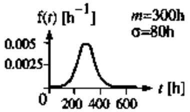
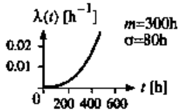
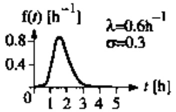
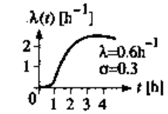
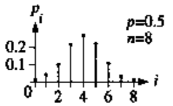
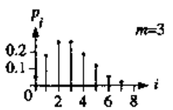
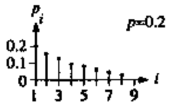
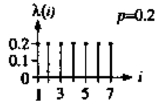
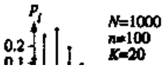


Changing the order of integration yields (see graph)

$$\int_0^{\infty} R(t)\, dt = \int_0^{\infty} \Big(\int_0^{x} dt \Big) f(x)\, dx = \int_0^{\infty} x\, f(x)\, dx.$$

Table A6.1 Distribution functions used in reliability analysis

Table A6.1 (cont)

Name	Distribution Function $F(t) = \Pr\{\tau \le t\}$	Density $f(t) = dF(t)/dt$	Parameter Range	Failure Rate $\lambda(t) = f(t)/(1-F(t))$	Mean $E[\tau]$	Variance $Var[\tau]$	Properties
Exponential	$1-e^{-\lambda t}$		$t \ge 0$ $\lambda > 0$		$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Memoryless: $\Pr\{\tau > t + x_0 \mid \tau > x_0\} = \Pr\{\tau > t\} = e^{-\lambda t}$
Weibull	$1-e^{-(\lambda t)^\beta}$		$t \ge 0$ $\lambda, \beta > 0$		$\frac{\Gamma(1+\frac{1}{\beta})}{\lambda}$	$\frac{\Gamma(1+\frac{2}{\beta}) - \Gamma^2(1+\frac{1}{\beta})}{\lambda^2}$	Monotonic failure rate: increasing for $\beta > 1$ $(\lambda(0) = 0,\ \lambda(\infty) = \infty)$, decreasing for $\beta < 1$ $(\lambda(0) = \infty,\ \lambda(\infty) = 0)$
Gamma	$\frac{1}{\Gamma(\beta)} \int_0^{\lambda t} x^{\beta-1} e^{-x} dx$		$t \ge 0$ $\lambda, \beta > 0$		$\frac{\beta}{\lambda}$	$\frac{\beta}{\lambda^2}$	Laplace transf. exists: $\tilde{f}(s) = \lambda^\beta / (s+\lambda)^\beta$; Monotonic failure rate with $\lambda(\infty) = \lambda$; Erlangian for $\beta = n = 2, 3, \ldots$ (distribution of the sum of n exp. dist. random variables)
Chi-square (χ^2)	$\frac{\int_0^t x^{\nu/2-1} e^{-x/2} dx}{2^{\nu/2}\,\Gamma(\nu/2)}$		$t \ge 0$ $\nu = 1, 2, \ldots$ (degrees of freedom)		ν	2ν	Gamma with $\beta = \frac{\nu}{2} = 1, 2, \ldots$ and $\lambda = \frac{1}{2}$; for $\nu = 2, 4, \ldots \Rightarrow F(t) = 1 - \sum_{i=0}^{(\nu/2)-1} \frac{(t/2)}{i!} e^{-t/2}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$		$\infty < t,\ m < \infty$ $\sigma > 0$		m	σ^2	$F(t) = \Phi(\frac{t-m}{\sigma})$ $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$
Lognormal	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln(\lambda t)}{\sigma}} e^{-x^2/2} dx$		$t \ge 0$ $\lambda, \sigma > 0$		$\frac{e^{\sigma^2/2}}{\lambda}$	$\frac{e^{2\sigma^2} - e^{\sigma^2}}{\lambda^2}$	$\ln\tau$ has a normal distribution; $F(t) = \Phi(\ln(\lambda t)/\sigma)$
Binomial	$\Pr\{\zeta \le k\} = \sum_{i=0}^{k} p_i$ $p_i = \binom{n}{i} p^i (1-p)^{n-i}$		$k = 0, \ldots, n$ $0 < p < 1$	not relevant	$n\,p$	$n\,p(1-p)$	$p_i = \Pr\{i$ successes in n Bernoulli trials$\}$ (n independent trials with $\Pr\{A\} = p$); Random sample with replacement
Poisson	$\Pr\{\zeta \le k\} = \sum_{i=0}^{k} p_i$ $p_i = \frac{m^i}{i!} e^{-m}$		$k = 0, 1, \ldots$ $m > 0$	not relevant	m	m	$\binom{n}{i} p^i (1-p)^{n-i} \approx \frac{(n\,p)^i}{i!} e^{-n\,p}$; $m = \lambda t \Rightarrow (\lambda t)^i e^{-\lambda t} / i! = \Pr\{i$ failures in $(0, t] \mid$ exponentially distributed failure-free times with param. $\lambda\}$
Geometric	$\Pr\{\zeta \le k\} = \sum_{i=1}^{k} p_i = 1-(1-p)^k$ $p_i = p(1-p)^{i-1}$		$k = 1, 2, \ldots$ $0 < p < 1$		$\frac{1}{p}$	$\frac{1-p}{p^2}$	Memoryless: $\Pr\{\zeta > i + j \mid \zeta > i\} = (1-p)^j$; $p_i = \Pr\{$first success in a sequence of Bernoulli trials occurs first at the i-th trial$\}$; ($p_i = p(1-p)^i$, with $i = 0, 1, \ldots$, would give $E[\zeta] = (1-p)/p$ and $Var[\zeta] = (1-p)/p^2$)
Hyper-geometric	$\Pr\{\zeta \le k\} = \sum^{k} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$		$i = 0, 1, \ldots$ $\ldots, \min(K, n)$	not relevant	$n\frac{K}{N}$	$\frac{K\,n(N-K)(N-n)}{N^2(N-1)}$	Random sample without replacement

For the expected value of the random variable $\eta = \mathrm{u}(\tau)$

$$\mathrm{E}[\eta] = \sum_k \mathrm{u}(t_k)\, p_k \qquad \text{or} \qquad \mathrm{E}[\eta] = \int_{-\infty}^{\infty} \mathrm{u}(t)\mathrm{f}(t)\,dt \tag{A6.39}$$

holds, provided that the series and the integral converge absolutely. Two particular cases of Eq. (A6.39) are:

1. $\mathrm{u}(x) = Cx$,

$$\mathrm{E}[C\tau] = \int_{-\infty}^{\infty} Ct\,\mathrm{f}(t)\,dt = C\,\mathrm{E}[\tau]. \tag{A6.40}$$

2. $\mathrm{u}(x) = x^k$, which leads to the k^{th} *moment* of τ,

$$\mathrm{E}[\tau^k] = \int_{-\infty}^{\infty} t^k\,\mathrm{f}(t)\,dt, \qquad k > 1. \tag{A6.41}$$

Further important properties of the mean are given by Eqs. (A6.68) and (A6.69).

A6.6.2 Variance

The *variance* of a random variable τ is a measure of the spread (or dispersion) of the random variable around its mean $\mathrm{E}[\tau]$. Variance is defined as

$$\mathrm{Var}[\tau] = \mathrm{E}[(\tau - \mathrm{E}[\tau])^2], \tag{A6.42}$$

and can be computed as

$$\mathrm{Var}[\tau] = \sum_k (t_k - \mathrm{E}[\tau])^2\, p_k \tag{A6.43}$$

for a discrete random variable, and as

$$\mathrm{Var}[\tau] = \int_{-\infty}^{\infty} (\tau - \mathrm{E}[\tau])^2\,\mathrm{f}(t)\,dt \tag{A6.44}$$

for a continuous random variable. In both cases,

$$\mathrm{Var}[\tau] = \mathrm{E}[\tau^2] - (\mathrm{E}[\tau])^2. \tag{A6.45}$$

If $\mathrm{E}[\tau]$ or $\mathrm{Var}[\tau]$ is infinite, τ is said to have an infinite variance. For arbitrary constants C and A, Eqs. (A6.45) and (A6.40) yield

$$\mathrm{Var}[C\tau - A] = C^2\,\mathrm{Var}[\tau] \tag{A6.46}$$

and

$$\mathrm{Var}[C] = 0.$$

The quantity

$$\sigma = \sqrt{\mathrm{Var}[\tau]} \tag{A6.47}$$

is the *standard deviation* of τ and, for $\tau \ge 0$,

$$\kappa = \frac{\sigma}{\mathrm{E}[\tau]} \tag{A6.48}$$

is the *coefficient of variation* of τ. The random variable

$$(\tau \mathrm{E}[\tau])/\sigma$$

has mean 0 and variance 1, and is a *standardized random variable.*

A good understanding of the variance as a measure of dispersion is given by the *Chebyshev inequality*, which states (Example A6.10) that for every $\varepsilon > 0$

$$\Pr\{|\tau - \mathrm{E}[\tau]| \ge \varepsilon\} \le \frac{\mathrm{Var}[\tau]}{\varepsilon^2}. \tag{A6.49}$$

The Chebyshev inequality is more useful in proving convergence than as an approximation. Further important properties of the variance are given by Eqs. (A6.70) and (A6.71).

Example A6.10

Prove the Chebyshev inequality for a continuous random variable (Eq. (A6.49)).

Solution

For a continuous random variable τ with density $\mathrm{f}(t)$, the definition of the variance implies

$$\Pr\{|\tau - \mathrm{E}[\tau]| > \varepsilon\} = \int_{|t-E[\tau]|>\varepsilon} \mathrm{f}(t)\,dt \le \int_{|t-E[\tau]|>\varepsilon} \frac{(t-\mathrm{E}[\tau])^2}{\varepsilon^2}\mathrm{f}(t)\,dt$$
$$\le \int_{-\infty}^{\infty} \frac{(t-\mathrm{E}[\tau])^2}{\varepsilon^2}\mathrm{f}(t)\,dt = \frac{1}{\varepsilon^2}\mathrm{Var}[\tau],$$

which proves Eq. (A6.49).

Generalization of the exponent in Eqs. (A6.43) and (A6.44) leads to the k^{th} *central moment* of τ

$$\mathrm{E}[(\tau - \mathrm{E}[\tau])^k] = \int_{-\infty}^{\infty} (t - \mathrm{E}[\tau])^k\,\mathrm{f}(t)\,dt, \qquad k > 1. \tag{A6.50}$$

A6.6.3 Modal Value, Quantile, Median

In addition to the moments discussed in Appendices A6.6.1 and A6.6.2, the modal value, quantile, and median are defined as follows:

1. For a continuous random variable τ, the *modal value* is the value of t for which $f(t)$ reaches its maximum, the distribution of τ is *multimodal* if $f(t)$ exhibits more than one maximum.
2. The q *quantile* is the value t_q for which $F(t)$ reaches the value q, $t_q = \inf\{t : F(t) \geq q\}$; in general, $F(t_q) = q$ for a continuous random variable.
3. The 0.5 quantile is the *median*.

A6.7 Multidimensional Random Variables, Conditional Distributions

Multidimensional random variables (random vectors) are often required in reliability and availability investigations of repairable systems. Here, the outcome of an experiment is an element of the n-dimensional space $\mathcal{R}^n$. The probability space $\{\Omega, \mathcal{F}, \Pr]$ introduced in Appendix A6.1 becomes $[\mathcal{R}^n, \mathcal{B}^n, T]$, where $\mathcal{B}^n$ is the smallest event field which contains all "intervals" of the form $(a_1, b_1] \cdot \ldots \cdot (a_n, b_n] = \{(t_1, \ldots, t_n) : t_i \in (a_i, b_i], i = 1, \ldots, n\}$. *Random vectors* are designated by Greek letters with an arrow, i.e. $\vec{\tau} = (\tau_1, \ldots, \tau_n)$, $\vec{\xi} = (\xi_1, \ldots, \xi_n)$, etc. The probabilities $\Pr\{A\} = \Pr\{\vec{\tau} \in A\}$, $A \in \mathcal{B}^n$ define the *distribution* law of $\vec{\tau}$. The function

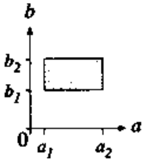
$$F(t_1, \ldots, t_n) = \Pr\{\tau_1 \leq t_1, \ldots, \tau_n \leq t_n\}, \tag{A6.51}$$

where

$$\{\tau_1 \leq t_1, \ldots, \tau_n \leq t_n\} \equiv \{(\tau_1 \leq t_1) \cap \ldots \cap (\tau_n \leq t_n)\}$$

is the *distribution function* of the *random vector* $\vec{\tau}$. $F(t_1, \ldots, t_n)$ is:

- monotonically nondecreasing in each variable,
- zero (in the limit) if at least one variable goes to $-\infty$,
- one (in the limit) if all variables go to ∞,
- continuous from the right in each variable,
- such that the probabilities $\Pr\{a_1 < \tau_1 \leq b_1, \ldots, a_n < \tau_n \leq b_n\}$, computed for arbitrary $a_1, \ldots, a_n$, $b_1, \ldots, b_n$ with $a_i < b_i$, are not negative ($n = 2$ yields for example, to $\Pr\{a_1 < \tau_1 \leq b_1, a_2 < \tau_2 \leq b_2\} = F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1)$, see graph).



It can be shown that every component τ_i of $\vec{\tau} = (\tau_1, \ldots, \tau_n)$ is a random variable with distribution function, ***marginal distribution function***,

$$F_i(t_i) = \Pr\{\tau_i \leq t_i\} = F(\infty, \ldots, \infty, t_i, \infty, \ldots, \infty). \tag{A6.52}$$

The components $\tau_1, \ldots, \tau_n$ of $\vec{\tau}$ are called *(stochastically) independent* if for any $(t_1, \ldots, t_n) \in \mathcal{R}^n$

$$F(t_1, \ldots, t_n) = \prod_{i=1}^{n} F_i(t_i). \tag{A6.53}$$

It can be shown that Eq. (A6.53) is equivalent to

$$\Pr\{\bigcap_{i=1}^{n} (\tau_i \in B_i)\} = \prod_{i=1}^{n} \Pr\{(\tau_i \in B_i)\}$$

for every $B_i \in \mathcal{B}^n$.

The random vector $\vec{\tau} = (\tau_1, \ldots, \tau_n)$ is (absolutely) *continuous* if a function $f(x_1, \ldots, x_n) \geq 0$ exists such that for any $t_1, \ldots, t_n$

$$F(t_1, \ldots, t_n) = \int_{-\infty}^{t_1} \ldots \int_{-\infty}^{t_n} f(x_1, \ldots, x_n) dx_1 \ldots dx_n. \tag{A6.54}$$

$f(x_1, \ldots, x_n)$ is the *density* of $\vec{\tau}$ and satisfies the condition

$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1.$$

For any subset $A \in \mathcal{B}^n$, it follows that

$$\Pr\{(\tau_1, \ldots, \tau_n) \in A\} = \int_A \ldots \int f(t_1, \ldots, t_n) dt_1 \ldots dt_n. \tag{A6.55}$$

The density of τ_i, *marginal density*, can be obtained from $f(t_1, \ldots, t_n)$ as

$$f_i(t_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(t_1, \ldots, t_n) dt_1 \ldots dt_{i-1} dt_{i+1} \ldots dt_n. \tag{A6.56}$$

The components $\tau_1, \ldots, \tau_n$ of a continuous random vector $\vec{\tau}$ are (stochastically) *independent* if and only if, for any $t_1, \ldots, t_n \in \mathcal{R}^n$

$$f(t_1, \ldots, t_n) = \prod_{i=1}^{n} f_i(t_i). \tag{A6.57}$$

For a two dimensional continuous random vector $\vec{\tau} = (\tau_1, \tau_2)$, the function

$$f_2(t_2 \mid t_1) = \frac{f(t_1, t_2)}{f_1(t_1)} \tag{A6.58}$$

is the *conditional density* of τ_2 under the condition $\tau_1 = t_1$ and with $f_1(t_1) > 0$. Similarly $f_1(t_1|t_2) = f(t_1, t_2)/f_2(t_2)$ is the conditional density for τ_1 given $\tau_2 = t_2$. For the marginal density of τ_2 it follows then that

$$f_2(t_2) = \int_{-\infty}^{\infty} f(t_1, t_2)\,dt_1 = \int_{-\infty}^{\infty} f_1(t_1) f_2(t_2 \mid t_1)\,dt_1. \tag{A6.59}$$

Therefore, for any $A \in \mathcal{B}^2$

$$\Pr\{\tau_2 \in A\} = \int_A f_2(t_2)\,dt_2 = \int_A \Big(\int_{-\infty}^{\infty} f_1(t_1) f_2(t_2|t_1)\,dt_1 \Big) dt_2, \tag{A6.60}$$

and in particular

$$F_2(t) = \Pr\{\tau_2 \le t\} = \int_{-\infty}^{t} f_2(t_2)\,dt_2 = \int_{-\infty}^{t} \int_{-\infty}^{\infty} f_1(t_1) f_2(t_2 \mid t_1)\,dt_1\,dt_2. \tag{A6.61}$$

A6.8 Numerical Parameters of Random Vectors

Let $\vec{\tau} = (\tau_1, \ldots, \tau_n)$ be a random vector, and u a real-valued function in $\mathcal{R}^n$. The *expected value* or *mean* of the random variable $u(\vec{\tau})$ is

$$E[u(\vec{\tau})] = \sum_{i_1=1}^{k_1} \ldots \sum_{i_n=1}^{k_n} u(t_{1,i_1}, \ldots, t_{n,i_n})\, p(i_1, \ldots, i_n) \tag{A6.62}$$

for the discrete case and

$$E[u(\vec{\tau})] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} u(t_1, \ldots, t_n)\, f(t_1, \ldots, t_n)\,dt_1 \ldots dt_n \tag{A6.63}$$

for the continuous case [A6.7, A6.14], assuming that the series and the integral converge absolutely.

The *conditional expected value* of τ_2 given $\tau_1 = t_1$ follows, in the continuous case, from Eqs. (A6.36) and (A6.58) as

$$E[\tau_2 \mid \tau_1 = t_1] = \int_{-\infty}^{\infty} t_2\, f(t_2 \mid t_1)\,dt_2. \tag{A6.64}$$

Thus the *unconditional expected value* of τ_2 can be obtained from

$$E[\tau_2] = \int_{-\infty}^{\infty} E[\tau_2 \mid \tau_1 = t_1]\, f_1(t_1)\,dt_1. \tag{A6.65}$$

Equation (A6.65) is known as the formula of *total expectation* and is useful in many practical applications.

A6.8.1 Covariance Matrix, Correlation Coefficient

Assuming for $\vec{\tau} = (\tau_1, \ldots, \tau_n)$ that $\mathrm{Var}[\tau_i] < \infty$, $i = 1, \ldots, n$, an important rough characterization of a random vector is the *covariance matrix* (a_{ij}), where

$$a_{ij} = \mathrm{Cov}[\tau_i, \tau_j] = E[(\tau_i - E[\tau_i])(\tau_j - E[\tau_j])]$$

are given in the continuous case by

$$a_{ij} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} (t_i - E[\tau_i])(t_j - E[\tau_j])\, f(t_1, \ldots, t_n)\,dt_1 \ldots dt_n. \tag{A6.66}$$

The diagonal elements of the covariance matrix are the *variances* of components τ_i, $i = 1, \ldots, n$. Elements outside the diagonal give a measure of the degree of dependency between components (obviously $a_{ij} = a_{ji}$). For τ_i *independent* of τ_j, $a_{ij} = a_{ji} = 0$ holds.

For a two dimensional random vector $\vec{\tau} = (\tau_1, \tau_n)$, the quantity

$$\rho(\tau_1, \tau_2) = \frac{\mathrm{Cov}[\tau_1, \tau_2]}{\sigma_1 \cdot \sigma_2}, \qquad \text{with} \qquad \sigma_i = \sqrt{\mathrm{Var}[\tau_i]}, \quad i = 1, 2, \tag{A6.67}$$

is the *correlation coefficient* of the random variables τ_1 and τ_2, provided $\sigma_i < \infty$, $i = 1, 2$. The main properties of the correlation coefficient are:

1. $|\rho| \le 1$.
2. if τ_1 and τ_2 are *independent*, then $\rho = 0$.
3. $\rho = \pm 1$ if and only if τ_1 and τ_2 are linearly dependent.

A6.8.2 Further Properties of Expected Value and Variance

Let $\tau_1, \ldots, \tau_n$ be *arbitrary* random variables (components of a random vector $\vec{\tau}$) having finite variances and $C_1, \ldots, C_n$ constants. From Eqs. (A6.62) or (A6.63) and (A6.40) it follows that

$$\mathrm{E}[C_1\tau_1 + \ldots + C_n\tau_n] = C_1\,\mathrm{E}[\tau_1] + \ldots + C_n\,\mathrm{E}[\tau_n]. \tag{A6.68}$$

If τ_1 and τ_2 are *independent* random variables then, from Eq. (A6.62) or (A6.63),

$$\mathrm{E}[\tau_1\tau_2] = \mathrm{E}[\tau_1]\,\mathrm{E}[\tau_2]. \tag{A6.69}$$

The variance of a sum of *independent* random variables $\tau_1, \ldots, \tau_n$ is obtained from Eqs. (A6.62) or (A6.63) and (A6.69) as

$$\mathrm{Var}[\tau_1 + \ldots + \tau_n] = \mathrm{Var}[\tau_1] + \ldots + \mathrm{Var}[\tau_n]. \tag{A6.70}$$

For a sum of *arbitrary* random variables $\tau_1, \ldots, \tau_n$, the variance can be obtained for $i, j \in \{1, \ldots, n\}$ as

$$\mathrm{Var}[\tau_1 + \ldots + \tau_n] = \mathrm{Var}[\tau_1] + \ldots + \mathrm{Var}[\tau_n] + \sum_{i \neq j} \mathrm{Cov}[\tau_i, \tau_j]. \tag{A6.71}$$

A6.9 Distribution of the Sum of Independent Positive Random Variables and of $\tau_{\min}$, $\tau_{\max}$

Let τ_1 and τ_2 be *independent* non-negative arithmetic random variables with $a_i = \Pr\{\tau_1 = i\}$, $b_i = \Pr\{\tau_2 = i\}$, $i = 0, 1, \ldots$. Obviously, $\tau_1 + \tau_2$ is also arithmetic, and therefore

$$\begin{aligned} c_k = \Pr\{\tau_1 + \tau_2 = k\} &= \Pr\{\bigcup_{i=0}^{k} \{\tau_1 = i \cap \tau_2 = k - i\}\} \\ &= \sum_{i=0}^{k} \Pr\{\tau_1 = i\} \Pr\{\tau_2 = k - i\} = \sum_{i=0}^{k} a_i\, b_{k-i}. \end{aligned} \tag{A6.72}$$

The sequence $c_0, c_1, \ldots$ is the *convolution* of the sequences $a_0, a_1, \ldots$ and $b_0, b_1, \ldots$.

Now, let τ_1 and τ_2 be two *independent* positive continuous random variables with distribution functions $\mathrm{F}_1(t)$, $\mathrm{F}_2(t)$ and densities $\mathrm{f}_1(t)$, $\mathrm{f}_2(t)$, respectively ($\mathrm{F}_1(0) = \mathrm{F}_2(0) = 0$). Using Eq. (A6.55), it can be shown (Example A6.11 and Fig. A6.4) that for the distribution of $\eta = \tau_1 + \tau_2$

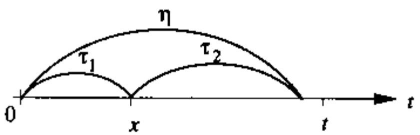


Figure A6.4 To compute the distribution of $\eta = \tau_1 + \tau_2$ $(\tau_1, \tau_2 > 0)$

$$\mathrm{F}_\eta(t) = \Pr\{\eta \le t\} = \int_0^t \mathrm{f}_1(x)\mathrm{F}_2(t - x)\,dx \tag{A6.73}$$

holds, and

$$\mathrm{f}_\eta(t) = \int_0^t \mathrm{f}_1(x)\mathrm{f}_2(t - x)\,dx. \tag{A6.74}$$

The extension to two independent continuous random variables τ_1 and τ_2 defined over $(-\infty, \infty)$ leads to

$$\mathrm{F}_\eta(t) = \int_{-\infty}^{\infty} \mathrm{f}_1(x)\mathrm{F}_2(t - x)\,dx \qquad \text{and} \qquad \mathrm{f}_\eta(t) = \int_{-\infty}^{\infty} \mathrm{f}_1(x)\mathrm{f}_2(t - x)\,dx.$$

The right-hand side of Eq. (A6.74) represents the *convolution* of the densities $\mathrm{f}_1(t)$ and $\mathrm{f}_2(t)$, and will be denoted by

$$\int_0^t \mathrm{f}_1(x)\mathrm{f}_2(t - x)\,dx = \mathrm{f}_1(t) * \mathrm{f}_2(t) \tag{A6.75}$$

The *Laplace transform* (Appendix A9.7) of $\mathrm{f}_\eta(t)$ is thus the product of the Laplace transforms of $\mathrm{f}_1(t)$ and $\mathrm{f}_2(t)$

$$\tilde{\mathrm{f}}_\eta(s) = \tilde{\mathrm{f}}_1(s)\tilde{\mathrm{f}}_2(s). \tag{A6.76}$$

Example A6.11

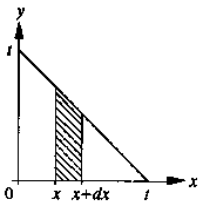
Prove Eq. (A6.74).

Solution

Let τ_1 and τ_2 be two independent positive continuous random variables with distribution functions $\mathrm{F}_1(t)$, $\mathrm{F}_2(t)$ and densities $\mathrm{f}_1(t)$, $\mathrm{f}_2(t)$, respectively. From Eq. (A6.55) one obtains using $\mathrm{f}(x, y) = \mathrm{f}_1(x)\mathrm{f}_2(y)$, see also the graph,

$$F_\eta(t) = \Pr\{\eta = \tau_1 + \tau_2 \le t\} = \iint\limits_{x+y\le t} f_1(x) f_2(y)\,dx\,dy$$

$$= \int_0^t \Big(\int_0^{t-x} f_2(y)\,dy \Big) f_1(x)\,dx = \int_0^t F_2(t-x) f_1(x)\,dx$$



which proves Eq. (A6.73). Eq. (A6.74) follows with $F_2(0) = 0$. A further demonstration of Eq. (A6.74) can be obtained using the formula for total expectation (Eq. (A6.65)).

Sums of positive random variables often occur in reliability theory when investigating repairable systems. For $n \ge 2$, the density $f_\eta(t)$ of $\eta = \tau_1 + \ldots + \tau_n$ for independent positive continuous random variables $\tau_1, \ldots, \tau_n$ follows as

$$f_\eta(t) = f_1(t) * \ldots * f_n(t). \qquad \text{(A6.77)}$$

Example A6.12

Two machines are used to increase the reliability of a system. The first is switched on at time $t = 0$, and the second at the time of failure of the first one, *standby redundancy*. The failure-free operating times of the machines, denoted by τ_1 and τ_2 are independent exponentially distributed with parameter λ (Eq. A6.81)). What is the reliability function of the system?

Solution

From $R_S(t) = \Pr\{\tau_1 + \tau_2 > t\} = 1 - \Pr\{\tau_1 + \tau_2 \le t\}$ and Eq. (A6.73) it follows that

$$R_S(t) = 1 - \int_0^t \lambda e^{-\lambda x} (1 - e^{-\lambda (t-x)})\,dx = e^{-\lambda t} + \lambda t e^{-\lambda t}.$$

$R_S(t)$ gives the probability that no failures ($e^{-\lambda t}$) or exactly one failure ($\lambda t e^{-\lambda t}$) occurs in the interval $(0, t]$, see also Eq. (A7.39) with n=0 and n=1.

Other important distribution functions for reliability analyses are the *minimum* $\tau_{\min}$ and *maximum* $\tau_{\max}$ of a finite set of positive, independent random variables $\tau_1, \ldots, \tau_n$ (for instance as failure-free operating time of a series and of a parallel 1-out-of-n system, respectively). If $\tau_1, \ldots, \tau_n$ are independent positive random variables with distribution functions $F_i(t) = \Pr\{\tau_i \le t\}$, $i = 1, \ldots, n$, then

$$\Pr\{\tau_{\min} > t\} = \Pr\{\tau_1 > t \cap \ldots \cap \tau_n > t\} = \prod_{i=1}^{n} (1 - F_i(t)), \qquad \text{(A6.78)}$$

and

$$\Pr\{\tau_{\max} \le t\} = \Pr\{\tau_1 \le t \cap \ldots \cap \tau_n \le t\} = \prod_{i=1}^{n} F_i(t). \qquad \text{(A6.79)}$$

It can be noted that the failure rate related to $\tau_{\min}$ is given by

$$\lambda_S(t) = \lambda_1(t) + \ldots + \lambda_n(t), \qquad \text{(A6.80)}$$

where $\lambda_i(t)$ is the failure rate related to $F_i(t)$, see Eqs. (A6.78) and (A6.28).

A6.10 Distribution Functions used in Reliability Analyses

This section introduces the most important distribution functions used in quality control and reliability analyses, see Table A6.1 for a summary.

A6.10.1 Exponential Distribution

A continuous positive random variable τ has an *exponential distribution* if

$$F(t) = 1 - e^{-\lambda t}, \qquad t \ge 0, \quad \lambda > 0. \qquad \text{(A6.81)}$$

The *density* is given by

$$f(t) = \lambda e^{-\lambda t}, \qquad t \ge 0, \quad \lambda > 0, \qquad \text{(A6.82)}$$

and the *failure rate* (Eq. (A6.27)) by

$$\lambda(t) = \lambda. \qquad \text{(A6.83)}$$

The *mean* and the *variance* can be obtained from Eqs. (A6.38) and (A6.44) as

$$E[\tau] = \frac{1}{\lambda} \qquad \text{(A6.84)}$$

and

$$\mathrm{Var}[\tau] = \frac{1}{\lambda^2}. \qquad \text{(A6.85)}$$

The *Laplace transform* of $f(t)$ is, according to Table A9.7,

$$\tilde{f}(s) = \frac{\lambda}{s + \lambda}. \qquad \text{(A6.86)}$$

Example A6.13

The failure-free operating time τ of an assembly is exponentially distributed with $\lambda = 10^{-5}\,h^{-1}$. What is the probability of τ being (i) over 2,000 h, (ii) over 20,000 h, (iii) over 100,000 h, (iv) between 20,000 h and 100,000 h?

Solution

From Eqs. (A6.81), (A6.26) and (A6.20) one obtains

(i) $\Pr\{\tau > 2{,}000\,h\} = e^{-0.02} \approx 0.98$,
(ii) $\Pr\{\tau > 20{,}000\,h\} = e^{-0.2} \approx 0.819$,
(iii) $\Pr\{\tau > 100{,}000\,h\} = \Pr\{\tau > 1/\lambda = E[\tau]\} = e^{-1} \approx 0.368$,
(iv) $\Pr\{20{,}000\,h < \tau \le 100{,}000\,h\} = e^{-0.2} - e^{-1} \approx 0.451$.

For an exponential distribution, the failure rate is *constant* (time-independent) and equal to λ. This important property is a *characteristic* of the exponential distribution and does not appear with any other continuous distribution. It greatly simplifies calculation because of the following properties:

1. *Memoryless property*: Assuming that the failure-free operating time is exponentially distributed and knowing that the item is functioning at the present time, its behavior in the future *will not depend on how long it has already been operating*. In particular, the probability that it will fail in the next time interval δt is *constant* and equal to $\lambda \delta t$. This is a consequence of Eq. (A6.29)

$$\Pr\{\tau > t + x_0 \mid \tau > x_0\} = e^{-\lambda t}. \qquad \text{(A6.87)}$$

2. *Constant failure rate at system level*: If a system *without redundancy* consists of the elements $E_1, \ldots, E_n$ and the failure-free operating times $\tau_1, \ldots, \tau_n$ of these elements are *independent* and *exponentially distributed* with parameters $\lambda_1, \ldots, \lambda_n$ then, according to Eq. (A6.78), the system failure rate is also *constant* (time-independent) and equal to the sum of the failure rates of its elements

$$R_S(t) = e^{-\lambda_1 t} \ldots e^{-\lambda_n t} = e^{-\lambda_S t}, \qquad \text{with } \lambda_S = \lambda_1 + \ldots + \lambda_n. \qquad \text{(A6.88)}$$

It should be noted that the expression $\lambda_S = \sum \lambda_i$ is a characteristic of the *series model* with independent elements, and also remains true for the time-dependent failure rates $\lambda_i = \lambda_i(t)$, see Eqs. (A6.80) and (2.16).

A6.10.2 Weibull Distribution

The *Weibull distribution* can be considered as a generalization of the exponential distribution. A continuous positive random variable τ has a Weibull distribution if

$$F(t) = 1 - e^{-(\lambda t)^\beta}, \qquad t \geq 0, \quad \lambda, \beta > 0. \qquad \text{(A6.89)}$$

The *density* is given by

$$f(t) = \lambda \beta (\lambda t)^{\beta - 1} e^{-(\lambda t)^\beta}, \qquad t \geq 0, \quad \lambda, \beta > 0, \qquad \text{(A6.90)}$$

and the *failure rate* (Eq. (A6.27)) by

$$\lambda(t) = \beta \lambda (\lambda t)^{\beta - 1}. \qquad \text{(A6.91)}$$

λ is the *scale parameter* (F(t) depends on λt only) and β the *shape parameter*. $\beta = 1$ yields the exponential distribution. For $\beta > 1$, the failure rate $\lambda(t)$ *increases monotonically*, with $\lambda(0) = 0$ and $\lambda(\infty) = \infty$. For $\beta < 1$, $\lambda(t)$ *decreases monotonically*, with $\lambda(0) = \infty$ and $\lambda(\infty) = 0$. The *mean* and the *variance* are given by

$$E[\tau] = \frac{\Gamma(1 + 1/\beta)}{\lambda} \qquad \text{(A6.92)}$$

and

$$\mathrm{Var}[\tau] = \frac{\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)}{\lambda^2}, \qquad \text{(A6.93)}$$

where

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} \, dx \qquad \text{(A6.94)}$$

is the *complete gamma function* (Appendix A9.6). The coefficient of variation $\kappa = \sqrt{\mathrm{Var}[\tau]} / E[\tau] = \sigma / E[\tau]$ is plotted in Fig.4.5. For a given $E[\tau]$, the density of the Weibull distribution becomes peaked with increasing β. An analytical expression for the *Laplace transform* of the Weibull distribution function does not exist.

For a system without redundancy (*series model*) whose elements have independent failure-free operating times $\tau_1, \ldots, \tau_n$ distributed according to Eq. (A6.89), the reliability function is given by

$$R_S(t) = (e^{-(\lambda t)^\beta})^n = e^{-(\lambda' t)^\beta}, \qquad \text{(A6.95)}$$

with $\lambda' = \lambda \sqrt[\beta]{n}$. Thus, the failure-free operating time of the system has a Weibull distribution with parameters λ' and β.

The Weibull distribution with $\beta > 1$ often occurs in applications as a distribution of the failure-free operating times of components which are subject to *wearout* and/or *fatigue* (lamps, relays, mechanical components, etc.). It was introduced by W. Weibull in 1951, related to investigations on fatigue in metals [A6.20]. B.W. Gnedenko showed that a Weibull distribution occurs as one of the *extreme value distributions* for the smallest of n ($n \to \infty$) independent random variables with the same distribution function (Weibull-Gnedenko distribution [A6.7, A6.8]).

The Weibull distribution is often given with the parameter $\alpha = \lambda^\beta$ instead of λ or also with *three parameters*

$$F(t) = 1 - e^{-(\lambda (t - \psi))^\beta}, \qquad t \geq \psi, \quad \psi, \lambda, \beta > 0. \qquad \text{(A6.96)}$$

Example A6.14

Shows that for a three parameter Weibull distribution the time scale parameter ψ too can be determined (graphically) on a Weibull probability chart, e.g. for the *empirical evaluation of data*.

Solution

In the system of coordinates $\log_{10}(t)$ and $\log_{10}\log_{10}(1/(1 - F(t)))$ the two parameter Weibull distribution function (Eq. (A6.89)) appears as a straight line, allowing a graphical determination of λ and β (see Eq.(A8.16) and Fig.A8.2). The three parameter Weibull distribution (Eq.(A6.96)) leads to a concave curve. In this case, for two arbitrary points t_1 and $t_2 > t_1$ it holds for the mean point on the scale $\log_{10}\log_{10}(1/(1 - F(t)))$, defining t_m, that $\log_{10}(t_2 - \psi) + \log_{10}(t_1 - \psi) = 2\log_{10}(t_m - \psi)$, see Eq. (A8.16), the identity $a + (b - a)/2 = (a + b)/2$, and Fig. A8.2. From this, $(t_2 - \psi)(t_1 - \psi) = (t_m - \psi)^2$ and $\psi = (t_1 t_2 - t_m^2)/(t_1 + t_2 - 2t_m)$, as function of t_1, t_2, t_m.

A6.10.3 Gamma Distribution, Erlangian Distribution, and χ^2 Distribution

A continuous positive random variable τ has a *Gamma distribution* if

$$F(t) = \Pr\{\tau \le t\} = \frac{1}{\Gamma(\beta)} \int_0^{\lambda t} x^{\beta-1} e^{-x} dx, \qquad t \ge 0, \quad \lambda, \beta > 0. \qquad (A6.97)$$

Γ is the *complete Gamma function* defined by Eq. (A6.94). The right-hand side of Eq. (A6.97) is the *incomplete Gamma function* with argument λt (Appendix A9.6). The *density* of the Gamma distribution is given by

$$f(t) = \lambda \frac{(\lambda t)^{\beta-1}}{\Gamma(\beta)} e^{-\lambda t}, \qquad t \ge 0, \quad \lambda, \beta > 0 \qquad (A6.98)$$

and the *failure rate* is calculated from $\lambda(t) = f(t)/(1 - F(t))$. $\lambda(t)$ is *constant* (time-independent) for $\beta = 1$, *monotonically decreasing* for $\beta < 1$ and *monotonically increasing* for $\beta > 1$. However, in contrast to the Weibull distribution, $\lambda(t)$ always converges to λ for $t \to \infty$, see Table A6.1 for an example. A Gamma distribution with $\beta < 1$ mixed with a three-parameter Weibull distribution (Eq. (A6.34)) can be used as an approximation to the distribution function for an item with failure rate as the *bathtub curve* given in Fig. 1.2.

The *mean* and the *variance* of the Gamma distribution are given by

$$E[\tau] = \frac{\beta}{\lambda} \qquad (A6.99)$$

and

$$\mathrm{Var}[\tau] = \frac{\beta}{\lambda^2}. \qquad (A6.100)$$

The *Laplace transform* (Table A9.7) of the Gamma distribution density is

$$\tilde{f}(s) = \frac{\lambda^\beta}{(s+\lambda)^\beta}. \qquad (A6.101)$$

From Eqs. (A6.101) and (A6.76), it follows that the sum of two *independent* Gamma-distributed random variables with parameters λ, β_1 and λ, β_2 has a Gamma distribution with parameters λ, $\beta_1 + \beta_2$.

Example A6.15

Let the random variables τ_1 and τ_2 be independent and distributed according to a Gamma distribution with the parameters λ and β. Determine the density of the sum $\eta = \tau_1 + \tau_2$.

Solution

According Eq. (A6.98), τ_1 and τ_2 have the density $f(t) = \lambda(\lambda t)^{\beta-1} e^{-\lambda t} / \Gamma(\beta)$. The Laplace transform of $f(t)$ is $\tilde{f}(s) = \lambda^\beta / (s+\lambda)^\beta$ (Table A9.7). From Eq. (A6.76), the Laplace transform of the density of $\eta = \tau_1 + \tau_2$ follows as $\tilde{f}_\eta(s) = \lambda^{2\beta} / (s+\lambda)^{2\beta}$. The random variable $\eta = \tau_1 + \tau_2$ thus has a Gamma distribution with parameters λ and 2β. The generalization to the sum of $\tau_1 + \ldots + \tau_n$ leads to a Gamma distribution with parameters λ and $n\beta$.

For $\beta = n = 2, 3, \ldots$, the Gamma distribution given by Eq. (A6.97) leads to an *Erlangian distribution* with parameters λ and n. Taking into account Eq. (A6.77) and comparing the Laplace transform of the exponential distribution $\lambda/(s+\lambda)$ with that of the Erlangian distribution $(\lambda/(s+\lambda))^n$, leads to the following conclusion:

> *If τ is Erlang distributed with parameters λ and n, then τ can be considered as the sum of n independent, exponentially distributed random variables with parameter λ, i.e. $\tau = \tau_1 + \ldots + \tau_n$ with $\Pr\{\tau_i \le t\} = 1 - e^{-\lambda t}$, $i = 1, \ldots, n$.*

The distribution function $F(t)$ of the *Erlangian distribution* is obtained by partial integration of the right-hand side of Eq. (A6.97), with $\beta = n$. This leads to

$$F(t) = \Pr\{\tau_1 + \ldots + \tau_n \le t\} = 1 - \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t}, \qquad t \ge 0, \quad \lambda > 0. \quad (A6.102)$$

When $\lambda = 1/2$ and $\beta = \nu/2$, $\nu = 1, 2, \ldots$, then the Gamma distribution given by Eq. (A6.97) is a *chi-square distribution* (χ^2 distribution) with ν degrees of freedom. The corresponding random variable is denoted χ^2_ν. The chi-square distribution with ν degrees of freedom is thus given by

$$F(t) = \Pr\{\chi^2_\nu \le t\} = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} \int_0^t x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} dx, \qquad t \ge 0, \ \nu = 1, 2, \ldots. \quad (A6.103)$$

From Eqs. (A6.97), (A6.102), and (A6.103) it follows that

$$2\lambda(\tau_1 + \ldots + \tau_n) \qquad (A6.104)$$

has a χ^2 distribution with $\nu = 2n$ degrees of freedom. If $\xi_1, \ldots, \xi_n$ are *independent, normally distributed* random variables with mean m and variance σ^2, then

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (\xi_i - m)^2$$

is χ^2 distributed with n degrees of freedom. The above considerations show the importance of the χ^2 distribution in mathematical statistics. The χ^2 distribution is also used to compute the *Poisson distribution* (Eq.(A6.102) with $n = \nu/2$ and $\lambda = 1/2$ or Eq. (A6.126) with $k = \nu/2 - 1$ and $m = t/2$, see also Table A9.2).

A6.10.4 Normal Distribution

A commonly used distribution function, in theory and practice, is the *normal distribution*, or Gaussian distribution. The random variable τ has a normal distribution if

$$F(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{(y-m)^2}{2\sigma^2}}dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{t-m}{\sigma}} e^{-\frac{x^2}{2}}dx, \qquad -\infty<t,m<\infty,\ \sigma>0. \tag{A6.105}$$

The *density* of the normal distribution is given by

$$f(t)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-m)^2}{2\sigma^2}}, \qquad -\infty<t,\ m<\infty, \quad \sigma>0. \tag{A6.106}$$

The failure rate is calculated from $\lambda(t)=f(t)/(1-F(t))$. The *mean* and *variance* are

$$E[\tau]=m \tag{A6.107}$$

and

$$Var[\tau]=\sigma^2, \tag{A6.108}$$

respectively. The density of the normal distribution is *symmetric* with respect to the line $x=m$. Its width depends upon the variance. The area under the density curve is equal to (Table A9.1)

- 0.683 for the interval $m\pm\sigma$
- 0.954 for the interval $m\pm 2\sigma$
- 0.997 for the interval $m\pm 3\sigma$.

Obviously, a normally distributed random variable assumes values from $-\infty$ to $+\infty$. However, for $m>3\sigma$ it can be considered as a positive random variable in practical applications. If τ is normally distributed with parameters m and σ^2, then $(\tau-m)/\sigma$ is normally distributed with parameters 0 and 1, which is the *standard normal distribution*, often represented by $\Phi(t)$

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}dx. \tag{A6.109}$$

If τ_1 and τ_2 are independent, normally distributed random variables with parameters m_1 and σ_1^2, m_2 and σ_2^2, respectively, then $\eta=\tau_1+\tau_2$ is normally distributed with parameters m_1+m_2 and $\sigma_1^2+\sigma_2^2$ (Example A6.16). This rule can be generalized to the sum of n independent normally distributed random variables, a result which can be extended to dependent normally distributed random variables (Example A6.16).

Example A6.16

Let the random variables τ_1 and τ_2 be statistically independent and normally distributed with means m_1 and m_2 and variances σ_1^2 and σ_2^2. Determine the density of the sum $\eta=\tau_1+\tau_2$.

Solution

According to Eq. (A6.74), the density of $\eta=\tau_1+\tau_2$ follows as

$$f_\eta(t)=\frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{-\left(\frac{(x-m_1)^2}{2\sigma_1^2}+\frac{(t-x-m_2)^2}{2\sigma_2^2}\right)}dx.$$

With $u=x-m_1$, $v=t-m_1-m_2$, and taking into consideration

$$\frac{u^2}{\sigma_1^2}+\frac{(v-u)^2}{\sigma_2^2}=\left[\frac{u\sqrt{\sigma_1^2+\sigma_2^2}}{\sigma_1\sigma_2}-\frac{v\sigma_1}{\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}\right]^2+\frac{v^2}{\sigma_1^2+\sigma_2^2},$$

the result

$$f_\eta(t)=\frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}e^{-\frac{(t-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}}$$

is obtained. Thus the sum of independent normally distributed random variables is also normally distributed with mean m_1+m_2 and variance $\sigma_1^2+\sigma_2^2$. If τ_1 and τ_2 are statistically dependent, then the distribution function of $\tau_1+\tau_2$ is still a normal distribution with $m=m_1+m_2$, but with variance $\sigma^2=\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2$, where ρ is the correlation coefficient as defined in Eq. (A6.67).

The normal distribution often occurs in practical applications because the distribution function of the sum of a large number of statistically independent random variables converges to a normal distribution under relatively general conditions (central limit theorem, Eq. (A6.148)).

A6.10.5 Lognormal Distribution

A continuous positive random variable τ has a *lognormal distribution* if its logarithm is normally distributed (Example A6.17). For the lognormal distribution,

$$F(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_{0}^{t}\frac{1}{y}e^{-\frac{(\ln(\lambda y))^2}{2\sigma^2}}dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\ln(\lambda t)}{\sigma}} e^{-\frac{x^2}{2}}dx=\Phi(\ln(\lambda t)/\sigma), \qquad t\ge 0,\ \lambda,\sigma>0. \tag{A6.110}$$

The *density* is given by

$$\mathrm{f}(t)=\frac{1}{t\,\sigma\sqrt{2\pi}}\,e^{-\frac{(\ln\lambda t)^2}{2\sigma^2}},\qquad t\geq 0,\quad \lambda,\sigma>0. \tag{A6.111}$$

The *failure rate* is calculated from $\lambda(t)=\mathrm{f}(t)/(1-\mathrm{F}(t))$, see Table A6.1 for an example. The *mean* and the *variance* of τ are

$$\mathrm{E}[\tau]=\frac{e^{\sigma^2/2}}{\lambda} \tag{A6.112}$$

and

$$\mathrm{Var}[\tau]=\frac{e^{2\sigma^2}-e^{\sigma^2}}{\lambda^2}, \tag{A6.113}$$

respectively. The density of the lognormal distribution has the important property that it is practically zero at the origin, increases rapidly to a maximum, and then decreases relatively quickly (Fig. 4.2). The lognormal distribution function is therefore suitable for modeling *repair times* (Section 4.1). It is also often used as a distribution function for the failure-free operating time of components in *accelerated reliability testing* (Section 7.4) as well as in cases where a large number of statistically independent random variables are combined together in a *multiplicative* fashion (*additive* for $\eta=\ln\tau$, i.e. for the normal distribution).

Example A6.17

Show that the logarithm of a lognormally distributed random variable $\eta=\ln\tau$ is normally distributed.

Solution

For

$$\mathrm{f}_\tau(t)=\frac{1}{t\,\sigma\sqrt{2\pi}}\,e^{-\frac{(\ln t+\ln\lambda)^2}{2\sigma^2}}$$

Equation (A6.33), with $\mathrm{u}(t)=\ln t$ and $\mathrm{u}^{-1}(t)=e^t$, yields

$$\mathrm{f}_\eta(t)=\frac{1}{e^t\sigma\sqrt{2\pi}}\,e^{-\frac{(t+\ln\lambda)^2}{2\sigma^2}}\,e^t=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(t+\ln\lambda)^2}{2\sigma^2}}=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(t-m)^2}{2\sigma^2}},$$

with $\ln(1/\lambda)=m$. This method can be used for other *transformations*, for example:

$\mathrm{u}(t)=e^t$: Normal distribution → lognormal distribution
$\mathrm{u}(t)=t^\beta$: Weibull distribution → Exponential distribution
$\mathrm{u}(t)=\sqrt[\beta]{t}$: Exponential distribution → Weibull distribution
$\mathrm{u}(t)=\mathrm{F}_\eta^{-1}(t)$: Uniform distribution (Eq. (A6.114)) in the interval [0, 1] → $\mathrm{F}_\eta(t)$.

In *Monte Carlo simulations*, algorithms more sophisticated than $\mathrm{F}_\eta^{-1}(t)$ are often used.

A6.10.6 Uniform Distribution

A random variable τ is *uniformly* (or rectangularly) *distributed* in the interval $[a, b]$ if it has the distribution function

$$\mathrm{F}(t)=\Pr\{\tau\leq t\}=\begin{cases}0 & \text{if } t\leq a\\ \dfrac{t-a}{b-a} & \text{if } a<t\leq b\\ 1 & \text{otherwise.}\end{cases} \tag{A6.114}$$

The *density* is then given by

$$\mathrm{f}(t)=\frac{1}{b-a}\qquad\text{for}\quad a\leq t\leq b.$$

The uniform distribution is a particular case of the geometric probabilities introduced by Eq. (A6.3), but for $\mathcal{R}^1$ instead of $\mathcal{R}^2$. Because of the property mentioned at the end of Example A6.17, the *uniform distribution* in the interval [0, 1] plays an important role in simulation problems.

A6.10.7 Binomial Distribution

Consider a trial in which the only outcome is either a given event A or its complement $\overline{A}$. This outcome can be represented by a random variable of the form

$$\delta=\begin{cases}1 & \text{if } A \text{ occurs}\\ 0 & \text{otherwise.}\end{cases} \tag{A6.115}$$

δ is called a *Bernoulli variable.* If

$$\Pr\{\delta=1\}=p\qquad\text{and}\qquad\Pr\{\delta=0\}=1-p, \tag{A6.116}$$

then

$$\mathrm{E}[\delta]=1\cdot p+0\cdot(1-p)=p, \tag{A6.117}$$

and

$$\mathrm{Var}[\delta]=\mathrm{E}[\delta^2]-\mathrm{E}^2[\delta]=p-p^2=p(1-p). \tag{A6.118}$$

An infinite sequence of independent Bernoulli variables

$$\delta_1,\delta_2,\ldots$$

with the same probability $\Pr\{\delta_i=1\}=p$, $i\geq 1$, is called a *Bernoulli model* or a *sequence of Bernoulli trials*. The sequence $\delta_1,\delta_2,\ldots$ describes, for example, the model of the repeated sampling of a component from a lot of size N, with K defective components ($p=K/N$) such that the component is returned to the lot after testing (sample with replacement). The random variable

$$\zeta = \delta_1 + \ldots + \delta_n \tag{A6.119}$$

is the number of ones occurring in n Bernoulli trials. The distribution of ζ is given by

$$p_k = \Pr\{\zeta = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, \ldots, n, \quad 0 < p < 1. \tag{A6.120}$$

Equation (A6.120) is the *binomial distribution*. ζ is obviously an *arithmetic random variable* taking on values in $\{0, 1, \ldots, n\}$ with probabilities p_k. To prove Eq. (A6.120), consider that

$$p^k (1-p)^{n-k} = \Pr\{\delta_1 = 1 \cap \ldots \cap \delta_k = 1 \cap \delta_{k+1} = 0 \cap \ldots \cap \delta_n = 0\}$$

is the probability of the event A occurring in the first k trials and not occurring in the $n-k$ following trials ($\delta_1, \ldots, \delta_n$ are independent); furthermore in n trials there are

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k!}$$

different possibilities of occurrence of k ones and $n-k$ zeros, the addition theorem (Eq. (A6.11)) then leads to Eq. (A6.120).

Example A6.18

A populated printed circuit board (PCB) contains 30 ICs. These are taken from a shipment in which the probability of each IC being defective is constant and equal to 1%. What are the probabilities that the PCB contains (i) no defective ICs, (ii) exactly one defective IC, and (iii) more than one defective IC?

Solution

From Eq. (A6.120) with $p = 0.01$,

(i) $p_0 = 0.99^{30} \approx 0.74$,
(ii) $p_1 = 30 \cdot 0.01 \cdot 0.99^{29} \approx 0.224$,
(iii) $p_2 + \ldots + p_{30} = 1 - p_0 - p_1 \approx 0.036$.

Knowing p_i and assuming C_i = cost for i repairs (because of i defective ICs) it is easy to calculate the *mean* C of the total cost caused by the defective ICs ($C = p_1 C_1 + \ldots + p_{30} C_{30}$) and thus to develop a *test strategy* based on cost considerations (Section 8.5).

For the random variable ζ defined by Eq. (A6.119) it follows that

$$\Pr\{\zeta \le k\} = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}, \qquad k = 0, \ldots, n, \quad 0 < p < 1, \tag{A6.121}$$

$$\mathrm{E}[\zeta] = n\,p, \tag{A6.122}$$

$$\mathrm{Var}[\zeta] = n\,p\,(1-p). \tag{A6.123}$$

Example A6.19

Determine the mean and the variance of a binomially distributed random variable with parameters n and p.

Solution

Considering the independence of $\delta_1, \ldots, \delta_n$, the definition of ζ (Eq. (A6.119)), and from Eqs. (A6.117) and (A6.118) it follows that

$$\mathrm{E}[\zeta] = \mathrm{E}[\delta_1] + \ldots + \mathrm{E}[\delta_n] = n\,p$$

and

$$\mathrm{Var}[\zeta] = \mathrm{Var}[\delta_1] + \ldots + \mathrm{Var}[\delta_n] = n\,p\,(1-p).$$

A further demonstration follows, as for Example A6.20, by considering that

$$\sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = n\,p \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = n\,p \sum_{i=0}^{m} \binom{m}{i} p^i (1-p)^{m-i} = n\,p.$$

For large n, the binomial distribution converges to the normal distribution (Eq. (A6.149)). The convergence is good for $\min(n\,p,\, n(1-p)) \ge 5$. For small values of p, the *Poisson approximation* (Eq. (A6.129)) can be used. Calculations of Eq. (A6.120) can be based upon the relationship between the binomial and the *beta* or the *Fisher distribution* (Appendix A9.4).

Generalization of Eq. (A6.120) for the case where one of the events $A_1, \ldots, A_m$ can occur with probability $p_1, \ldots, p_m$ at every trial, leads to the *multinomial distribution*

$$\Pr\{\text{in } n \text{ trials } A_1 \text{ occurs } k_1 \text{ times}, \ldots, A_m \text{ occurs } k_m \text{ times}\} = \frac{n!}{k_1! \ldots k_m!}\, p_1^{k_1} \ldots p_m^{k_m}, \tag{A6.124}$$

with $k_1 + \ldots + k_m = n$ and $p_1 + \ldots + p_m = 1$.

A6.10.8 Poisson Distribution

The arithmetic random variable ζ has a *Poisson distribution* if

$$p_k = \Pr\{\zeta = k\} = \frac{m^k}{k!} e^{-m}, \qquad k = 0, 1, \ldots, \quad m > 0 \tag{A6.125}$$

and thus

$$\Pr\{\zeta \le k\} = \sum_{i=0}^{k} \frac{m^i}{i!} e^{-m}, \qquad k = 0, 1, \ldots, \quad m > 0. \tag{A6.126}$$

The *mean* and the *variance* of ζ are

$$E[\zeta] = m \qquad (A6.127)$$

and

$$\mathrm{Var}[\zeta] = m. \qquad (A6.128)$$

The Poisson distribution often occurs in connection with *exponentially distributed failure-free operating times*, since Eq. (A6.125) with $m = \lambda t$ gives the probability of exactly k failures in the time interval $(0, t]$, see Eq. (A7.39).

The Poisson distribution is also used as an *approximation* of the binomial distribution for $n \to \infty$ and $p \to 0$ such that $n p = m < \infty$. To prove this convergence, called the *Poisson approximation*, set $m = n p$, Eq. (A6.120) then yields

$$p_k = \frac{n!}{k!(n-k)!}\left(\frac{m}{n}\right)^k\left(1-\frac{m}{n}\right)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{n^k}\cdot\frac{m^k}{k!}\left(1-\frac{m}{n}\right)^{n-k}$$

$$= 1\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\cdot\frac{m^k}{k!}\left(1-\frac{m}{n}\right)^{n-k},$$

from which (for $k < \infty$ and $m = n p < \infty$) it follows that

$$\lim_{n\to\infty} p_k = \frac{m^k}{k!}e^{-m}, \qquad m = n p. \qquad (A6.129)$$

Using partial integration one can show that

$$\sum_{i=0}^{k}\frac{m^i}{i!}e^{-m} = 1 - \frac{1}{k!}\int_0^m y^k e^{-y}dy = 1 - \frac{1}{k!\,2^{k+1}}\int_0^{2m} x^k e^{-\frac{x}{2}}dx. \qquad (A6.130)$$

The right-hand side of Eq. (A6.130) is a special case of the *chi-square distribution* (Eq. (A6.103) with $\nu/2 = k+1$ and $t = 2m$). A table of the chi-square distribution can then be used for numerical evaluation of the Poisson distribution (Table A9.2).

Example A6.20

Determine the mean and the variance of a Poisson-distributed random variable.

Solution

From Eqs. (A6.35) and (A6.125),

$$E[\zeta] = \sum_{k=0}^{\infty} k\frac{m^k}{k!}e^{-m} = \sum_{k=1}^{\infty} m\frac{m^{k-1}}{(k-1)!}e^{-m} = m\sum_{i=0}^{\infty}\frac{m^i}{i!}e^{-m} = m.$$

Similarly, from Eqs. (A6.45), (A6.41), and (A6.125),

$$\mathrm{Var}[\zeta] = \sum_{k=0}^{\infty} k^2\frac{m^k}{k!}e^{-m} - m^2 = \sum_{k=0}^{\infty}[k(k-1)+k]\frac{m^k}{k!}e^{-m} - m^2$$

$$= \sum_{k=2}^{\infty} m^2\frac{m^{k-2}}{(k-2)!}e^{-m} + m - m^2 = m^2\sum_{i=0}^{\infty}\frac{m^i}{i!}e^{-m} + m - m^2 = m.$$

A6.10.9 Geometric Distribution

Let $\delta_1, \delta_2, \dots$ be a sequence of independent Bernoulli variables resulting from Bernoulli trials. The arithmetic random variable ζ defining the number of trials to the *first occurrence* of the event A has a *geometric distribution* given by

$$p_k = \Pr\{\zeta = k\} = p(1-p)^{k-1}, \qquad k = 1, 2, \dots, \qquad 0 < p < 1. \qquad (A6.131)$$

Equation (A6.131) follows from the definition of Bernoulli variables δ_i (Eq. (A6.115))

$$p_k = \Pr\{\zeta = k\} = \Pr\{\delta_1 = 0 \cap \dots \cap \delta_{k-1} = 0 \cap \delta_k = 1\} = (1-p)^{k-1}p.$$

The geometric distribution is the only discrete distribution which exhibits the *memoryless property*, as does the exponential distribution for the continuous case. In fact, from $\Pr\{\zeta > k\} = \Pr\{\delta_1 = 0 \cap \dots \cap \delta_k = 0\} = (1-p)^k$ and, for any k and $j > 0$, it follows that

$$\Pr\{\zeta > k + j \mid \zeta > k\} = \frac{(1-p)^{k+j}}{(1-p)^k} = (1-p)^j.$$

The *failure rate* (Eq. (A6.31)) is time independent and given by

$$\lambda(k) = \frac{p(1-p)^{k-1}}{(1-p)^{k-1}} = p. \qquad (A6.132)$$

For the distribution function of the random variable ζ defined by Eq. (A6.131) one obtains

$$\Pr\{\zeta \le k\} = \sum_{i=1}^{k} p_i = 1 - \Pr\{\zeta > k\} = 1 - (1-p)^k. \qquad (A6.133)$$

Mean and variance are then (with $\sum_{n=1}^{\infty} n x^n = x/(1-x)^2$ and $\sum_{n=1}^{\infty} n^2 x^n = x(1+x)/(1-x)^3, x < 1$)

$$E[\zeta] = \frac{1}{p} \qquad (A6.134)$$

and

$$\mathrm{Var}[\zeta] = \frac{1-p}{p^2}. \qquad (A6.135)$$

If Bernoulli trials are carried out at regular intervals Δt, then Eq. (A6.133) provides the distribution function of the *number of time units Δt between successive occurrences of the event A under consideration*; for example, breakdown of a capacitor, interference pulse in a digital network, etc.

Often the geometric distribution is considered with $p_k = p(1-p)^k$, $k = 0, 1, \dots$, in this case $E[\zeta] = (1-p)/p$ and $\mathrm{Var}[\zeta] = (1-p)/p^2$.

A6.10.10 Hypergeometric Distribution

The *hypergeometric distribution* describes the model of a *random sample without replacement*. For example, if it is known that there are exactly K defective components in a lot of size N, then the probability of finding k defective components in a random sample of size n is given by

$$p_k = \Pr\{\zeta = k\} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \qquad k = 0, \ldots, \min(K,n). \tag{A6.136}$$

Equation (A6.136) defines the *hypergeometric distribution*. Since for fixed n and k $(0 \le k \le n)$

$$\lim_{N\to\infty} \Pr\{\zeta = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \text{with } p = \frac{K}{N},$$

the hypergeometric distribution can, for large N, be approximated by the binomial distribution with $p = K/N$. For the random variable ζ defined by Eq. (A6.136) it follows that

$$\Pr\{\zeta \le k\} = \sum_{i=0}^{k} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}, \qquad k = 0, \ldots, n, \quad 0 < p < 1, \tag{A6.137}$$

$$E[\zeta] = n\frac{K}{N}, \tag{A6.138}$$

and

$$\mathrm{Var}[\zeta] = \frac{K\,n(N-K)(N-n)}{N^2(N-1)}. \tag{A6.139}$$

A6.11 Limit Theorems

Limit theorems are of great importance in practical applications because they can be used to find *approximations* with the help of known (tabulated) distributions. Two important cases will be discussed in this section, the *law of large numbers* and the *central limit theorem*. The law of large numbers provides additional justification for the construction of probability theory on the basis of relative frequencies. The central limit theorem shows that the normal distribution can be used as an approximation in many practical situations.

A6.11.1 Law of Large Numbers

Two notions used with the law of large numbers are *convergence in probability* and *convergence with probability one*. Let $\zeta_1, \zeta_2, \ldots$, and ζ be random variables on a probability space $[\Omega, \mathcal{F}, \Pr]$. ζ_n *converge in probability* to ζ if for arbitrary $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\{\,|\,\xi_n - \xi\,| > \varepsilon\} = 0 \tag{A6.140}$$

holds. ζ_n *converge* to ζ *with probability one* if

$$\Pr\{\lim_{n\to\infty} \xi_n = \xi\} = 1. \tag{A6.141}$$

The convergence with probability one is also called *convergence almost sure* (a.s.). An equivalent condition for Eq. (A6.141) is

$$\lim_{n\to\infty} \Pr\{\sup_{k\ge n} |\,\zeta_k - \zeta\,| > \varepsilon\} = 0, \tag{A6.142}$$

for any $\varepsilon > 0$. This clarifies the difference between Eq. (A6.140) and the stronger condition given by Eq. (A6.141).

Let us now consider an infinite sequence of *Bernoulli trials* (Eqs. (A6.115), (A6.119), and (A6.120)), with parameter $p = \Pr\{A\}$, and let S_n be the number of occurrences of the event A in n trials

$$S_n = \delta_1 + \ldots + \delta_n. \tag{A6.143}$$

The quantity S_n/n is the *relative frequency* of the occurrence of A in n independent trials. The *weak law of large numbers* states that for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\{\,|\,\frac{S_n}{n} - p\,| > \varepsilon\} = 0. \tag{A6.144}$$

Equation (A6.144) is a direct consequence of Chebyshev's inequality (Eq. (A6.49)).. Similarly, for a sequence of *independent identically distributed* random variables $\tau_1, \ldots, \tau_n$, with mean $E[\tau_i] = a$ and variance $\mathrm{Var}[\tau_i] = \sigma^2 < \infty$ $(i = 1, \ldots, n)$,

$$\lim_{n\to\infty} \Pr\{\,|\,(\frac{1}{n}\sum_{i=1}^{n}\tau_i) - a\,| > \varepsilon\} = 0. \tag{A6.145}$$

According to Eq. (A6.144), the sequence S_n/n *converges in probability* to $p = \Pr\{A\}$. Moreover, according to the Eq. (A6.145), the *arithmetic mean*

$(t_1 + \ldots + t_n)/n$ of *n independent observations* of the random variable τ (with a finite variance) *converges in probability* to $E[\tau]$. Therefore, $\hat{p} = S_n/n$ and $\hat{a} = (t_1 + \ldots + t_n)/n$ are *consistent estimates* of $p = \Pr\{A\}$ and $a = E[\tau]$, respectively (Appendix A8.1 and A8.2). Equation (A6.145) is also a direct consequence of Chebyshev's inequality.

A firmer statement than the weak law of large numbers is given by the *strong law of large numbers,*

$$\Pr\{\lim_{n\to\infty} \frac{S_n}{n} = p\} = 1. \tag{A6.146}$$

According to Eq. (A6.146), the relative frequency S_n/n *converges with probability one* (a.s.) to $p = \Pr\{A\}$. Similarly, for a sequence of independent identically distributed random variables $\tau_1, \ldots, \tau_n$, with mean $E[\tau_i] = a$ and variance $\mathrm{Var}[\tau_i] = \sigma^2 < \infty$ $(i = 1, 2, \ldots)$,

$$\Pr\{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \tau_i = a\} = 1. \tag{A6.147}$$

The proof of the strong law of large numbers (A6.146) and (A6.147) is more laborious than that of the weak law of large numbers, see e.g. [A6.6 (vol. II), A6.7, A6.12].

A6.11.2 Central Limit Theorem

Let $\tau_1, \tau_2, \ldots$ be independent, identically distributed random variables with mean $E[\tau_i] = a$ and variance $\mathrm{Var}[\tau_i] = \sigma^2 < \infty$, $i = 1, 2, \ldots$. For every $t < \infty$,

$$\lim_{n\to\infty} \Pr\{\frac{\sum_{i=1}^{n} \tau_i - na}{\sigma\sqrt{n}} \le t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx. \tag{A6.148}$$

Equation (A6.148) is the *central limit theorem.* It says that for large values of n, the distribution function of the sum $\tau_1 + \ldots + \tau_n$ can be approximated by the normal distribution with mean $E[\tau_1 + \ldots + \tau_n] = nE[\tau_i] = na$ and variance $\mathrm{Var}[\tau_1 + \ldots + \tau_n] = n\mathrm{Var}[\tau_i] = n\sigma^2$. The central limit theorem is of great theoretical and practical importance, in probability theory and mathematical statistics. It includes the *integral Laplace theorem* (also known as the *De Moivre-Laplace theorem*) for the case where $\tau_i = \delta_i$ are Bernoulli variables,

$$\lim_{n\to\infty} \Pr\{\frac{\sum_{i=1}^{n} \delta_i - np}{\sqrt{np(1-p)}} \le t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx. \tag{A6.149}$$

$\sum_{i=1}^{n} \delta_i$ is the random variable ζ in Eq. (A6.120) for the binomial distribution, i.e it is the total number of occurrences of the event considered in n *Bernoulli trials*. From Eq. (A6.149) it follows that for $n \to \infty$

$$\Pr\{(\frac{\sum_{i=1}^{n} \delta_i}{n} - p) \le \psi\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{n\psi}{\sqrt{np(1-p)}}} e^{-\frac{x^2}{2}} dx, \qquad n \to \infty,$$

or (for each given $\varepsilon > 0$)

$$\Pr\{|\frac{\sum_{i=1}^{n} \delta_i}{n} - p| \le \varepsilon\} \to \frac{2}{\sqrt{2\pi}} \int_{0}^{\frac{n\varepsilon}{\sqrt{np(1-p)}}} e^{-\frac{x^2}{2}} dx, \qquad n \to \infty. \tag{A6.150}$$

Setting the right-hand side of Eq. (A6.150) equal to γ allows determination of the number of trials n for given γ, p, and ε which are necessary to fulfill the inequality $|(\delta_1 + \ldots + \delta_n)/n - p| \le \varepsilon$ with a probability γ. This result is important for reliability investigations using *Monte Carlo simulations*, see also Eq. (A8.45).

The central limit theorem can be generalized under weak conditions to the sum of independent random variables with different distribution functions [A6.6 (Vol. II), A6.7], the meaning of these conditions being that each individual standardized random variable $(\tau_i - E[\tau_i])/\sqrt{\mathrm{Var}[\tau_i]}$ provides a small contribution to the standardized sum (Lindeberg conditions).

Example A6.21

Determine (for instance in the context of *Monte Carlo simulations*) the number of trials necessary to estimate an unknown probability p with a confidence interval of length $\le 2\varepsilon$ and a confidence level approximately equal to (but not lower than) γ.

Solution

From Eq. (A6.150) it follows that for $2\varepsilon = \hat{p}_u - \hat{p}_l$ and $n \to \infty$

$$\Pr\{|\frac{\sum_{i=1}^{n} \delta_i}{n} - p| \le \varepsilon\} \approx \frac{2}{\sqrt{2\pi}} \int_{0}^{\frac{n\varepsilon}{\sqrt{np(1-p)}}} e^{-\frac{x^2}{2}} dx = \gamma.$$

Therefore,

$$\frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{n\varepsilon}{\sqrt{np(1-p)}}} e^{-\frac{x^2}{2}} dx = \frac{\gamma}{2} \qquad \text{or} \qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{n\varepsilon}{\sqrt{np(1-p)}}} e^{-\frac{x^2}{2}} dx = 0.5 + \frac{\gamma}{2} = \frac{1+\gamma}{2}$$

and thus $n\varepsilon/\sqrt{np(1-p)} = t_{(1+\gamma)/2}$, from which

$$n = (\frac{t_{(1+\gamma)/2}}{\varepsilon})^2 p(1-p), \tag{A6.151}$$

where $t_{(1+\gamma)/2}$ is the $(1+\gamma)/2$ quantile of the standard normal distribution $\Phi(t)$ (Eq. (A6.109),

Appendix A9.1). The number of trials n depend on the value of p and is a maximum ($n_{\max}$) for $p = 0.5$. The following table gives $n_{\max}$ for different values of ε and γ

$2\varepsilon = \hat{p}_u - \hat{p}_l$	0.1 ($\varepsilon = 0.05$)			0.05 ($\varepsilon = 0.025$)		
γ	0.8	0.9	0.95	0.8	0.9	0.95
$t_{(1+\gamma)/2}$	1.282	1.645	1.960	1.282	1.645	1.960
$n_{\max}$	164	271	384	657	1,082	1,537

Example A6.22

The series production of a given assembly requires 5,000 ICs of a particular type. 0.5% of these ICs are defective. How many ICs must be bought in order to be able to produce the series with a probability of $\gamma = 0.99$?

Solution

Setting $p = \Pr\{\text{IC good}\} = 0.995$, the minimum value of n satisfying

$$\Pr\{\sum_{i=1}^{n} \delta_i > 5{,}000\} \ge 0.99 = \gamma$$

must be found. Rearrangement of Eq. (A6.149), setting $t = t_{1-\gamma}$, leads to

$$\lim_{n\to\infty} \Pr\{\sum_{i=1}^{n} \delta_i > t_{1-\gamma}\sqrt{n\,p(1-p)} + n\,p\} = \frac{1}{\sqrt{2\pi}} \int_{t_{1-\gamma}}^{\infty} e^{-\frac{x^2}{2}} dx = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_{1-\gamma}} e^{-\frac{x^2}{2}} dx = \gamma,$$

where $t_{1-\gamma}$ denotes the $1-\gamma$ quantile of the standard normal distribution $\Phi(t)$ given by Eq. (A6.109) or Table A9.1. For $\gamma = 0.99$ one obtains from Table A9.1 $t_{1-\gamma} = t_{0.01} = -2.33$. With $p = 0.995$, it follows that

$$-2.33\sqrt{n \cdot 0.995 \cdot 0.005} + 0.995 n \ge 5{,}000 .$$

Thus, $n = 5{,}037$ ICs must be bought (if only $5{,}025 = 5{,}000 + 5{,}000 \cdot 0.005$ ICs were ordered, then $t_{1-\gamma} \approx 0$ and $\gamma \approx 0.5$).

Example A6.23

Electronic components are delivered with a defective probability $p = 0.1\%$. (i) How large is the probability of having exactly 8 defective components in a (homogeneous) lot of size $n = 5{,}000$? (ii) In which interval $[k_1, k_2]$ around the mean value $n\,p = 5$ will the number of defective components lie in a lot of size $n = 5{,}000$ with a probability γ as near as possible to 0.95 ?

Solution

(i) The use of the Poisson approximation (Eq. (A6.129)) leads to

$$p_8 \approx \frac{5^8}{8!} e^{-5} \approx 0.06528,$$

the exact value (obtained with Eq. (A6.120)) being 0.06527. For comparison, the following are the values of p_k obtained with the Poisson approximation (Eq. (A6.129)) in the first row and the exact values from Eq. (A6.120) in the second row

$k =$	0	1	2	3	4	5	6	7	8	9
$p_k =$	0.007	0.034	0.084	0.140	0.175	0.175	0.146	0.104	0.065	0.036
$p_k =$	0.007	0.034	0.084	0.140	0.176	0.176	0.146	0.104	0.065	0.036

(ii) From the above table one recognizes that the interval $[k_1, k_2] = [1, 9]$ is centered on the mean value $n\,p=5$ and satisfy the condition "γ as near as possible to 0.95" ($\gamma = p_1 + p_2 + \ldots + p_9 \approx 0.96$). A good approximation for k_1 and k_2 can also be obtained using Eq. (A6.151) to determine ε by given p, n, and $t_{(1+\gamma)/2}$

$$\varepsilon = \frac{\sqrt{n\,p(1-p)}}{n}\, t_{(1+\gamma)/2}, \tag{A6.152}$$

where $t_{(1+\gamma)/2}$ is the $(1+\gamma)/2$ quantile of the standard normal distribution $\Phi(t)$ (Eq. (A6.109)). Equation (A6.152) is a consequence of Eq. (A6.150) by considering that

$$\frac{2}{\sqrt{2\pi}} \int_0^A e^{-\frac{x^2}{2}} d\,x = \gamma \quad \text{yield to} \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{A} e^{-\frac{x^2}{2}} dx = 0.5 + \gamma/2 = \frac{1+\gamma}{2},$$

from which,

$$n\varepsilon / \sqrt{n\,p(1-p)} = A = t_{(1+\gamma)/2}.$$

With $\gamma = 0.95$, $t_{(1+\gamma)/2} = t_{0.975} = 1.96$ (Table A9.1), $n = 5{,}000$, and $p = 0.001$ one obtains $n\varepsilon = 4.38$, yielding $k_1 = np - n\varepsilon = 0.62$ (≥ 0) and $k_2 = np + n\varepsilon = 9.38$ ($\le n$). The same solution is also given by Eq. (A8.45)

$$k_{2,1} = np \pm b\sqrt{np(1-p)},$$

considering $b = t_{(1+\gamma)/2}$.

A7 Basic Stochastic Process Theory

Stochastic processes are powerful tools for the investigation of the reliability and availability of repairable equipment and systems. They can be considered as families of time-dependent random variables or as random functions in time, and thus have a theoretical foundation based on *probability theory* (Appendix A6). The use of stochastic processes allows the analysis of the influence of the failure-free operating and repair time distributions of elements, as well as of the system's structure, repair strategy, and logistical support, on the *reliability and availability* of a given system. Considering the applications given in Chapter 6, and for reasons of mathematical tractability, this appendix mainly deals with *regenerative stochastic processes* with a finite state space, to which belong *renewal processes, Markov processes, semi-Markov processes*, and *semi-regenerative processes*. The theoretical presentation will be supported by examples taken from practical applications. This appendix is a compendium of the theory of stochastic processes, consistent from a mathematical point of view but still with engineering applications in mind.

A7.1 Introduction

Stochastic processes are mathematical models for random phenomena evolving over time, such as the time behavior of repairable systems, the number of calls in a telephone exchange, or the noise voltage of a diode. They are designated here by Greek letters $\xi(t)$, $\zeta(t)$, $\eta(t)$, $\nu(t)$etc.

Consider the time behavior of a system subject to random influences and let T be the time interval of interest, e.g. $T = [0, \infty)$. The set of possible states of the system, i.e. the *state space*, is assumed to be a subset of the set of real numbers. The state of the system at a given time t_0 is thus a *random variable* $\xi(t_0)$. The random variables $\xi(t)$, $t \in T$, may be arbitrarily coupled together. However, for any $n = 1, 2, \ldots$, and arbitrary values $t_1, \ldots, t_n \in T$, the existence of the *n-dimensional distribution function* (Eq. (A6.51))

$$\mathrm{F}(x_1, \ldots, x_n, t_1, \ldots, t_n) = \Pr\{\xi(t_1) \le x_1, \ldots, \xi(t_n) \le x_n\} \qquad \text{(A7.1)}$$

is assumed. $\xi(t_1), \ldots, \xi(t_n)$ are thus the components of a *random vector* $\vec{\xi}(t)$. It can be shown that the family of n-dimensional distribution functions (Eq. (A7.1)) satisfies the *consistency condition*

$$\mathrm{F}(x_1, \ldots, x_k, \infty, \ldots, \infty, t_1, \ldots, t_k, t_{k+1}, \ldots, t_n) = \mathrm{F}(x_1, \ldots, x_k, t_1, \ldots, t_k), \qquad k < n$$

and the *symmetry condition*

$$\mathrm{F}(x_{i_1}, \ldots, x_{i_n}, t_{i_1}, \ldots, t_{i_n}) = \mathrm{F}(x_1, \ldots, x_n, t_1, \ldots, t_n),$$
$$i_i \in \{1, \ldots, n\}, \quad i_j \ne i_i \ \text{ for } j \ne i.$$

Conversely, if a family of distribution functions $F(x_1, \ldots, x_n, t_1, \ldots, t_n)$ satisfying the above consistency and symmetry conditions is given, then according to a theorem of A.N. Kolmogorov [A6.10], a *distribution law* on a suitable event field $\mathcal{B}^T$ of the space $\mathcal{R}^T$ consisting of all real functions on T exists. This distribution law is the distribution of a *random function* $\xi(t)$, $t \in T$, usually referred to as a *stochastic process*. The time function resulting from a particular experiment is called a *sample path* or *realization* of the stochastic process. All sample paths are in $\mathcal{R}^T$, however the set of sample paths for a particular stochastic process can be significantly smaller than $\mathcal{R}^T$, e.g. consisting only of increasing step functions. In the case of discrete time, the notion of a *sequence of random variables* ξ_n, $n \in T$ is generally used. The concept of a stochastic process generalizes the concept of a random variable introduced in Appendix A6.5. If the random variables $\xi(t)$ are defined as measurable functions $\xi(t) = \xi(t, \omega)$, $t \in T$, on a given probability space $[\Omega, \mathcal{F}, \Pr]$ then

$$\mathrm{F}(x_1, \ldots, x_n, t_1, \ldots, t_n) = \Pr\{\omega : \ \xi(t_1, \omega) \le x_1, \ldots, \xi(t_n, \omega) \le x_n\},$$

and the consistency and symmetry conditions are fulfilled. ω represents the random influence. The function $\xi(t, \omega)$, $t \in T$, is for a given ω a *realization of the stochastic process*.

The Kolmogorov theorem assures the *existence* of a stochastic process. However, the determination of all n-dimensional distribution functions is pratically impossible in a general case. Sufficient for many applications are often some specific parameters of the stochastic process involved, such as state probabilities or sojourn times. The problem considered and the model assumed generally allow determination of

- the time domain T (continuous, discrete, finite, infinite)
- the structure of the state space (continuous, discrete)
- the dependency structure of the process under consideration (e.g. memoryless)
- invariance properties with respect to time shifts (time-homogeneous, stationary).

The simplest processes in discrete time are *sequences of independent random variables* $\xi_1, \xi_2, \ldots$. Also easy to describe are *processes with independent increments*, e.g. Poisson processes, for which

$$\Pr\{\xi(t_0) \le x_0, \xi(t_1) - \xi(t_0) \le x_1, \dots, \xi(t_n) - \xi(t_{n-1}) \le x_n\} = \Pr\{\xi(t_0) \le x_0\} \prod_{i=1}^{n} \Pr\{\xi(t_i) - \xi(t_{i-1}) \le x_i\} \qquad \text{(A7.2)}$$

holds for arbitrary $n = 1, 2, \dots,$ $t_0 < \dots < t_n \in T$, and $x_1, \dots, x_n$. For reliability investigations, processes with a *continuous time parameter* $t \ge 0$ and *discrete state space* $\{Z_0, \dots, Z_m\}$ are important. Among these, the following processes will be discussed in the following sections

- renewal processes
- Markov processes
- semi-Markov processes
- semi-regenerative processes (processes with an embedded semi-Markov process)
- regenerative processes with only one or just some few regeneration states.

In view of their *dependence structure*, *Markov processes* represent a straightforward generalization of sequences of independent random variables. A Markov process is characterized by the *memoryless property*, such that the evolution of the process after an arbitrary time point t *only depends on* t *and on the state occupied at* t, but not on the evolution of the process before t (in time-homogeneous Markov processes, the dependence on t also disappears). *Markov processes* are very simple *regenerative stochastic processes*, they are regenerative with respect to each state and, if time-homogeneous, also with respect to any time t. In a general *regenerative stochastic process* there is a sequence of random points (regeneration points), at which the process forgets its foregoing evolution and (from a probabilistic point of view) restarts anew. Typically, regeneration points occur when the process returns to some particular states (regeneration states). Between regeneration points, the dependency structure of the process can be very complicated. *Semi-Markov processes* have the Markov property at the time points of *any state change*, all states of a Semi-Markov process are thus regeneration states. In a semi-regenerative process, a subset $Z_0, \dots, Z_k$ of the states $Z_0, \dots, Z_m$ are regeneration states ($k < m$) and constitute an *embedded semi-Markov process*.

In order to describe the time behavior of systems which are in statistical equilibrium, i.e. in *steady-state* or *stationary state*, stationary and time-homogeneous processes are suitable. The process $\xi(t)$ is *stationary* (strictly stationary) if for any $n = 1, 2, \dots,$ arbitrary $t_1, \dots, t_n$, and arbitrary time span a ($t_i, t_i + a \in T$)

$$F(x_1, \dots, x_n, t_1 + a, \dots, t_n + a) = F(x_1, \dots, x_n, t_1, \dots, t_n). \qquad \text{(A7.3)}$$

For $n = 1$, Eq. (A7.3) shows that the distribution function of the random variable $\xi(t)$ is independent of t. Hence, $E[\xi(t)]$, $Var[\xi(t)]$, and all other moments are independent of time. For $n = 2$, the distribution function of the two-dimensional random variable $(\xi(t), \xi(t+u))$ is only a function of u. From this it follows that the *correlation coefficient* between $\xi(t)$ and $\xi(t+u)$ is also only a function of u

$$\rho_{\xi\xi}(t, t+u) = \frac{E[(\xi(t+u) - E[\xi(t+u)])(\xi(t) - E[\xi(t)])]}{\sqrt{Var[\xi(t+u)]\,Var[\xi(t)]}} = \frac{E[\xi(t)\xi(t+u)] - E^2[\xi(t)]}{Var[\xi(t)]} = \rho_{\xi\xi}(u). \qquad \text{(A7.4)}$$

Besides stationarity in the strict sense, stationarity is also defined in the wide sense. The process $\xi(t)$ is *stationary in the wide sense* if the mean $E[\xi(t)]$ the variance $Var[\xi(t)]$, and the correlation coefficient $\rho_{\xi\xi}(t, t+u)$ are finite and independent of t. Stationarity in the strict sense of a process having a finite variance implies stationarity in the wide sense, but the contrary is generally not true (it is however true for the normal process, i.e. for a process for which all *n-dimensional distribution function* (Eq. (A7.1) are n-dimensional normal distribution functions).

A process $\xi(t)$ is a *process with stationary increments*, often also called a *time-homogeneous process* if for any $n = 1, 2, \dots,$ arbitrary time intervals (b_i, t_i), arbitrary time span a ($(t_i, t_i + a, b_i, b_i + a \in T)$, and arbitrary values of $x_1, \dots, x_n$,

$$\Pr\{\xi(t_1 + a) - \xi(b_1 + a) \le x_1, \dots, \xi(t_n + a) - \xi(b_n + a) \le x_n\} = \Pr\{\xi(t_1) - \xi(b_1) \le x_1, \dots, \xi(t_n) - \xi(b_n) \le x_n\}. \qquad \text{(A7.5)}$$

If a process $\xi(t)$ is stationary, then it is also a process with stationary increments.

The stochastic processes discussed in this appendix evolve in time, and their state space is a subset of natural numbers. Both restrictions can be omitted, without particular difficulties, with a view to a general theory of stochastic processes.

A7.2 Renewal Processes

In reliability theory, *renewal processes* describe the model of an item in continuous operation which is replaced at each failure, in a negligible amount of time, by a new, statistically identical item. Results of renewal processes are basic and can be used to investigate a broad number of practical situations.

To define the renewal process, let $\tau_0, \tau_1, \dots$ be statistically independent non-negative random variables (e.g. failure-free operating times) distributed according to

$$F_A(x) = \Pr\{\tau_0 \le x\} \qquad \text{(A7.6)}$$

and

$$F(x) = \Pr\{\tau_i \le x\}, \qquad i = 1, 2, \dots . \qquad \text{(A7.7)}$$

The random variables

$$S_n = \sum_{i=0}^{n-1} \tau_i , \qquad n = 1, 2, \dots , \tag{A7.8}$$

or equivalently the sequence $\tau_0, \tau_1, \dots$ itself constitutes a *renewal process*. The points $S_1, S_2, \dots$ are *renewal points* (regeneration points). The renewal process is a particular *point process*, see Fig. A7.1a. A *counting process*

$$\nu(t) = \begin{cases} 0 & \text{for } t < \tau_0 \\ n & \text{for } S_n \le t < S_{n+1}, \qquad n = 1, 2, \dots , \end{cases}$$

can be associated with any renewal process, and gives the number of renewal points in the interval $(0, t]$, see Fig. A7.1b. Renewal processes are *ordinary* for $F_A(x) = F(x)$, otherwise they are *modified* (*stationary* for $F_A(x)$ as in Eq. (A7.34)). To simplify the analysis, let us assume in the following that

$$F_A(0) = F(0) = 0, \tag{A7.9}$$

$$f_A(x) = \frac{dF_A(x)}{dx} \quad \text{and} \quad f(x) = \frac{dF(x)}{dx} \quad \text{exist,} \tag{A7.10}$$

$$MTTF = E[\tau_i] = \int_0^\infty (1 - F(x))dx < \infty, \qquad i \ge 1. \tag{A7.11}$$

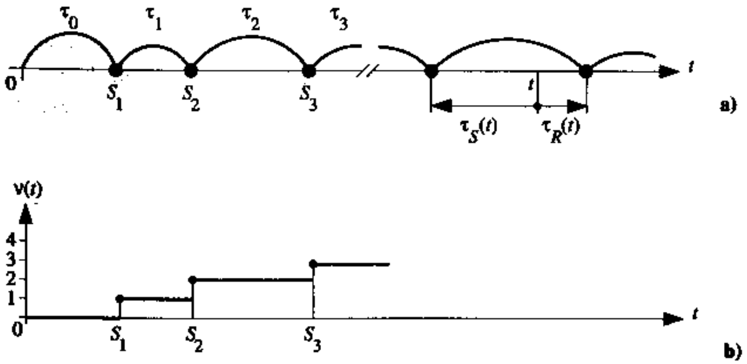


Figure A7.1 a) Time evolution of a renewal process; b) Corresponding count function $\nu(t)$

A7.2.1 Renewal Function, Renewal Density

Consider first the distribution function of the number of renewal points $\nu(t)$ in the time interval $(0, t]$. From Fig. A7.1,

$$\begin{aligned} \Pr\{\nu(t) \le n-1\} &= \Pr\{S_n > t\} = 1 - \Pr\{S_n \le t\} \\ &= 1 - \Pr\{\tau_0 + \dots + \tau_{n-1} \le t\} = 1 - F_n(t), \qquad n = 1, 2, \dots . \end{aligned} \tag{A7.12}$$

The functions $F_n(t)$ can be calculated recursively (Eq. A6.73))

$$F_1(t) = F_A(t),$$

$$F_{n+1}(t) = \int_0^t F_n(t-x) f(x) dx, \qquad n = 1, 2, \dots . \tag{A7.13}$$

From Eq. (A7.12) it follows that

$$\Pr\{\nu(t) = n\} = \Pr\{\nu(t) \le n\} - \Pr\{\nu(t) \le n-1\} = F_n(t) - F_{n+1}(t), \qquad n = 1, 2, \dots \tag{A7.14}$$

and thus, for the expected value (mean) of $\nu(t)$,

$$E[\nu(t)] = \sum_{n=1}^{\infty} n [F_n(t) - F_{n+1}(t)] = \sum_{n=1}^{\infty} F_n(t) = H(t). \tag{A7.15}$$

The function $H(t)$ defined by Eq. (A7.15) is the *renewal function*. Due to $F(0) = 0$, one has $H(0) = 0$. The distribution functions $F_n(t)$ have densities (Eq. (A6.74))

$$f_1(t) = f_A(t) \qquad \text{and} \qquad f_n(t) = \int_0^t f(x) f_{n-1}(t-x) dx, \qquad n = 2, 3, \dots , \tag{A7.16}$$

and are thus the *convolutions* of $f(x)$ with $f_{n-1}(x)$. Changing the order of summation and integration one obtains from Eq. (A7.15)

$$H(t) = \sum_{n=1}^{\infty} \int_0^t f_n(x) dx = \int_0^t \sum_{n=1}^{\infty} f_n(x) dx. \tag{A7.17}$$

The function

$$h(t) = \frac{dH(t)}{dt} = \sum_{n=1}^{\infty} f_n(t) \tag{A7.18}$$

is the *renewal density*. Using the iteration formula (A7.13), Eq. (A7.17) can be written in the form

$$H(t) = F_A(t) + \int_0^t H(x) f(t-x)\,dx. \tag{A7.19}$$

Equation (A7.19) is the *renewal equation*. The corresponding equation for the renewal density is

$$h(t) = f_A(t) + \int_0^t h(x) f(t-x)\,dx. \tag{A7.20}$$

It can be shown that Eq. (A7.20) has exactly one solution whose Laplace transform $\tilde{h}(s)$ exists and is given by (Appendix A9.7)

$$\tilde{h}(s) = \frac{\tilde{f}_A(s)}{1-\tilde{f}(s)}. \tag{A7.21}$$

For an ordinary renewal process ($F_A(x) = F(x)$),

$$\tilde{h}(s) = \frac{\tilde{f}(s)}{1-\tilde{f}(s)}. \tag{A7.22}$$

Thus, an ordinary renewal process is *completely determined* by its renewal density $h(t)$ or renewal function $H(t)$. In particular, it can be shown [6.3 (1983)] that

$$\mathrm{Var}[\nu(t)] = H(t) + 2\int_0^t h(x) H(t-x)\,dx - (H(t))^2. \tag{A7.23}$$

It is not difficult to see that $H(t) = E[\nu(t)]$ and $\mathrm{Var}[\nu(t)]$ are finite for all $t < \infty$. The renewal density $h(t)$ has the following important meaning:

Due to the assumption $F_A(0) = F(0) = 0$, *it follows that*

$$\lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{\nu(t+\delta t) - \nu(t) > 1\} = 0$$

and thus, for $\delta t \downarrow 0$,

$$\Pr\{\text{any one of the renewal points } S_1 \text{ or } S_2 \text{ or } \dots \text{ lies in } (t, t+\delta t]\} \approx h(t)\,\delta t. \tag{A7.24}$$

This interpretation of the renewal density is useful in practical applications. From Eq. (A7.24) it follows that the *renewal density* $h(t)$ *differs basically* from the *failure rate* $\lambda(t)$ defined by Eq. (A6.27)

$$\lambda(t) = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{t < \tau_0 < t + \delta t \mid \tau_0 > t\} = \frac{f_A(t)}{1 - F_A(t)}.$$

However, for the *Poisson process* $F_A(x) = F(x) = 1 - e^{-\lambda x}$ and thus $h(t) = \lambda(t) = \lambda$, see Appendix A7.2.5.

Example A7.1

Determine the renewal function $H(t)$, analytically for

(i) $f_A(t) = f(t) = \lambda e^{-\lambda t}$ (Exponential)

(ii) $f_A(t) = f(t) = 0.5\lambda(\lambda t)^2 e^{-\lambda t}$ (Erlang with $n = 3$)

(iii) $f_A(t) = f(t) = \lambda \frac{(\lambda t)^{\beta-1}}{\Gamma(\beta)} e^{-\lambda t}$ (Gamma),

and numerically for a failure rate $\lambda(t) = \lambda$ for $0 \le t < \Psi$ and $\lambda(t) = \lambda + \beta\lambda_w^\beta (t-\Psi)^{\beta-1}$ for $t \ge \Psi$, i.e. for

(iv) $$F_A(t) = F(t) = \int_0^t f(x)\,dx = \begin{cases} 1 - e^{-\lambda t} & \text{for } 0 \le t < \Psi \\ 1 - e^{-(\lambda t + \lambda_w^\beta (t-\Psi)^\beta)} & \text{for } t \ge \Psi \end{cases}$$

with $\lambda = 4 \cdot 10^{-6}\,\mathrm{h}^{-1}$, $\lambda_w = 10^{-5}\,\mathrm{h}^{-1}$, $\beta = 5$, $\psi = 2 \cdot 10^5\,\mathrm{h}$ (wearout), and for

(v) $F_A(t) = F(t)$ as in case (iv) but with $\beta = 0.3$ and $\psi = 0$ (early failures).

Give the solution in a graphical form for cases (iv) and (v).

Solution

The Laplace transformations of $f_A(t)$ and $f(t)$ for the cases (i) to (iii) are (Table A9.7b)

(i) $\tilde{f}_A(s) = \tilde{f}(s) = \lambda/(s+\lambda)$

(ii) $\tilde{f}_A(s) = \tilde{f}(s) = \lambda^3/(s+\lambda)^3$

(iii) $\tilde{f}_A(s) = \tilde{f}(s) = \lambda^\beta/(s+\lambda)^\beta$,

$\tilde{h}(s)$ follows then from Eq. (A7.22) yielding $h(t)$ or directly $H(t) = \int_0^t h(x)\,dx$

(i) $\tilde{h}(s) = \lambda/s$ and $H(t) = \lambda t$

(ii) $\tilde{h}(s) = \lambda^3/s(s^2 + 3\lambda s + 3\lambda^2) = \lambda^3/s[(s + \frac{3}{2}\lambda)^2 + \frac{3}{4}\lambda^2]$

and $H(t) = \frac{1}{3}[\lambda t - 1 + \frac{2}{\sqrt{3}} e^{-3\lambda t/2} \sin(\sqrt{3}\lambda t/2 + \frac{\pi}{3})]$

(iii) $$\tilde{h}(s) = \frac{\lambda^\beta/(s+\lambda)^\beta}{1 - \lambda^\beta/(s+\lambda)^\beta} = \sum_{n=1}^{\infty} \left[\lambda^\beta/(s+\lambda)^\beta\right]^n = \sum_{n=1}^{\infty} \frac{\lambda^{n\beta}}{(s+\lambda)^{n\beta}}$$

and $$H(t) = \sum_{n=1}^{\infty} \int_0^t \frac{\lambda^{n\beta} x^{n\beta-1}}{\Gamma(n\beta)} e^{-\lambda x}\,dx.$$

Cases (iv) and (v) can only be solved numerically or by simulation. Figure A7.2 gives the results for these two cases in a graphical form (see Eq. (A7.28) for the asymptotic behavior of $H(t)$, represented by the dashed line in Fig. A7.2a). Figure A7.2 shows that the convergence of $H(t)$ to its asymptotic value is reasonably fast, as for many practical applications. The shape of $H(t)$ can allow the recognition of the presence of wearout (iv) or early failures (v), but can not deliver (in general) a precise interpretation of the failure rate shape.

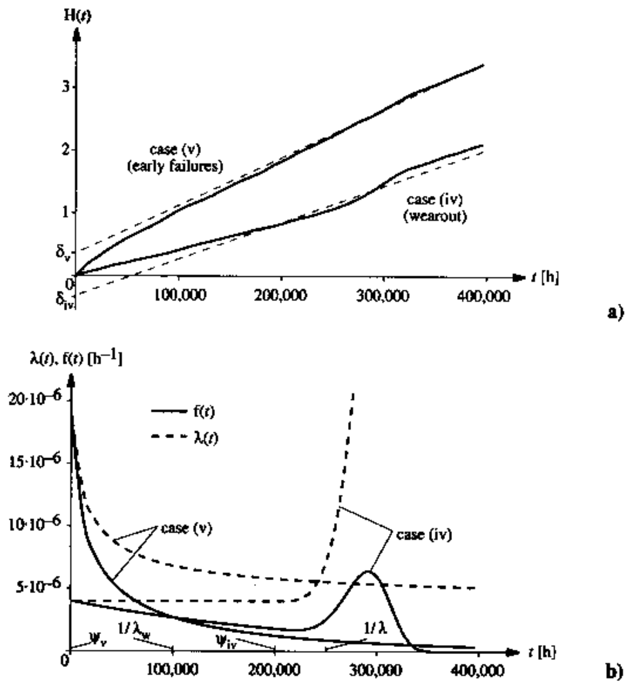


Figure A7.2 a) Renewal function $H(t)$ and b) failure rate $\lambda(t)$ and density function $f(t)$ for cases (iv) and (v) in Example A7.1 ($H(t)$ was obtained empirically, simulating 1000 failure-free times and plotting $H(t)$ as a continuous curve; $\delta = [(\sigma / MTTF)^2 - 1]/2$ according to Eq. (A7.28))

A7.2.2 Recurrence Times

Consider now the distribution functions of the forward recurrence time $\tau_R(t)$ and the backward recurrence time $\tau_S(t)$. As shown in Fig. A7.1a, $\tau_R(t)$ and $\tau_S(t)$ are the time intervals from *an arbitrary time point t* forward to the next renewal point and backward to the last renewal point (or to the time origin), respectively. It follows from Fig. A7.1a that the event $\tau_R(t) > x$ occurs with one of the following mutually exclusive events

$$A_0 = S_1 > t + x$$

$$A_n = (S_n \le t) \cap (\tau_n > t + x - S_n), \qquad n = 1, 2, \ldots.$$

Obviously, $\Pr\{A_0\} = 1 - F_A(t + x)$. The event A_n means that exactly n renewal points have occurred before t and the $(n+1)$th renewal point occurs after $t + x$. Because S_n and τ_n are independent, it follows that

$$\Pr\{A_n \mid S_n = y\} = \Pr\{\tau_n > t + x - y\}, \qquad n = 1, 2, \ldots,$$

and thus, from the theorem of total probability (Eq. (A6.17))

$$\Pr\{\tau_R(t) > x\} = 1 - F_A(t + x) + \int_0^t h(y)(1 - F(t + x - y))\,dy.$$

For the *forward recurrence time* $\tau_R(t)$ one obtains then

$$\Pr\{\tau_R(t) \le x\} = F_A(t + x) - \int_0^t h(y)(1 - F(t + x - y))\,dy. \tag{A7.25}$$

The distribution function of the *backward recurrence time* $\tau_S(t)$ can be obtained as

$$\Pr\{\tau_S(t) \le x\} = \begin{cases} \int_{t-x}^{t} h(y)(1 - F(t - y))\,dy & \text{for } x < t \\ 1 & \text{for } x \ge t. \end{cases} \tag{A7.26}$$

Since $\Pr\{S_0 > t\} = 1 - F_A(t)$, the distribution function of $\tau_S(t)$ makes a jump of height $1 - F_A(t)$ at the point $x = t$.

A7.2.3 Asymptotic Behavior

Asymptotic behavior of a renewal process (generally of a stochastic process) is understood to be the behavior of the process for $t \to \infty$. The following theorems hold with *MTTF* as in Eq. (A7.11):

1. *Elementary Renewal Theorem* [A6.6 (vol. II), A7.24]: If the conditions (A7.9)-(A7.11) are fulfilled, then

$$\lim_{t \to \infty} \frac{H(t)}{t} = \frac{1}{MTTF}, \qquad \text{where } H(t) = E[\nu(t)]. \tag{A7.27}$$

   For $Var[\nu(t)]$ it holds that $\lim_{t\to\infty} Var[\nu(t)]/t = \sigma^2/MTTF^3$, with $\sigma^2 = Var[\tau_i] < \infty$, $i \ge 1$.

2. *Tightened Elementary Renewal Theorem* [6.3, A7.4, A7.24]: If the conditions (A7.9)-(7.11) are fulfilled, $E[\tau_A] = MTTF_A < \infty$ and $\sigma^2 = Var[\tau_i] < \infty$, $i \ge 1$, then

$$\lim_{t \to \infty} \left(H(t) - \frac{t}{MTTF}\right) = \frac{\sigma^2}{2\,MTTF^2} - \frac{MTTF_A}{MTTF} + \frac{1}{2}. \tag{A7.28}$$

3. *Key Renewal Theorem* [A7.9(vol. II), A7.24]: If the conditions (A7.9) to (A7.11) are fulfilled, $\mathrm{U}(z) \geq 0$ is bounded, nonincreasing, and Riemann integrable over the interval $(0, \infty)$, and $\mathrm{h}(t)$ is a renewal density, then

$$\lim_{t \to \infty} \int_0^t \mathrm{U}(t-y)\,\mathrm{h}(y)\,dy = \frac{1}{MTTF} \int_0^\infty \mathrm{U}(z)\,dz. \tag{A7.29}$$

   For any $a > 0$, the key renewal theorem leads, with

$$\mathrm{U}(z) = \begin{cases} 1 & \text{for } 0 < z < a \\ 0 & \text{otherwise,} \end{cases}$$

   to *Blackwell's Theorem* [A7.9 (vol. II), A7.24]

$$\lim_{t \to \infty} \frac{\mathrm{H}(t+a) - \mathrm{H}(t)}{a} = \frac{1}{MTTF} .$$

   Conversely, the key renewal theorem can be obtained from Blackwell's theorem.

4. *Renewal Density Theorem* [A7.9(1941), A7.24]: If the conditions (A7.9)-(A7.11) are fulfilled, $\mathrm{f}_A(x)$ and $\mathrm{f}(x)$ go to 0 as $x \to \infty$, and $\mathrm{Var}[\tau_i] < \infty$, $i \geq 1$, then

$$\lim_{t \to \infty} \mathrm{h}(t) = \frac{1}{MTTF}. \tag{A7.30}$$

5. *Recurrence Time Limit Theorems*: Assuming $\mathrm{U}(z) = 1 - F(x+z)$ in Eq. (A7.29) and considering $\mathrm{F}_A(\infty) = 1$ as well as $1 = \int_0^\infty (1 - \mathrm{F}(y))\,dy \,/\, MTTF$, Eq. (A7.25) yields

$$\lim_{t \to \infty} \Pr\{\tau_R(t) \leq x\} = 1 - \frac{1}{MTTF} \int_0^\infty (1 - \mathrm{F}(x+z))\,dz = \frac{1}{MTTF} \int_0^x (1 - \mathrm{F}(y))\,dy. \tag{A7.31}$$

   For $t \to \infty$, the density of the *forward recurrence time* $\tau_R(t)$ is thus given by $\mathrm{f}_{\tau_R}(x) = (1 - \mathrm{F}(x)) / MTTF$. Assuming $\mathrm{E}[\tau_i] = MTTF < \infty$, $\mathrm{Var}[\tau_i] = \sigma^2 < \infty$ $(i \geq 1)$, and $\mathrm{E}[\tau_R(t)] < \infty$ it follows that $\lim_{x \to \infty} (x^2 (1 - \mathrm{F}(x))) = 0$. Integration by parts then yields

$$\lim_{t \to \infty} \mathrm{E}[\tau_R(t)] = \frac{1}{MTTF} \int_0^\infty x(1 - \mathrm{F}(x))\,dx = \frac{MTTF}{2} + \frac{\sigma^2}{2\,MTTF} . \tag{A7.32}$$

   The result of Eq. (A7.32) is important to explain the *waiting time paradox*: $\lim_{t \to \infty} \mathrm{E}[\tau_R(t)] = MTTF/2$ holds for $\sigma^2 = 0$ (i.e. for $\tau_i = MTTF$, $i \geq 0$) and $\lim_{t \to \infty} \mathrm{E}[\tau_R(t)] = \mathrm{E}[\tau_i] = 1/\lambda$, $i \geq 0$, for $F_A(x) = F(x) = 1 - e^{-\lambda x}$ (*memoryless property* of the Poisson process). Similarly, for the *backward recurrence time* $\tau_S(t)$ it follows that

$$\lim_{t \to \infty} \Pr\{\tau_S(t) \leq x\} = \frac{1}{MTTF} \int_0^x (1 - \mathrm{F}(y))\,dy \quad \text{and} \quad \lim_{t \to \infty} \mathrm{E}[\tau_S(t)] = \frac{MTTF}{2} + \frac{\sigma^2}{2\,MTTF} .$$

   For a simultaneous observation, note that $\tau_R(t)$ and $\tau_S(t)$ *belong to the same* τ_i.

6. *Central Limit Theorem for Renewal Processes* [6.3, A7.24]: If the conditions (A7.9) and (A7.11) are fulfilled and $\sigma^2 = Var[\tau_i] < \infty$, $i \geq 1$, then

$$\lim_{t \to \infty} \Pr\{\frac{\nu(t) - t / MTTF}{\sigma \sqrt{t / MTTF^3}} < x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2}\,dy. \tag{A7.33}$$

Equations (A7.27) to (A7.33) state that the renewal process with an arbitrary initial distribution function $\mathrm{F}_A(x)$ converges to its *statistical equilibrium as* $t \to \infty$, see Appendix A7.2.4 for a discussion in connection with the concept of stationary renewal process.

A7.2.4 Stationary Renewal Processes

The results of Appendix A7.2.3 allow a *stationary renewal process* to be defined as follows:

> *A renewal process is stationary (in steady-state) if for all* $t > 0$ *the distribution function of* $\tau_R(t)$ *in Eq. (A7.25) does not depend on t.*

It is intuitively clear that such a situation can only occur if a particular relationship exists between the distribution functions $\mathrm{F}_A(x)$ and $\mathrm{F}(x)$ given by Eqs. (A7.6) and (A7.7). Assuming

$$\mathrm{F}_A(x) = \frac{1}{MTTF} \int_0^x (1 - \mathrm{F}(y))\,dy, \tag{A7.34}$$

it follows that $\mathrm{f}_A(x) = (1 - \mathrm{F}(x)) / MTTF$, $\tilde{\mathrm{f}}_A(s) = (1 - \tilde{\mathrm{f}}(s)) / (s\,MTTF)$, and thus from Eq. (A7.21)

$$\tilde{\mathrm{h}}(s) = \frac{1}{s\,MTTF}$$

yielding

$$\mathrm{h}(t) = \frac{1}{MTTF}. \tag{A7.35}$$

With $\mathrm{F}_A(x)$ as in Eq. (A7.34) and $\mathrm{h}(x)$ from Eq. (A7.35), Eq. (A7.25) yields, for any $t \geq 0$,

$$\Pr\{\tau_R(t) \le x\} = \frac{1}{MTTF}\int_0^{t+x}(1-\mathrm{F}(y))\,dy - \int_0^{t}\frac{1}{MTTF}(1-\mathrm{F}(t+x-y))\,dy$$
$$= \frac{1}{MTTF}\int_0^{x}(1-\mathrm{F}(y))\,dy. \qquad \text{(A7.36)}$$

Equation (A7.34) is thus a *necessary and sufficient* condition for stationarity of the renewal process with $\Pr\{\tau_i \le x\} = \mathrm{F}(x)$, $i \ge 1$.

It is not difficult to show that the counting process $\nu(t)$ as in Fig. 7.1b, belonging to a *stationary renewal process*, is a *process with stationary increments*. For any t, $a > 0$, and $n = 1, 2, \ldots$ it follows that

$$\Pr\{\nu(t+a)-\nu(t)=n\} = \Pr\{\nu(a)=n\} = \mathrm{F}_n(a) - \mathrm{F}_{n+1}(a),$$

with $\mathrm{F}_{n+1}(a)$ as in Eq. (A7.13) and $\mathrm{F}_A(x)$ as in Eq. (A7.34). Moreover, for a stationary renewal process, $\mathrm{H}(t) = t\,/\,MTTF$ and the mean number of renewals within an arbitrary interval $(t, t+a]$ is

$$\mathrm{H}(t+a) - \mathrm{H}(t) = \frac{a}{MTTF}.$$

Comparing Eq. (A7.31) with Eq. (A7.36) it follows that under very general conditions as $t \to \infty$ every renewal process becomes stationary. From this, the following interpretation can be made which is useful for practical applications:

> *A stationary renewal process can be regarded as a renewal process with arbitrary initial condition $\mathrm{F}_A(x)$, which has been started at $t = -\infty$ and will only be considered for $t \ge 0$ ($t = 0$ being an arbitrary time point).*

The most important properties of stationary renewal processes are summarized in Table A7.1. Equation (A7.31) also obviously holds for $\tau_R(t)$ and for $\tau_S(t)$ in the case of a stationary renewal process.

A7.2.5 Poisson Processes

The renewal process defined by Eq. (A7.8) with

$$\mathrm{F}_A(x) = \mathrm{F}(x) = 1 - e^{-\lambda x} \qquad \text{(A7.37)}$$

is a (homogeneous) *Poisson process*. $\mathrm{F}_A(x)$ as in Eq. (A7.37) fulfills Eq. (A7.34). Thus, the Poisson process is stationary. From Eqs. (A7.12), (A7.14), (A7.22), (A7.25), and (A7.26) it follows that

$$\Pr\{\tau_0 + \ldots + \tau_{n-1} \le t\} = \mathrm{F}_n(t) = 1 - \sum_{i=0}^{n-1}\frac{(\lambda t)^i}{i!}e^{-\lambda t}, \qquad n = 1, 2, \ldots, \qquad \text{(A7.38)}$$

$$\Pr\{\nu(t)=n\} = \mathrm{F}_n(t) - \mathrm{F}_{n+1}(t) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}, \qquad n = 1, 2, \ldots, \qquad \text{(A7.39)}$$

$$\mathrm{H}(t) = \mathrm{E}[\nu(t)] = \lambda t, \qquad \mathrm{h}(t) = \lambda, \qquad \mathrm{Var}[\nu(t)] = \lambda t, \qquad \text{(A7.40)}$$

$$\Pr\{\tau_R(t) \le x\} = 1 - e^{-\lambda x}, \qquad t \ge 0, \qquad \text{(A7.41)}$$

$$\Pr\{\tau_S(t) \le x\} = \begin{cases} 1 - e^{-\lambda x} & \text{for } x < t \\ 1 & \text{for } x \ge t. \end{cases} \qquad \text{(A7.42)}$$

As a result of the *memoryless property* of the exponential distribution, the counting process $\nu(t)$ as in Fig. A7.1b has *independent increments*. A Poisson process can also be defined as a process with independent, stationary increments for which Eq. (A7.39) holds.

Substituting a nondecreasing (generally increasing) function $\mathrm{M}(t) > 0$ for λt in Eq. (A7.39), a *nonhomogeneous Poisson process* (NHPP) is obtained. The nonhomogeneous Poisson process is a *process with independent increments* (Eq. (A7.2)) for which

$$\Pr\{\nu(t)=n\} = \frac{(\mathrm{M}(t))^n}{n!}e^{-\mathrm{M}(t)}, \qquad \text{(A7.43)}$$

where $\mathrm{M}(t) = \mathrm{E}[\nu(t)]$ is tacitly assumed to be continuous. If

$$\mathrm{m}(t) = \frac{d\mathrm{M}(t)}{dt} \qquad \text{(A7.44)}$$

exists, then $\mathrm{m}(t)$ is the *intensity* of the nonhomogenous Poisson process. For $\mathrm{M}(t)$ continuous it follows that $\Pr\{\nu(t+\delta t)-\nu(t)=1\} = \mathrm{m}(t)\delta t + o(\delta t)$. No distinction is thus made here between *arrival rate* and *intensity*.

Table A7.1 Main properties of a stationary renewal process

	Expression	Comments, assumptions
1. Distribution function of τ_0	$\mathrm{F}_A(x) = \frac{1}{T}\int_0^x(1-\mathrm{F}(y))\,dy$	$\mathrm{f}_A(x) = dF_A(x)/dx = (1-F(x))/T$ $T = \mathrm{E}[\tau_i]$, $i \ge 1$
2. Distribution function of τ_i, $i \ge 1$	$\mathrm{F}(x)$	$\mathrm{f}(x) = \frac{d\,\mathrm{F}(x)}{dx}$, $x \ge 0$
3. Renewal function	$\mathrm{H}(t) = \frac{t}{T}$, $t \ge 0$	$\mathrm{H}(t) = \mathrm{E}[\nu(t)] = \mathrm{E}$[number of renewal points in $(0, t]$]
4. Renewal density	$\mathrm{h}(t) = \frac{1}{T}$, $t \ge 0$	$\mathrm{h}(t) = \frac{d\mathrm{H}(t)}{dt}$, $\mathrm{h}(t)\delta t \approx \lim_{\delta t \downarrow 0}\Pr\{S_1$ or S_2 or … lies in $(t, t+\delta t]\}$
5. Distribution function & mean of the forward recurrence time	$\Pr\{\tau_R(t) \le x\} = \mathrm{F}_A(x)$, $t \ge 0$ $\mathrm{E}[\tau_R(t)] = T/2 + \mathrm{Var}[\tau_i]/2T$	$\mathrm{F}_A(x)$ as in point 1, same results for $\tau_S(t)$

A7.3 Alternating Renewal Processes

The generalization of the renewal process given in Fig. A7.1a by introducing a positive random replacement time, distributed according to $G(x)$, leads to the *alternating renewal process*. An alternating renewal process is a process with two states, which alternate from one state to the other after a sojourn time distributed according to $F(x)$ and $G(x)$, respectively. Considering the reliability and availability analysis of a repairable item in Section 6.2 and in order to simplify the notation, these two states will be referred to as the *up state* and the *down state*, abbreviated as u and d, respectively.

To define an alternating renewal process, consider two *independent* renewal processes $\{\tau_i\}$ and $\{\tau'_i\}$, $i = 0, 1, \ldots$. For reliability applications, τ_i denotes the i-th *failure-free operating time* and τ'_i the i-th *repair time*. These random variables are distributed according to

$$F_A(x) \quad \text{for } \tau_0 \qquad \text{and} \qquad F(x) \quad \text{for } \tau_i, \quad i \geq 1, \tag{A7.45}$$

and

$$G_A(x) \quad \text{for } \tau'_0 \qquad \text{and} \qquad G(x) \quad \text{for } \tau'_i, \quad i \geq 1, \tag{A7.46}$$

with densities $f_A(x)$, $f(x)$, $g_A(x)$, and $g(x)$, with finite means

$$MTTF = E[\tau_i] = \int_0^\infty (1 - F(t))\,dt, \qquad i \geq 1, \tag{A7.47}$$

and

$$MTTR = E[\tau'_i] = \int_0^\infty (1 - G(t))\,dt, \qquad i \geq 1, \tag{A7.48}$$

where *MTTF* and *MTTR* are used for *mean time to failure* and *mean time to repair*. The sequences

$$\tau_0, \tau'_1, \tau_1, \tau'_2, \tau_2, \tau'_3, \ldots \qquad \text{and} \qquad \tau'_0, \tau_1, \tau'_1, \tau_2, \tau'_2, \tau_3, \ldots \tag{A7.49}$$

form two modified *alternating renewal processes*, starting at $t = 0$ with τ_0 and τ'_0, respectively. Figure A7.3 shows a time evolution of these two alternating renewal processes. *Embedded* in every of these processes are two renewal processes with renewal points S_{udui} or S_{uddi} marked with ▲ and S_{duui} or S_{dudi} marked with •, where *udu* denotes a *transition from up to down, given up at* $t = 0$, i.e.

$$S_{udu_1} = \tau_0 \qquad \text{and} \qquad S_{udu_i} = \tau_0 + (\tau'_1 + \tau_1) + \ldots + (\tau'_{i-1} + \tau_{i-1}), \qquad i \geq 1.$$

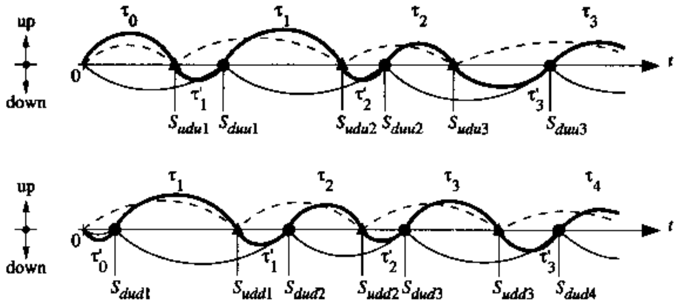


Figure A7.3 Time evolution of two alternating renewal processes starting at $t = 0$ with τ_0 and τ'_0 respectively (the four embedded renewal processes with renewal points • and ▲ are also shown)

These four *embedded renewal processes* are statistically identical up to the time intervals starting at $t = 0$, i.e. up to

$$\tau_0, \quad \tau_0 + \tau'_1, \quad \tau'_0 + \tau_1, \quad \tau'_0.$$

The corresponding densities are

$$f_A(x), \quad f_A(x) * g(x), \quad g_A(x) * f(x), \quad g_A(x)$$

for the time interval starting at $t = 0$, and

$$f(x) * g(x)$$

for all others. The symbol $*$ denotes *convolution* (Eq. (A6.75)).

The results of Section A7.2 can be used to investigate the embedded renewal processes of Fig. A7.3. Equation (A7.22) yields Laplace transforms of the renewal densities $h_{udu}(t)$, $h_{duu}(t)$, $h_{udd}(t)$, and $h_{dud}(t)$

$$\tilde{h}_{udu}(s) = \frac{\tilde{f}_A(s)}{1 - \tilde{f}(s)\tilde{g}(s)}, \qquad \tilde{h}_{duu}(s) = \frac{\tilde{f}_A(s)\tilde{g}(s)}{1 - \tilde{f}(s)\tilde{g}(s)},$$

$$\tilde{h}_{udd}(s) = \frac{\tilde{g}_A(s)\tilde{f}(s)}{1 - \tilde{f}(s)\tilde{g}(s)}, \qquad \tilde{h}_{dud}(s) = \frac{\tilde{g}_A(s)}{1 - \tilde{f}(s)\tilde{g}(s)}. \tag{A7.50}$$

To describe the alternating renewal process defined above (Fig. A7.3), let us introduce the two-dimensional stochastic process $(\zeta(t),\ \tau_{R\zeta(t)}(t))$ where $\zeta(t)$ denotes the state of the process (repairable item in reliability application)

$$\zeta(t) = \begin{cases} u & \text{if the item is up at time } t \\ d & \text{if the item is down at time } t. \end{cases}$$

$\tau_{Ru}(t)$ and $\tau_{Rd}(t)$ are thus the forward recurrence times in the up and down states, respectively, provided that the item is *up* or *down* at the time t, see Fig. 6.3.

To investigate the general case, both alternating renewal processes of Fig. A7.3 must be combined. For this let

$$p = \Pr\{\text{item up at } t=0\} \quad \text{and} \quad 1-p = \Pr\{\text{item down at } t=0\}. \qquad (A7.51)$$

In terms of the process $(\zeta(t),\ \tau_{R\zeta(t)}(t))$,

$$p=\Pr\{\zeta(0)=u\}, \qquad \mathrm{F}_A(x)=\Pr\{\tau_{Ru}(0)\le x \mid \zeta(0)=u\},$$

$$1-p=\Pr\{\zeta(0)=d\}, \qquad \mathrm{G}_A(x)=\Pr\{\tau_{Rd}(0)\le x \mid \zeta(0)=d\}.$$

Consecutive jumps from *up to down* form a renewal process with renewal density

$$\mathrm{h}_{ud}(t)=p\,\mathrm{h}_{udu}(t)+(1-p)\mathrm{h}_{udd}(t). \qquad (A7.52)$$

Similarly, the renewal density for consecutive jumps from *down to up* is given by

$$\mathrm{h}_{du}(t)=p\,\mathrm{h}_{duu}(t)+(1-p)\mathrm{h}_{dud}(t). \qquad (A7.53)$$

Using Eqs. (A7.52) and (A7.53), and considering Eq. (A7.25), it follows that

$$\Pr\{\zeta(t)=u \cap \tau_{Ru}(t)>\theta\}$$
$$= p(1-\mathrm{F}_A(t+\theta))+\int_0^t \mathrm{h}_{du}(x)(1-\mathrm{F}(t-x+\theta))dx \qquad (A7.54)$$

and

$$\Pr\{\zeta(t)=d \cap \tau_{Rd}(t)>\theta\}$$
$$=(1-p)(1-\mathrm{G}_A(t+\theta))+\int_0^t \mathrm{h}_{ud}(x)(1-\mathrm{G}(t-x+\theta))dx. \qquad (A7.55)$$

Setting $\theta=0$ in Eq. (A7.54) yields

$$\Pr\{\zeta(t)=u\}=p(1-\mathrm{F}_A(t))+\int_0^t \mathrm{h}_{du}(x)(1-\mathrm{F}(t-x))dx. \qquad (A7.56)$$

The probability $\mathrm{PA}(t)=\Pr\{\zeta(t)=u\}$ is called the *point availability* and $\mathrm{IR}(t,t+\theta]=\Pr\{\zeta(t)=u\cap\tau_{Ru}(t)>\theta\}$ the *interval reliability* of the given item (Section 6.2).

An alternating renewal process, characterized by the parameters p, $\mathrm{F}_A(x)$, $\mathrm{F}(x)$, $\mathrm{G}_A(x)$, and $\mathrm{G}(x)$ is *stationary* if the two-dimensional process $(\zeta(t),\ \tau_{R\zeta(t)}(t))$ is stationary. As with the renewal process it can be shown (e.g. using Laplace transforms) that an alternating renewal process is *stationary* if and only if

$$p=\frac{MTTF}{MTTF+MTTR}, \quad \mathrm{F}_A(x)=\frac{1}{MTTF}\int_0^x(1-\mathrm{F}(y))dy, \quad G_A(x)=\frac{1}{MTTR}\int_0^x(1-\mathrm{G}(y))dy, \qquad (A7.57)$$

with $MTTF$ and $MTTR$ as in Eqs. (A7.47) and (A7.48). In particular, for $t\ge 0$ the following relationships apply for the *stationary alternating renewal process* (Examples 6.3 and 6.4)

$$\mathrm{PA}(t)=\Pr\{\text{item up at } t\}=\frac{MTTF}{MTTF+MTTR}=PA, \qquad (A7.58)$$

$$\mathrm{IR}(t,t+\theta)=\Pr\{\text{item up at } t \text{ and remains up until } t+\theta\}$$
$$=\frac{1}{MTTF+MTTR}\int_\theta^\infty(1-\mathrm{F}(y))dy. \qquad (A7.59)$$

Condition (A7.57) is equivalent to

$$\mathrm{h}_{ud}(t)=\mathrm{h}_{du}(t)=\frac{1}{MTTF+MTTR}, \qquad t\ge 0. \qquad (A7.60)$$

Moreover, application of the *key renewal theorem* (Eq. (A7.29)) to Eqs. (A7.54) to (A7.56) yields (Example 6.4)

$$\lim_{t\to\infty}\Pr\{\zeta(t)=u\cap\tau_{Ru}(t)>\theta\}=\frac{1}{MTTF+MTTR}\int_\theta^\infty(1-\mathrm{F}(y))dy, \qquad (A7.61)$$

$$\lim_{t\to\infty}\Pr\{\zeta(t)=d\cap\tau_{Rd}(t)>\theta\}=\frac{1}{MTTF+MTTR}\int_\theta^\infty(1-\mathrm{G}(y))dy, \qquad (A7.62)$$

$$\lim_{t\to\infty}\Pr\{\zeta(t)=u\}=\lim_{t\to\infty}\mathrm{PA}(t)=\mathrm{PA}=\frac{MTTF}{MTTF+MTTR}. \qquad (A7.63)$$

Thus, irrespective of its initial conditions p, $\mathrm{F}_A(x)$, and $\mathrm{G}_A(x)$, an alternating renewal process has for $t\to\infty$ an *asymptotic behavior* which is identical to the stationary state (*steady-state*). In other words:

> *A stationary alternating renewal process can be regarded as an alternating renewal process with arbitrary initial conditions p, $\mathrm{F}_A(x)$, and $\mathrm{G}_A(x)$, which has been started at $t=-\infty$ and will only be considered for $t\ge 0$ ($t=0$ being an arbitrary time point).*

It should be noted that the results of this section remain valid even if independence between τ_i and τ_i' within a *cycle* (e.g. $\tau_0+\tau_1'$, $\tau_1+\tau_2'$, ...) is dropped; only independence between cycles is necessary. For *exponentially distributed* τ_i and τ_i', i.e. for *constant failure rate* λ and *repair rate* μ in reliability applications, the *convergence* of $\mathrm{PA}(t)$ towards PA stated by Eq. (A7.63) is of the form $\mathrm{PA}(t) - PA = (\lambda/(\lambda+\mu))e^{-(\lambda+\mu)t} \approx (\lambda/\mu)e^{-\mu t}$, see Eq. (6.20) and Section 6.2.4 for further considerations.

A7.4 Regenerative Processes *)

A *regenerative process* is characterized by the property that there is a sequence of random points on the time axis, *regeneration points*, at which the process *forgets its foregoing evolution* and, from a probabilistic point of view, restarts anew. The times at which a regenerative process restarts occur when the process returns to some states, defined as *regeneration states*. The sequence of these time points for a specific regeneration state is a *renewal process embedded* in the original stochastic process. For example, both the states up and down of an alternating renewal process are regeneration states. All states of time homogeneous Markov processes and of semi-Markov processes, defined by Eqs. (A7.95) and (A7.144), are regenerative. However there are processes in discrete state space with only few (two in Fig. A7.10, one in Fig. 6.10) or even with no (Appendix A7.8) regeneration states. A regenerative process must have *at least one regeneration state.*

A regenerative process thus consists of *independent cycles* which describe the time behavior of the process between two consecutive regeneration points *of the same type* (same regeneration state). The *i*-th *cycle* is characterized by a positive random variable τ_{c_i} (duration of cycle *i*) and a stochastic process $\xi_i(t)$ defined for $0 \le t < \tau_{c_i}$ (content of the cycle). Let $\xi_n(t)$, $0 \le t < \tau_{c_n}$, $n = 0, 1, \ldots$ be independent and for $n \ge 1$ identically distributed cycles. For simplicity, let us assume that the time points $S_1 = \tau_{c_0}$, $S_2 = \tau_{c_0} + \tau_{c_1}$, ... form a *renewal process.* The random variables τ_{c_0} and τ_{c_i}, $i \ge 1$, have distribution functions $\mathrm{F}_A(x)$ for τ_{c_0} and $\mathrm{F}(x)$ for τ_{c_i}, densities $\mathrm{f}_A(x)$ and $\mathrm{f}(x)$, and finite means T_A and T_c, respectively. The regenerative process $\xi(t)$ is then given by

$$\xi(t) = \begin{cases} \xi_0(t) & \text{for } 0 \le t < S_1 \\ \xi_n(t - S_n) & \text{for } S_n \le t < S_{n+1}, \quad n = 1, 2, \ldots . \end{cases}$$

The regenerative structure is sufficient for the existence of an *asymptotic behavior* (limiting distribution) for the process as $t \to \infty$ (provided that the mean time between regeneration points is finite). This limiting distribution is determined by the behavior of the process between two consecutive regeneration points of the same regeneration state.

---

*) *Appendix A7.4 can be omitted at a first reading.*

Defining $\mathrm{h}(t)$ as the renewal density of the renewal process given by $S_1, S_2, \ldots$ and setting

$$\mathrm{U}(t,\mathrm{B}) = \Pr\{\xi_i(t) \in \mathrm{B} \cap \tau_{c_i} > t\}, \qquad i = 1, 2, \ldots ,$$

it follows, similarly to Eq. (A7.25), that

$$\Pr\{\xi(t) \in \mathrm{B}\} = \Pr\{\xi_0(t) \in \mathrm{B} \cap \tau_{c_0} > t\} + \int_0^t \mathrm{h}(x)\mathrm{U}(t-x,\mathrm{B})\,dx. \tag{A7.64}$$

For any given distribution of the cycle $\xi_i(t)$, $0 \le t < \tau_{c_i}$, $i \ge 1$, with $T_c = E[\tau_{c_i}] < \infty$, there exists a stationary regenerative process $\xi_e(t)$ with regeneration points S_{e_i}, $i \ge 1$. The cycles $\xi_{e_n}(t)$, $0 \le t < \tau_{e_n}$, have for $n \ge 1$ the same distribution law as $\xi_i(t)$, $0 \le t < \tau_{c_i}$. The distribution law of the starting cycle $\xi_{e_0}(t)$, $0 \le t < \tau_{e_0}$, can be calculated from the distribution law of $\xi_i(t)$, $0 \le t < \tau_{c_i}$, see Eq. (A7.57) for alternating renewal processes. In particular,

$$\Pr\{\xi_e(0) \in \mathrm{B}\} = \frac{1}{T_c}\int_0^\infty \mathrm{U}(t,\mathrm{B})\,dt, \tag{A7.65}$$

with $T_c = E[\tau_{c_i}] < \infty$, $i \ge 1$. Furthermore, for every non-negative function $\mathrm{g}(t)$ and $S_1 = 0$,

$$\mathrm{E}[\mathrm{g}(\xi_e(0))] = \frac{1}{T_c}\mathrm{E}[\int_0^{\tau_{c_1}} \mathrm{g}(\xi_1(x))\,dx]. \tag{A7.66}$$

Eqation (A7.66) is known as the *stochastic mean value theorem.*

Since $\mathrm{U}(t,B)$ is nonincreasing and $\le 1 - \mathrm{F}(t)$ for all $t \ge 0$, it follows from Eq. (A7.64) and the *key renewal theorem* (Eq. (A7.29)) that

$$\lim_{t \to \infty} \Pr\{\xi(t) \in \mathrm{B}\} = \frac{1}{T_c}\int_0^\infty \mathrm{U}(t,\mathrm{B})\,dt . \tag{A7.67}$$

Equations (A7.65) and (A7.67) show that under very general conditions as $t \to \infty$ a regenerative process becomes *stationary.* As in the case of renewal and alternating renewal processes, the following interpretation is true:

> *A stationary regenerative process can be considered as a regenerative process with arbitrary distribution of the starting cycle, which has been started at $t = -\infty$ and will only be considered for $t \ge 0$ ($t = 0$ being an arbitrary time point).*

A7.5 Markov Processes with Finitely Many States

Markov processes are processes *without memory*. They are characterized by the property that for any arbitrarily chosen time point t their evolution after t depends on t and the state occupied at t, but not *on the process evolution up to the time t*. In the case of *time-homogeneous* Markov processes, dependence on t also disappears, such that *future evolution of the process depends only on the current state*. In reliability theory, these processes describe the behavior of repairable systems with *constant failure and repair rates* for all elements (constant *during the sojourn time in each state*, but not necessarily at a state change, for instance because of load sharing, see for example Figs A7.4, 6.8, 6.13, and 2.12). After a short introduction to *Markov chains* this section deals with time-homogeneous *Markov processes* with finitely many states, as basis for reliability investigations in Chaptere 6.

A7.5.1 Markov Chains with Finitely Many States

A stochastic process in discrete time ξ_n, $n = 0, 1, \ldots$ with finitely many states $Z_0, \ldots, Z_m$ is a *Markov chain* if for $n = 1, 2, \ldots$ and arbitrary $i, j, i_0, \ldots, i_{n-1} \in \{0, \ldots, m\}$,

$$\Pr\{\xi_{n+1} = Z_j \mid (\xi_n = Z_i \cap \xi_{n-1} = Z_{i_{n-1}} \cap \ldots \cap \xi_0 = Z_{i_0})\}$$
$$= \Pr\{\xi_{n+1} = Z_j \mid \xi_n = Z_i\} = p_{ij}(n). \qquad \text{(A7.68)}$$

The quantities $\mathrm{p}_{ij}(n)$ are the (one step) *transition probabilities of the Markov chain*. Investigation will be limited here to *time-homogeneous* Markov chains, for which the transition probabilities $\mathrm{p}_{ij}(n)$ are independent of n

$$p_{ij}(n) = p_{ij} = \Pr\{\xi_{n+1} = Z_j \mid \xi_n = Z_i\}, \qquad n = 0, 1, \ldots . \qquad \text{(A7.69)}$$

The probabilities p_{ij} satisfy the relationships

$$p_{ij} \ge 0 \qquad \text{and} \qquad \sum_{j=0}^{m} p_{ij} = 1, \qquad i, j \in \{0, \ldots, m\}. \qquad \text{(A7.70)}$$

A matrix with elements p_{ij} as in Eq. (A7.70) is a *stochastic matrix*. The k-step transition probabilities are the elements of the kth power of the stochastic matrix with elements p_{ij}. For example, $k = 2$ leads to (Example A7.2)

$$p_{ij}^{(2)} = \Pr\{\xi_{n+2} = Z_j \mid \xi_n = Z_i\} = \sum_{k=0}^{m} \Pr\{(\xi_{n+2} = Z_j \cap \xi_{n+1} = Z_k) \mid \xi_n = Z_i\}$$
$$= \sum_{k=0}^{m} \Pr\{\xi_{n+1} = Z_k \mid \xi_n = Z_i\} \Pr\{\xi_{n+2} = Z_j \mid (\xi_n = Z_i \cap \xi_{n+1} = Z_k)\},$$

from which, considering the Markov property (A7.68)

$$p_{ij}^{(2)} = \sum_{k=0}^{m} \Pr\{\xi_{n+1} = Z_k \mid \xi_n = Z_i\} \Pr\{\xi_{n+2} = Z_j \mid \xi_{n+1} = Z_k\} = \sum_{k=0}^{m} p_{ik}\, p_{kj} . \qquad \text{(A7.71)}$$

Results for $k > 2$ follow by induction.

Example A7.2

Prove that $\mathrm{Q}_{ij}(x)$ as in Eq. (A7.95) can be expressed by $\mathrm{Q}_{ij}(x) = p_{ij}\,\mathrm{F}_{ij}(x)$.

Solution

Let $(\xi_{n+1} = Z_j) = B$, $(\eta_n \le x) = A$, and $(\xi_n = Z_i) = C$. For $\Pr\{C\} > 0$ it follows that

$$\Pr\{(A \cap B) \mid C\} = \frac{\Pr\{A \cap B \cap C\}}{\Pr\{C\}} = \frac{\Pr\{B \cap C\} \Pr\{A \mid (B \cap C)\}}{\Pr\{C\}} = \Pr\{B \mid C\} \Pr\{A \mid (B \cap C)\}.$$

The *distribution law* of a Markov chain is completely determined by the *initial distribution*

$$A_i = \Pr\{\xi_0 = Z_i\}, \qquad i = 0, \ldots, m, \qquad \text{(A7.72)}$$

and the *transition probabilities* p_{ij}, since for every $n > 0$ and arbitrary indices $i_0, \ldots, i_n \in \{0, \ldots, m\}$,

$$\Pr\{\xi_0 = Z_{i_0} \cap \xi_1 = Z_{i_1} \cap \ldots \cap \xi_n = Z_{i_n}\} = A_{i_0}\, p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$$

and thus (theorem of total probability)

$$\Pr\{\xi_n = Z_j\} = \sum_{i=0}^{m} A_i\, p_{ij}^{(n)}, \qquad n = 0, 1, \ldots . \qquad \text{(A7.73)}$$

A Markov chain with transition probabilities p_{ij} is *stationary* if and only if the state probabilities $\Pr\{\xi_n = Z_j\}$, $j = 0, \ldots, m$, are independent of n, i.e. if the initial distribution A_i according to Eq. (A7.72) is a solution (p_j) of the system

$$p_j = \sum_{i=0}^{m} p_i\, p_{ij}, \qquad \text{with} \quad p_j \ge 0 \quad \text{and} \quad \sum_{j=1}^{m} p_j = 1, \qquad j = 0, \ldots, m. \qquad \text{(A7.74)}$$

The values $p_0, \ldots, p_m$ satisfying Eq. (A7.74) define the *stationary distribution* of the Markov chain with transition probabilities p_{ij}, see Eqs. (A7.75) and (A7.76).

A Markov chain with transition probabilities p_{ij} is *irreducible* if every state can be reached from every other state, i.e. if for each couple i, j there is an index $n = \mathrm{n}(i, j)$ such that

$$p_{ij}^{(n)} > 0, \qquad i, j \in \{0, \ldots, m\}, \qquad n \ge 1. \qquad \text{(A7.75)}$$

It can be shown that the system (A7.74) possesses a *unique solution* with

$$p_j > 0 \qquad \text{and} \qquad \sum_{j=1}^{m} p_j = 1, \qquad j = 0, \ldots, m, \tag{A7.76}$$

if and only if the Markov chain is irreducible, see e.g. [A7.3, A7.13, A7.27, A7.29].

A7.5.2 Markov Processes with Finitely Many States

A stochastic process $\xi(t)$ in continuous time with finitely many states $Z_0, \ldots, Z_m$ is a *Markov process* if for $n = 1, 2, \ldots$, arbitrary time points $t + a > t > t_n > \ldots > t_1$, and arbitrary $i, j, i_1, \ldots, i_n \in \{0, \ldots, m\}$,

$$\begin{aligned} \Pr\{\xi(t+a) &= Z_j \mid (\xi(t) = Z_i \cap \xi(t_n) = Z_{i_n} \cap \ldots \cap \xi(t_1) = Z_{i_1})\} \\ &= \Pr\{\xi(t+a) = Z_j \mid \xi(t) = Z_i\}. \end{aligned} \tag{A7.77}$$

The conditional state probabilities in Eq. (A7.77) are the *transition probabilities* of the Markov process and they will be designated by $P_{ij}(t, t+a)$

$$P_{ij}(t, t+a) = \Pr\{\xi(t+a) = Z_j \mid \xi(t) = Z_i\}. \tag{A7.78}$$

Equations (A7.77) and (A7.78) give merely the probability that $\xi(t+a)$ will be Z_j given that $\xi(t)$ was Z_i. Between t and $t+a$ the Markov process *can visit any other state* (this is not the case in Eq. (A7.95), in which Z_j is the *next state* visited after Z_i).

The Markov process is *time-homogeneous* if

$$P_{ij}(t, t+a) = P_{ij}(a) \tag{A7.79}$$

holds. In the following only time-homogeneous Markov processes will be considered. For arbitrary $t > 0$ and $a > 0$, $P_{ij}(t+a)$ satisfy the *Chapman-Kolmogorov equations*

$$P_{ij}(t+a) = \sum_{k=0}^{m} P_{ik}(t) P_{kj}(a), \qquad i, j \in \{0, \ldots, m\}, \tag{A7.80}$$

the demonstration of which is similar to that for $p_{ij}^{(2)}$ as in Eq. (A7.71). Furthermore $P_{ij}(a)$ satisfy the conditions

$$P_{ij}(a) \geq 0 \qquad \text{and} \qquad \sum_{j=0}^{m} P_{ij}(a) = 1, \quad i = 0, \ldots, m. \tag{A7.81}$$

and thus form a *stochastic matrix*. Together with the *initial distribution*

$$P_i(0) = \Pr\{\xi(0) = Z_i\}, \qquad i = 0, \ldots, m, \tag{A7.82}$$

the transition probabilities $P_{ij}(a)$ completely determine the distribution law of the Markov process. In particular, the *state probabilities* for $t > 0$

$$P_j(t) = \Pr\{\xi(t) = Z_j\}, \qquad i = 0, \ldots, m \tag{A7.83}$$

can be obtained from

$$P_j(t) = \sum_{i=0}^{m} P_i(0) P_{ij}(t). \tag{A7.84}$$

Setting

$$P_{ij}(0) = \delta_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases} \tag{A7.85}$$

and assuming that the transition probabilities $P_{ij}(t)$ are *continuous* at $t = 0$, it can be shown that $P_{ij}(t)$ are also differentiable at $t = 0$. The limiting values

$$\lim_{\delta t \downarrow 0} \frac{P_{ij}(\delta t)}{\delta t} = \rho_{ij}, \quad \text{for } i \neq j, \qquad \text{and} \qquad \lim_{\delta t \downarrow 0} \frac{1 - P_{ii}(\delta t)}{\delta t} = \rho_i, \tag{A7.86}$$

exist and satisfy

$$\rho_i = \sum_{\substack{j=0 \\ j \neq i}}^{m} \rho_{ij}, \qquad i = 0, \ldots, m. \tag{A7.87}$$

Equation (A7.86) can be written in the form

$$P_{ij}(\delta t) = \rho_{ij}\,\delta t + o(\delta t) \qquad \text{and} \qquad 1 - P_{ii}(\delta t) = \rho_i\,\delta t + o(\delta t), \tag{A7.88}$$

where $o(\delta t)$ denotes a quantity having *an order higher than that of* δt, i.e.

$$\lim_{\delta t \downarrow 0} \frac{o(\delta t)}{\delta t} = 0. \tag{A7.89}$$

Considering for any $t \geq 0$

$$P_{ij}(\delta t) = \Pr\{\xi(t + \delta t] = Z_j \mid \xi(t) = Z_i\},$$

the following useful interpretation for ρ_{ij} and ρ_i can be obtained for $\delta t \downarrow 0$ and arbitrary t

$$\begin{aligned} &\rho_{ij}\,\delta t \approx \Pr\{\text{jump from } Z_i \text{ to } Z_j \text{ in } (t, t+\delta t] \mid \xi(t) = Z_i\} \\ &\rho_i\,\delta t \approx \Pr\{\text{leave } Z_i \text{ in } (t, t+\delta t] \mid \xi(t) = Z_i\}. \end{aligned} \tag{A7.90}$$

It is thus reasonable to define ρ_{ij} and ρ_i as *transition rates* (for Markov processes, ρ_{ij} play a similar role to that of the transition probabilities for Markov chains).

Setting $a = \delta t$ in Eq. (A7.80) and considering Eqs. A7.78) and (A7.79) yields

$$\mathrm{P}_{ij}(t+\delta t) = \sum_{\substack{k=0\\k\neq j}}^{m} \mathrm{P}_{ik}(t)\mathrm{P}_{kj}(\delta t) + \mathrm{P}_{ij}(t)\mathrm{P}_{jj}(\delta t)$$

or

$$\frac{\mathrm{P}_{ij}(t+\delta t)-\mathrm{P}_{ij}(t)}{\delta t} = \sum_{\substack{k=0\\k\neq j}}^{m} \mathrm{P}_{ik}(t)\frac{\mathrm{P}_{kj}(\delta t)}{\delta t} + \mathrm{P}_{ij}(t)\frac{\mathrm{P}_{jj}(\delta t)-1}{\delta t} ,$$

and then, taking into account Eq. (A7.86), it follows that

$$\dot{\mathrm{P}}_{ij}(t) = -\mathrm{P}_{ij}(t)\rho_j + \sum_{\substack{k=0\\k\neq j}}^{m} \mathrm{P}_{ik}(t)\rho_{kj}, \qquad i,j \in \{0, \ldots, m\}. \tag{A7.91}$$

Equations (A7.91) are the *Kolmogorov forward equations*. With initial conditions $\mathrm{P}_{ij}(0) = \delta_{ij}$ as in Eq. (A7.85), they have a *unique solution* which satisfies Eq. (A7.81). In other words, the transition rates according to Eq. (A7.86) or Eq. (A7.90) *uniquely determine* the transition probabilities $\mathrm{P}_{ij}(t)$ of the Markov process. Similarly as for Eq. (A7.91), it can be shown that the transition probabilities $\mathrm{P}_{ij}(t)$ also satisfy the *Kolmogorov backward equations*

$$\dot{\mathrm{P}}_{ij}(t) = -\rho_i \mathrm{P}_{ij}(t) + \sum_{\substack{k=0\\k\neq i}}^{m} \rho_{ik}\mathrm{P}_{kj}(t), \qquad i,j \in \{0, \ldots, m\}, \tag{A7.92}$$

The following description of the (time-homogeneous) *Markov process* with initial distribution $\mathrm{P}_i(0)$ and transition rates ρ_{ij}, $i, j \in \{0, \ldots, m\}$, provides a *better insight* into the structure of a Markov process as a pure jump process (process with piecewise constant sample paths (realization)). It is the basis for investigations of Markov processes by means of *integral equations* (Section A7.5.3.2), and is the motivation for the introduction of *semi-Markov processes* (Section A7.6). Let $\xi_0, \xi_1, \ldots$ be a sequence of random variables taking values in $\{Z_0, \ldots, Z_m\}$ denoting the states *successively* occupied and $\eta_0, \eta_1, \ldots$ a sequence of positive random variables denoting the *sojourn times between two consecutive* state transitions. Define

$$p_{ij} = \frac{\rho_{ij}}{\rho_i}, \quad i \neq j \qquad \text{and} \qquad p_{ii} = 0, \qquad i,j \in \{0, \ldots, m\}, \tag{A7.93}$$

and assume furthermore that

$$\Pr\{\xi_0 = Z_i\} = \mathrm{P}_i(0), \qquad i = 0, \ldots, m, \tag{A7.94}$$

and, for $n = 1, 2, \ldots$, arbitrary $i, j, i_0, \ldots, i_{n-1} \in \{0, \ldots, m\}$, and arbitrary positive values $x_0, \ldots, x_{n-1}$,

$$\Pr\{(\xi_{n+1} = Z_j \cap \eta_n \leq x) \,|\, (\xi_n = Z_i \cap \eta_{n-1} = x_{n-1} \cap \ldots \cap \xi_1 = Z_{i_1} \cap \eta_0 = x_0 \cap \xi_0 = Z_{i_0})\}$$
$$= \Pr\{(\xi_{n+1} = Z_j \cap \eta_n \leq x) \,|\, \xi_n = Z_i\} = \mathrm{Q}_{ij}(x) = p_{ij}\,\mathrm{F}_{ij}(x) = p_{ij}\,(1 - e^{-\rho_i x}). \tag{A7.95}$$

In Eq. (A7.95), as well as in Eq. (A7.144), Z_j is the *next state* visited after Z_i (this is not the case in Eq. (A7.77), see also the remark with Eq. (A7.106)). $\mathrm{Q}_{ij}(x)$ is thus defined only for $j \neq i$. $\xi_0, \xi_1, \ldots$ is a *Markov chain*, with an *initial distribution*

$$\mathrm{P}_i(0) = \Pr\{\xi_0 = Z_i\}$$

and *transition probabilities*

$$p_{ij} = \Pr\{\xi_{n+1} = Z_j \,|\, \xi_n = Z_i\}, \qquad \text{with} \qquad p_{ii} = 0,$$

embedded in the original process. From Eq. (A7.95), it follows that (Example A7.2)

$$\mathrm{F}_{ij}(x) = \Pr\{\eta_n \leq x \,|\, (\xi_n = Z_i \cap \xi_{n+1} = Z_j)\} = 1 - e^{-\rho_i x}. \tag{A7.96}$$

$\mathrm{Q}_{ij}(x)$ is a *semi-Markov transition probability* and will as such be introduced and discussed in Section A7.6. Now, define

$$S_0 = 0, \qquad S_n = \eta_0 + \ldots + \eta_{n-1}, \qquad n = 1, 2, \ldots, \tag{A7.97}$$

and

$$\xi(t) = \xi_n, \qquad \text{for} \quad S_n \leq t < S_{n+1}. \tag{A7.98}$$

From Eq. (A7.98) and the *memoryless property* of the exponential distribution (Eq. (A6.87)) it follows that $\xi(t)$, $t \geq 0$ is a *Markov process* with an *initial distribution*

$$\mathrm{P}_i(0) = \Pr\{\xi(0) = Z_i\}$$

and *transition rates*

$$\rho_{ij} = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{\text{jump from } Z_i \text{ to } Z_j \text{ in } (t, t+\delta t] \,|\, \xi(t) = Z_i\}, \qquad j \neq i$$

and

$$\rho_i = \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \Pr\{\text{leave } Z_i \text{ in } (t, t+\delta t] \,|\, \xi(t) = Z_i\} = \sum_{\substack{j=0\\j\neq i}}^{m} \rho_{ij}.$$

The evolution of a (time-homogeneous) *Markov process* with transition rates ρ_{ij}, and ρ_i can thus be described in the following way [A7.2 (1974, Ph.D.)]:

> *If at $t = 0$ the process enters the state Z_i, i.e. $\xi_0 = Z_i$, then the next state to be entered, say Z_j ($j \neq i$) is selected according to the probability p_{ij}, and the sojourn time in Z_i is a random variable η_0 with distribution function*
>
> $$\Pr\{\eta_0 \leq x \,|\, (\xi_0 = Z_i \cap \xi_1 = Z_j)\} = 1 - e^{-\rho_i x};$$

as the process enters Z_j, *the next state to be entered, say* Z_k $(k \neq j)$, *will be selected with probability* p_{jk} *and the sojourn time* η_1 *in* Z_j *will be distributed according to*

$$\Pr\{\eta_1 \le x \mid (\xi_1 = Z_j \cap \xi_2 = Z_k)\} = 1 - e^{-\rho_j x}$$

etc.

The sequence ξ_n, $n = 0, 1, \ldots$ of the states *successively occupied* by the process is that of the Markov chain *embedded* in $\xi(t)$, the so called *embedded Markov chain*. The random variable η_n is the *sojourn time* of the process in the state defined by ξ_n. From the above description it follows that each state Z_i, $i = 0, \ldots, m$, is a *regeneration state*.

In practical applications, the following technique can be used to *determine the quantities* $Q_{ij}(x)$, p_{ij}, *and* $F_{ij}(x)$ in Eq. (A7.95):

If the process enters the state Z_i *at an arbitrary time, say at* $t = 0$, *then a set of independent random times* τ_{ij}, $j \neq i$, *begin* (τ_{ij} *is the sojourn time in* Z_i *with the next jump to* Z_j); *the process will then jump to* Z_j *at the time* x *if* $\tau_{ij} = x$ *and* $\tau_{ik} > \tau_{ij}$ *for (all)* $k \neq j$.

In this interpretation, the quantities $Q_{ij}(x)$, p_{ij}, and $F_{ij}(x)$ are given by

$$Q_{ij}(x) = \Pr\{\tau_{ij} \le x \cap \tau_{ik} > \tau_{ij},\ k \neq j\}, \tag{A7.99}$$

$$p_{ij} = \Pr\{\tau_{ik} > \tau_{ij},\ k \neq j\}, \tag{A7.100}$$

$$F_{ij}(x) = \Pr\{\tau_{ij} \le x \mid \tau_{ik} > \tau_{ij},\ k \neq j\}. \tag{A7.101}$$

Assuming for the Markov process (memoryless property stated by Eq. (A7.77))

$$\Pr\{\tau_{ij} \le x\} = 1 - e^{-\rho_{ij} x}$$

one obtains, as in Eq. (A7.95),

$$Q_{ij}(x) = \int_0^x \rho_{ij}\, e^{-\rho_{ij} y} \prod_{\substack{k=0\\k\neq j}}^{m} e^{-\rho_{ik} y}\, dy = \frac{\rho_{ij}}{\rho_i}(1 - e^{-\rho_i x}), \qquad j \neq i, \tag{A7.102}$$

$$p_{ij} = \frac{\rho_{ij}}{\rho_i} = Q_{ij}(\infty) \quad \text{for } j \neq i, \qquad \rho_i = \sum_{\substack{j=0\\j\neq i}}^{m} \rho_{ij}, \qquad p_{ii} \equiv 0, \tag{A7.103}$$

$$F_{ij}(x) = 1 - e^{-\rho_i x}. \tag{A7.104}$$

It should be emphasized that due to the *memoryless property* of the Markov process, there is no difference whether the process enters Z_i at $t = 0$ or whether it is already there. However, this is not true for a *semi-Markov* process, see Appendix A7.6.

Quite generally, a repairable system can be described by a (time-homogeneous) Markov process if and only if *all random variables occurring* (failure-free operating times and repair times of all elements) are *independent* and *exponentially distributed*. If some failure-free operating times or repair times of elements are Erlang distributed (Appendix A6.10.3), the time evolution of the system can be described by means of a Markov process with appropriate extension of the state space (see Fig. 6.6 for an example).

A useful tool when investigating a Markov process is the *diagram of transition probabilities* in $(t, t + \delta t]$ where $\delta t \to 0$ ($\delta t > 0$, i.e. $\delta t \downarrow 0$) and t is an *arbitrary time point*, e.g. $t = 0$. This diagram is a directed graph with nodes labeled by states Z_i, $i = 0, \ldots, m$, and arcs labeled by transition probabilities $P_{ij}(\delta t)$, where terms of order $o(\delta t)$ are omitted (it is an extension of the *state transition diagram*, more appropriate for practical applications). Taking into account the properties of the random variables τ_{ij} introduced with Eq. (A7.99) yields for $\delta t \to 0$

$$\Pr\{(\xi(\delta t) = Z_j \cap \text{only one jump in } (0, \delta t]) \mid \xi(0) = Z_i\}$$
$$= (1 - e^{-\rho_{ij}\, \delta t}) \prod_{\substack{k=0\\k\neq j}}^{m} e^{-\rho_{ik}\, \delta t} = \rho_{ij}\delta t + o(\delta t), \tag{A7.105}$$

and

$$\Pr\{(\xi(\delta t) = Z_j \cap \text{more than one jump in } (0, \delta t]) \mid \xi(0) = Z_i\} = o(\delta t). \tag{A7.106}$$

From this,

$$P_{ij}(\delta t) = \rho_{ij}\, \delta t + o(\delta t), \qquad j \neq i$$

and

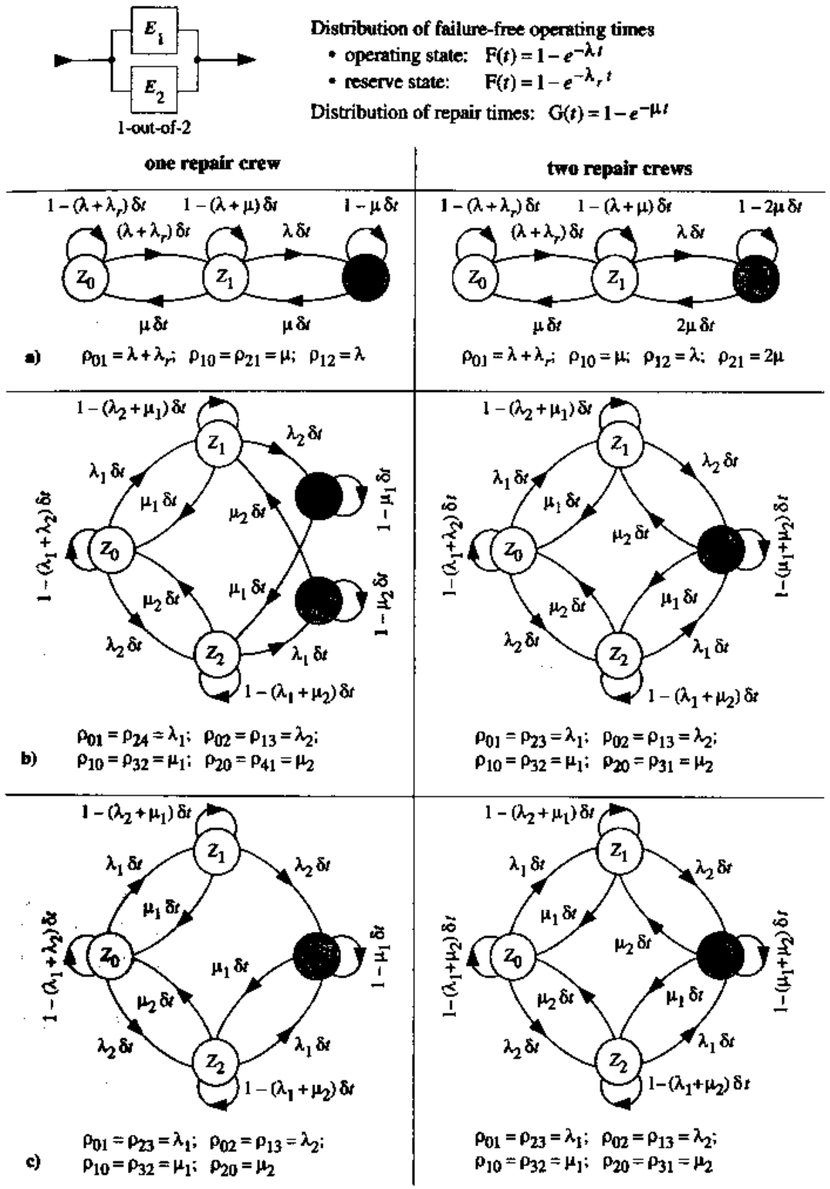
$$P_{ii}(\delta t) = 1 - \rho_i\, \delta t + o(\delta t),$$

as with Eq. (A7.88). Although for $\delta t \to 0$ it holds that $P_{ij}(\delta t) = Q_{ij}(\delta t)$, the meanings of $P_{ij}(\delta t)$ as in Eq. (A7.79) or Eq. (A7.78) and $Q_{ij}(\delta t)$ as in Eq. (A7.95) or Eq. (A7.144) are *basically different*; with $Q_{ij}(x)$, Z_j is the next state visited after Z_i, which is not the case for $P_{ij}(x)$.

Examples A7.3 to A7.5 give the diagram of transition probabilities in $(t + \delta t]$ for some typical structures for reliability applications. The states in which the system is down are hatched on the diagrams. In state Z_0 all elements are up (operating or in the reserve state).

Example A7.3

Figure A7.4 shows several cases of a *1-out-of-2 redundancy*. The difference with respect to the number of repair crews appears when leaving the states Z_2 and Z_3. Cases b) and c) are identical when two repair crews are available.



Distribution of failure-free operating times

- operating state: $F(t)=1-e^{-\lambda t}$
- reserve state: $F(t)=1-e^{-\lambda_r t}$

Distribution of repair times: $G(t)=1-e^{-\mu t}$

a) one repair crew: $\rho_{01}=\lambda+\lambda_r;\ \rho_{10}=\rho_{21}=\mu;\ \rho_{12}=\lambda$ — two repair crews: $\rho_{01}=\lambda+\lambda_r;\ \rho_{10}=\mu;\ \rho_{12}=\lambda;\ \rho_{21}=2\mu$

b) one repair crew: $\rho_{01}=\rho_{24}=\lambda_1;\ \rho_{02}=\rho_{13}=\lambda_2;\ \rho_{10}=\rho_{32}=\mu_1;\ \rho_{20}=\rho_{41}=\mu_2$ — two repair crews: $\rho_{01}=\rho_{23}=\lambda_1;\ \rho_{02}=\rho_{13}=\lambda_2;\ \rho_{10}=\rho_{32}=\mu_1;\ \rho_{20}=\rho_{31}=\mu_2$

c) one repair crew: $\rho_{01}=\rho_{23}=\lambda_1;\ \rho_{02}=\rho_{13}=\lambda_2;\ \rho_{10}=\rho_{32}=\mu_1;\ \rho_{20}=\mu_2$ — two repair crews: $\rho_{01}=\rho_{23}=\lambda_1;\ \rho_{02}=\rho_{13}=\lambda_2;\ \rho_{10}=\rho_{32}=\mu_1;\ \rho_{20}=\rho_{31}=\mu_2$

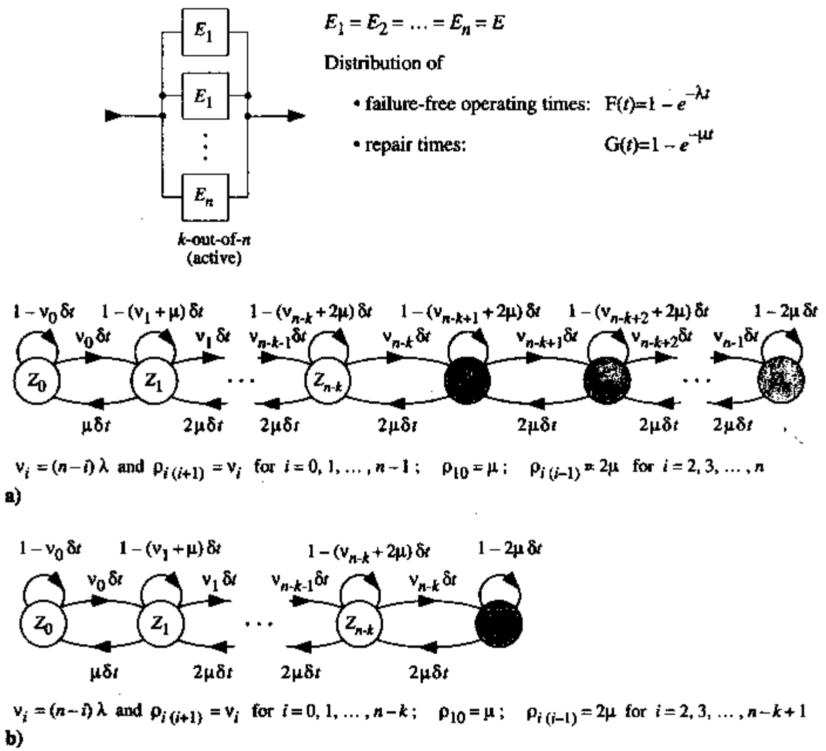
Figure A7.4 Diagram of transition probabilities in $(t, t+\delta t]$ for a repairable 1-out-of-2 redundancy (λ, λ_r = failure rates, μ = repair rate): **a)** Warm redundancy with $E_1 = E_2$ ($\lambda_r = \lambda \rightarrow$ active redundancy, $\lambda_r \equiv 0 \rightarrow$ standby redundancy); b) Active redundancy with $E_1 \neq E_2$; c) Active redundancy with $E_1 \neq E_2$ and repair priority on E_1 (t arbitrary, $\delta t \downarrow 0$, Markov process)

Example A7.4

Figure A7.5 shows two cases of a *k-out-of-n active redundancy* with two repair crews. In the first case, the system operates up to the failure of all elements (with reduced performance from state Z_{n-k+1}). In the second case no further failures can occur when the system is down.

Example A7.5

Figure A7.6 shows a *series/parallel structure* consisting of the series connection (in the reliability sense) of a 1-out-of-2 active redundancy, with elements E_2 and E_3 and a switching element E_1. The system has only one repair crew. Since one of the redundant elements E_2 or E_3 can be down without having a *system failure*, in cases a) and b) the repair of element E_1 is given *first priority*. This means that if a failure of E_1 occurs during a repair of E_2 or E_3, the repair is stopped and E_1 will be repaired. In cases c) and d) the *repair priority* on E_1 has been dropped.



$E_1 = E_2 = \ldots = E_n = E$

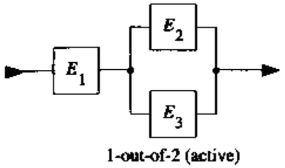
Distribution of

- failure-free operating times: $F(t)=1-e^{-\lambda t}$
- repair times: $G(t)=1-e^{-\mu t}$

a) $\nu_i=(n-i)\lambda$ and $\rho_{i\,(i+1)}=\nu_i$ for $i=0, 1, \ldots, n-1$; $\rho_{10}=\mu$; $\rho_{i\,(i-1)}=2\mu$ for $i=2, 3, \ldots, n$

b) $\nu_i=(n-i)\lambda$ and $\rho_{i\,(i+1)}=\nu_i$ for $i=0, 1, \ldots, n-k$; $\rho_{10}=\mu$; $\rho_{i\,(i-1)}=2\mu$ for $i=2, 3, \ldots, n-k+1$

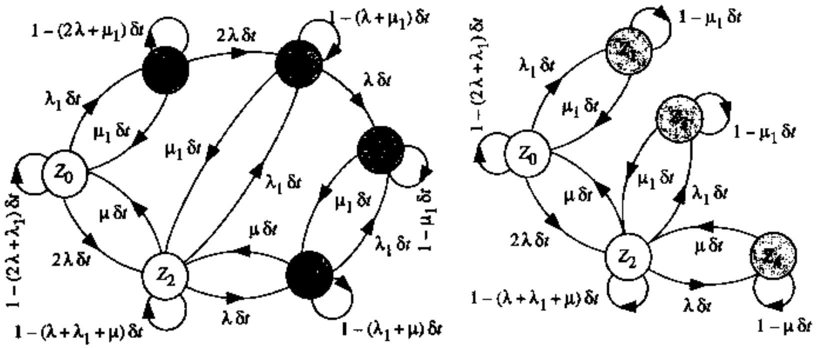
Figure A7.5 Diagram of transition probabilities in $(t, t+\delta t]$ for a repairable k-out-of-n redundancy with *two* repair crews (λ = failure rate, μ = repair rate): **a)** The system operates up to the failure of the last element; **b)** No further failures at system down (t arbitrary, $\delta t \downarrow 0$, Markov process; in a k-out-of-n redundancy the system is up if at least k elements are operating)



$E_2 = E_3 = E$

Distribution of

- failure-free times: $F(t) = 1 - e^{-\lambda t}$ for E, $F(t) = 1 - e^{-\lambda_1 t}$ for E_1
- repair times: $G(t) = 1 - e^{-\mu t}$ for E, $G(t) = 1 - e^{-\mu_1 t}$ for E_1

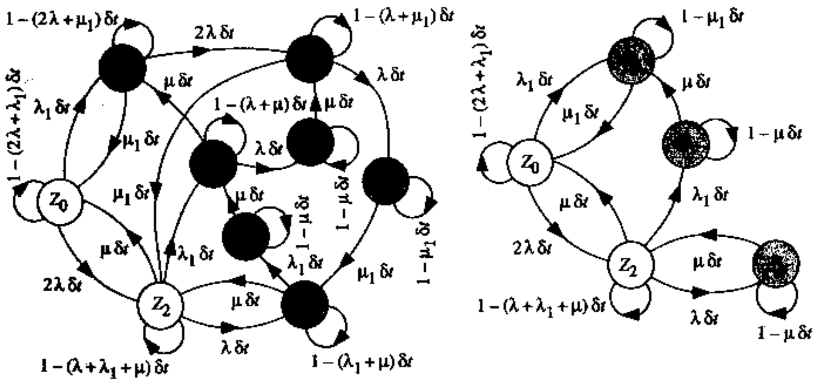


$\rho_{01} = \rho_{25} = \rho_{46} = \lambda_1$; $\rho_{02} = \rho_{15} = 2\lambda$; $\rho_{24} = \lambda$;
$\rho_{10} = \rho_{52} = \rho_{64} = \mu_1$; $\rho_{20} = \rho_{42} = \mu$; $\rho_{56} = \lambda$

$\rho_{01} = \rho_{23} = \lambda_1$; $\rho_{02} = 2\lambda$; $\rho_{24} = \lambda$;
$\rho_{10} = \rho_{32} = \mu_1$; $\rho_{20} = \rho_{42} = \mu$

a) Repair priority on E_1

b) As a), but no further failures at syst. down



$\rho_{01} = \rho_{23} = \rho_{47} = \lambda_1$; $\rho_{02} = \rho_{15} = 2\lambda$; $\rho_{10} = \rho_{52} = \rho_{64} = \mu_1$;
$\rho_{20} = \rho_{31} = \rho_{42} = \rho_{73} = \rho_{85} = \mu$; $\rho_{24} = \rho_{38} = \rho_{56} = \lambda$

$\rho_{01} = \rho_{23} = \lambda_1$; $\rho_{02} = 2\lambda$; $\rho_{24} = \lambda$;
$\rho_{10} = \mu_1$; $\rho_{20} = \rho_{31} = \rho_{42} = \mu$

c) No repair priority (first-in first-out)

d) As c), but no further failures at syst. down

Figure A7.6 Diagram of transition probabilities in $(t, t+\delta t]$ for a repairable *series parallel structure* with $E_2 = E_3 = E$ and *one* repair crew: **a)** Repair priority on E_1 and the system operates up to the failure of the last element; **b)** Repair priority on E_1 and at system failure no further failures can occur; **c)** and **d)** as a) and b), respectively, but without repair priority on E_1 (t arbitrary, $\delta t \downarrow 0$, Markov process)

A7.5.3 State Probabilities and Sojourn Times in a Given Class of States

In reliability theory, two important quantities are the state probabilities and the distribution function of the sojourn times in the set of system up states. The state probabilities allow calculation of the *point availability*. The *reliability function* can be obtained from the distribution function of the sojourn time in the set of system up states. Furthermore, a combination of these quantities allows for time-homogeneous Markov processes a simple calculation of the *interval reliability*.

It is useful in such an analysis to subdivide the system state space into two complementary sets U and $\overline{U}$

U = set of the system up states (up states at system level)

$\overline{U}$ = set of the system down states (down states at system level). (A7.107)

Calculation of state probabilities and of sojourn times can be carried out for Markov processes using the method of differential equations or of integral equations.

A7.5.3.1 Method of Differential Equations

The *method of differential equations* is the classical one used in investigating Markov processes. It is based on the *diagram of transition probabilities in* $(t, t+\delta t]$. Consider a time-homogeneous Markov process $\xi(t)$ with arbitrary initial probabilities $P_i(0) = \Pr\{\xi(0) = Z_i\}$ and transition rates ρ_{ij} and ρ_i. The state probabilities defined by Eq. (A7.83)

$$P_j(t) = \Pr\{\xi(t) = Z_j\}, \qquad j = 0, \ldots, m,$$

satisfy the system of differential equations

$$\dot{P}_j(t) = -\rho_j P_j(t) + \sum_{\substack{i=0 \\ i \neq j}}^{m} P_i(t)\rho_{ij}, \qquad \rho_j = \sum_{\substack{i=0 \\ i \neq j}}^{m} \rho_{ji}, \qquad j = 0, \ldots, m. \qquad \text{(A7.108)}$$

The proof of Eq. (A7.108) is the same as that of Eq. (A7.91). The *point availability* $PA_S(t)$, for arbitrary initial conditions at $t = 0$, follows then from

$$PA_S(t) = \Pr\{\xi(t) \in U\} = \sum_{Z_j \in U} P_j(t). \qquad \text{(A7.109)}$$

In reliability analysis, particular *initial conditions* (probabilities) at $t = 0$ are often of interest. Assume

$$P_i(0) = 1 \quad \text{and} \quad P_j(0) = 0 \quad \text{for } j \neq i, \qquad \text{(A7.110)}$$

i.e. that the system is in Z_i at $t=0$ (usually in state Z_0 denoting "all elements are *up*"). In this case, the state probabilities $\mathrm{P}_j(t)$ are the *transition probabilities* $\mathrm{P}_{ij}(t)$ defined by Eqs. (A7.78) and (A7.79) and can be obtained as

$$\mathrm{P}_{ij}(t) \equiv \mathrm{P}_j(t) \tag{A7.111}$$

with $\mathrm{P}_j(t)$ as the solution of Eq. (A7.108) with initial conditions as in Eq. (A7.110), or of Eq. (A7.92). The *point availability*, now designated with $\mathrm{PA}_{Si}(t)$, is then given by

$$\mathrm{PA}_{Si}(t) = \Pr\{\xi(t) \in U \mid \xi(0) = Z_i\} = \sum_{Z_j \in U} \mathrm{P}_{ij}(t), \qquad i = 0, \ldots, m. \tag{A7.112}$$

$\mathrm{PA}_{S_i}(t)$ is the probability that the system is in one of the up states at t, given it was in Z_i at $t=0$. Example A 7.6 illustrate calculation of the point-availability for a 1-out-of-2 active redundancy.

Example A7.6

Assume a *1-out-of-2 active redundancy*, consisting of 2 identical elements $E_1 = E_2 = E$ with constant failure rate λ and repair rate μ, and only one repair crew. Determine the state probabilities of the involved Markov process (E_1 and E_2 are new at $t=0$).

Solution

Figure A7.7 shows the diagram of transition probabilities in $(t, t+\delta t]$ for the investigation of the point availability. Because of the *memoryless property* of the Markov Process, Fig A7.7 and Eq. (A7.83) lead to (by omitting the terms in $o(\delta t)$, as per Eq. (A7.89))

$$\mathrm{P}_0(t+\delta t) = \mathrm{P}_0(t)(1-2\lambda\,\delta t) + \mathrm{P}_1(t)\mu\,\delta t$$

$$\mathrm{P}_1(t+\delta t) = \mathrm{P}_1(t)(1-(\lambda+\mu)\delta t) + \mathrm{P}_0(t)2\lambda\,\delta t + \mathrm{P}_2(t)\mu\,\delta t$$

$$\mathrm{P}_2(t+\delta t) = \mathrm{P}_2(t)(1-\mu\,\delta t) + \mathrm{P}_1(t)\lambda\,\delta t,$$

and then, as $\delta t \downarrow 0$,

$$\dot{\mathrm{P}}_0(t) = -2\lambda\,\mathrm{P}_0(t) + \mu\,\mathrm{P}_1(t)$$

$$\dot{\mathrm{P}}_1(t) = -(\lambda+\mu)\mathrm{P}_1(t) + 2\lambda\,\mathrm{P}_0(t) + \mu\,\mathrm{P}_2(t)$$

$$\dot{\mathrm{P}}_2(t) = -\mu\,\mathrm{P}_2(t) + \lambda\,\mathrm{P}_1(t). \tag{A7.113}$$

Equation (A7.113) also follows from Eq. (A7.108) with the ρ_{ij} from Fig. A7.7. The solution of Eq. (A7.113) with given initial conditions at $t=0$, e.g. $\mathrm{P}_0(0)=1$, $\mathrm{P}_1(0)=\mathrm{P}_2(0)=0$, leads to state probabilities $\mathrm{P}_0(t)$, $\mathrm{P}_1(t)$, and $\mathrm{P}_2(t)$, and then to the point availability according to Eqs. (A7.111) and (A7.112) with $i=0$ (see also Example A7.9).

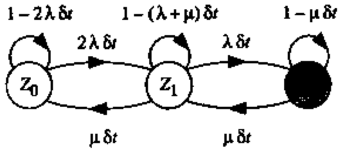


Figure A7.7 Diagram of the transition probabilities in $(t, t+\delta t]$ for a 1-out-of-2 active redundancy with $E_1 = E_2 = E$, one repair crew, calculation of the point availability (t arbitr., $\delta t \downarrow 0$, Markov proc.)

A further important quantity for reliability theory is the *reliability function* $\mathrm{R}_S(t)$, i.e. the probability of no system failure in $(0, t]$. $\mathrm{R}_S(t)$ can be calculated using the method of differential equations if all states in $\overline{U}$ are declared to be *absorbing states*. This means that the process will never leave Z_k if it jumps into a state $Z_k \in \overline{U}$. It is not difficult to see that in this case, the events

first system failure occurs before t

and

system is in one of the states $\overline{U}$ at t

are equivalent, so that the sum of the probabilities to be in one of the states in U is the required reliability function, i.e. the probability that up to the time t the process has never left the set of up states U. To make this analysis rigorous, consider the *modified Markov process* $\xi'(t)$ with transition probabilities $\mathrm{P}'_{ij}(t)$ and transition rates

$$\rho'_{ij} = \rho_{ij} \text{ if } Z_i \in U, \qquad \rho'_{ij} = 0 \text{ if } Z_i \in \overline{U}, \qquad \rho'_i = \sum_{\substack{j=0 \\ j \neq i}}^{m} \rho'_{ij}. \tag{A7.114}$$

The state probabilities $\mathrm{P}'_j(t)$ of $\xi'(t)$ satisfy the following system of differential equations

$$\dot{\mathrm{P}}'_j(t) = -\rho'_j\,\mathrm{P}'_j(t) + \sum_{\substack{i=0 \\ i \neq j}}^{m} \mathrm{P}'_i(t)\rho'_{ij}, \qquad \rho'_j = \sum_{\substack{i=0 \\ i \neq j}}^{m} \rho'_{ji}, \qquad j = 0, \ldots, m. \tag{A7.115}$$

Assuming as initial conditions $\mathrm{P}'_i(0)=1$ and $\mathrm{P}'_i(0)=0$ for $j \neq i$ (with $Z_i \in U$), the solution of Eq. (A7.115) leads to the state probabilities $\mathrm{P}'_j(t)$ and from these to the transition probabilities

$$\mathrm{P}'_{ij}(t) \equiv \mathrm{P}'_j(t). \tag{A7.116}$$

The *reliability function* $\mathrm{R}_{Si}(t)$ is then given by

$$\mathrm{R}_{Si}(t) = \Pr\{\xi(x) \in U \text{ for } 0 < x \leq t \mid \xi(0) = Z_i\} = \sum_{Z_j \in U} \mathrm{P}'_{ij}(t), \qquad Z_i \in U. \tag{A7.117}$$

The dashed probalibities ($\mathrm{P}'(t)$) are reserved for the computation of the reliability, when using the method of differential equations. This should avoid confusion with the corresponding quantities for the point availability. Example A7.7 illustrates the calculation of the reliability function for a 1-out-of-2 active redundancy.

Example A7.7

Determine the reliability function for the same case as in Example A7.6, i.e. the probability that the system has not left the states Z_0 and Z_1 up to time t.

Solution

The diagram of transition probabilities in $(t, t+\delta t]$ of Fig. A7.7 is modified as in Fig. A7.8 by making the down state Z_2 *absorbing*. For the state probabilities it follows that

$$\dot{\mathrm{P}}_0'(t) = -2\lambda\,\mathrm{P}_0'(t) + \mu\,\mathrm{P}_1'(t)$$

$$\dot{\mathrm{P}}_1'(t) = -(\lambda+\mu)\mathrm{P}_1'(t) + 2\lambda\,\mathrm{P}_0'(t)$$

$$\dot{\mathrm{P}}_2'(t) = -\lambda\,\mathrm{P}_1'(t). \tag{A7.118}$$

The solution of Eq. (A7.118) with the given initial conditions at $t=0$ ($\mathrm{P}_0'(0)=1$, $\mathrm{P}_1'(0)=\mathrm{P}_2'(0)=0$) leads to the state probabilities $\mathrm{P}_0'(t)$, $\mathrm{P}_1'(t)$ and $\mathrm{P}_2'(t)$, and then to the transition probabilities and to the reliability function according to Eqs. (A7.116) and (A7.117), respectively (the dashed state probabilities should avoid confusion with the solution given by Eq. (A7.113)).

Equations (A7.112) and (A7.117) can be combined to determine the probability that the process is in an up state (set U) at t and does not leave the set U in the time interval $[t, t+\theta]$, given $\xi(0)=Z_i$. This quantity is the *interval reliability* $\mathrm{IR}_{Si}(t,t+\theta)$. Due to the *memoryless property* of the Markov process,

$$\mathrm{IR}_{Si}(t,t+\theta) = \Pr\{\xi(x)\in U \text{ for } t\le x\le t+\theta \mid \xi(0)=Z_i\} = \sum_{Z_j\in U}\mathrm{P}_{ij}(t)\mathrm{R}_{Sj}(\theta),$$

$$i=0,\dots,m, \tag{A7.119}$$

with $\mathrm{P}_{ij}(t)$as given in Eq. (A7.111).

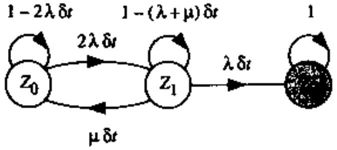


Figure A7.8 Diagram of the transition probabilities in $(t, t+\delta t]$ for a 1-out-of-2 active redundancy with $E_1=E_2=E$, one repair crew (for this case not mandatory), calculation of the reliability function (t arbitrary, $\delta t\downarrow 0$, Markov process)

A7.5.3.2 Method of Integral Equations

The *method of integral equations* is based on the representation of the Markov process $\xi(t)$ as a pure jump process by means of ξ_n and η_n as introduced in Appendix A7.5.2 (Eq. (A7.95)). From the *memoryless property* it uses *only* the fact that jump points (in a new state) are *regeneration points* of $\xi(t)$.

The *transition probabilities* $\mathrm{P}_{ij}(t) = \Pr\{\xi(t)=Z_j \mid \xi(0)=Z_i\}$ can be obtained by solving the following system of integral equations

$$\mathrm{P}_{ij}(t) = \delta_{ij}\,e^{-\rho_i t} + \sum_{\substack{k=0\\k\ne i}}^{m}\int_0^t \rho_{ik}\,e^{-\rho_i x}\,\mathrm{P}_{kj}(t-x)\,dx, \qquad \rho_i=\sum_{\substack{j=0\\j\ne i}}^{m}\rho_{ij}, \qquad i,j\in\{0,\dots,m\}, \tag{A7.120}$$

with $\delta_{ij}=0$ for $j\ne i$ and $\delta_{ii}=1$. To prove Eq. (A7.120), consider that

$$\begin{aligned}\mathrm{P}_{ij}(t) &= \Pr\{(\xi(t)=Z_j \cap \text{no jumps in } (0,t]) \mid \xi(0)=Z_i\}\\ &\quad+\sum_{\substack{k=0\\k\ne i}}^{m}\Pr\{(\xi(t)=Z_j \cap \text{first jump in } (0,t] \text{ in } Z_k) \mid \xi(0)=Z_i\}\\ &= \Pr\{\xi(t)=Z_j\cap\eta_0>t \mid \xi(0)=Z_i\}\\ &\quad+\sum_{\substack{k=0\\k\ne i}}^{m}\Pr\{\xi(t)=Z_j\cap\eta_0\le t\cap\xi_1=Z_k \mid \xi(0)=Z_i\}.\end{aligned} \tag{A7.121}$$

The first term of Eq. (A7.121) only holds for $j=i$ and it gives the probability that the process will *not leave* the state Z_i ($e^{-\rho_i t}=\Pr\{\tau_{ij}>t \text{ for all } j\ne i\}$ according to the interpretation given by Eqs. (A7.99) – (A7.104)). The second term holds for any $j\ne i$, it gives the probability that the process will move first from Z_i to Z_k and take into account that the occurrence of Z_k is a *regeneration point*. According to Eq. (A7.95), $\Pr\{\xi_1=Z_k\cap\eta_0\le x \mid \xi(0)=Z_i\}=Q_{ik}(x)=p_{ik}(1-e^{-\rho_i x})$ and $\Pr\{\xi(t)=Z_j \mid (\xi_0=Z_i\cap\eta_0=x\cap\xi_1=Z_k)\}=P_{kj}(t-x)$. Equation (A7.120) then follows from the *theorem of total probability* (Eq. (A6.17)).

In the same way as for Eq. (A.121), it can be shown that the *reliability function* $\mathrm{R}_{Si}(t)$, as defined in Eq. (A7.117), satisfies the following system of integral equations

$$\mathrm{R}_{Si}(t) = e^{-\rho_i t} + \sum_{\substack{Z_j\in U\\j\ne i}}\int_0^t \rho_{ij}\,e^{-\rho_i x}\,\mathrm{R}_{Sj}(t-x)\,dx, \qquad \rho_i=\sum_{\substack{j=0\\j\ne i}}^{m}\rho_{ij}, \qquad Z_i\in U. \tag{A7.122}$$

Point availability $\mathrm{PA}_{Si}(t)$ and *interval reliability* $\mathrm{IR}_{Si}(t,t+\theta)$ are again given by Eqs. (A7.112) and (A7.119), with $\mathrm{P}_{ij}(t)$ form Eq. (A7.120).

The systems of integral equations (A7.120) and (A7.122) can be solved using *Laplace transforms*. Referring to Appendix A9.7,

$$\bar{\mathrm{P}}_{ij}(s)=\frac{\delta_{ij}}{s+\rho_i}+\sum_{\substack{k=0\\k\neq i}}^{m}\frac{\rho_{ik}}{s+\rho_i}\bar{\mathrm{P}}_{kj}(s), \qquad \rho_i=\sum_{\substack{j=0\\j\neq i}}^{m}\rho_{ij}, \quad i,j\in\{0,\ldots,m\}, \tag{A7.123}$$

and

$$\tilde{\mathrm{R}}_{Si}(s)=\frac{1}{s+\rho_i}+\sum_{\substack{Z_j\in U\\j\neq i}}\frac{\rho_{ik}}{s+\rho_i}\tilde{\mathrm{R}}_{Sj}(s), \qquad \rho_i=\sum_{\substack{j=0\\j\neq i}}^{m}\rho_{ij}, \quad Z_i\in U. \tag{A7.124}$$

A direct advantage of the method based on integral equations appears in the computation of the mean of the sojourn time in the up states. Denoting by $MTTF_{S_i}$ the *system mean time to failure*, provided the system is in state $Z_i\in U$ at $t=0$, Eq. (A6.38) leads to

$$MTTF_{Si}=\int_0^\infty \mathrm{R}_{Si}(t)\,dt=\tilde{\mathrm{R}}_{Si}(0). \tag{A7.125}$$

Thus, according to Eq. (A7.124), $MTTF_{S_i}$ satisfies the following system of algebraic equations

$$MTTF_{Si}=\frac{1}{\rho_i}+\sum_{\substack{Z_j\in U\\j\neq i}}\frac{\rho_{ij}}{\rho_i}MTTF_{Sj}, \qquad \rho_i=\sum_{\substack{j=0\\j\neq i}}\rho_{ij}, \quad Z_i\in U. \tag{A7.126}$$

A7.5.3.3 Stationary State and Asymptotic Behavior

The determination of time-dependent state probabilities and the point availability of a system whose elements have constant failure and repair rates is still possible using differential or integral equations. However, it can become time-consuming. The situation is easier in which the state probabilities are independent of time, i.e. when the process involved is *stationary* (the system of differential equations reduces to a system of algebraic equations):

> *A time-homogeneous Markov process* $\xi(t)$ *with states* $Z_0,\ldots,Z_m$ *is stationary, if its state probabilities* $\mathrm{P}_i(t)=\Pr\{\xi(t)=Z_i\}$, $i=0,\ldots,m$ *do not depend on t.*

This can be seen from the following relationship

$$\begin{aligned}\Pr\{\xi(t_1)=Z_{i_1}\cap\ldots\cap\xi(t_n)=Z_{i_n}\}\\ =\Pr\{\xi(t_1)=Z_{i_1}\}\,\mathrm{P}_{i_1i_2}(t_2-t_1)\ldots\mathrm{P}_{i_{n-1}i_n}(t_n-t_{n-1})\end{aligned}$$

which, according to the Markov property (Eq. (A7.77)) must be valid for arbitrary $t_1<\ldots<t_n$ and $i_1,\ldots,i_n\in\{0,\ldots,m\}$. For any $a>0$ this leads to

$$\Pr\{\xi(t_1)=Z_{i_1}\cap\ldots\cap\xi(t_n)=Z_{i_n}\}=\Pr\{\xi(t_1+a)=Z_{i_1}\cap\ldots\cap\xi(t_n+a)=Z_{i_n}\}.$$

From $\mathrm{P}_i(t+a)=\mathrm{P}_i(t)$ it also follows that $\mathrm{P}_i(t)=\mathrm{P}_i(0)=P_i$ and in particular $\dot{\mathrm{P}}_i(t)=0$. Consequently, the process $\xi(t)$ is *stationary* if and only if its initial distribution $P_i=\mathrm{P}_i(0)=\Pr\{\xi(0)=Z_i\}$, $i=0,\ldots,m$, satisfies for any $t>0$ the system

$$\rho_j P_j=\sum_{\substack{i=0\\i\neq j}}^{m}P_i\rho_{ij}, \qquad \text{with} \quad P_j\geq 0, \quad \sum_{j=0}^{m}P_j=1, \quad \rho_i=\sum_{\substack{j=0\\j\neq i}}^{m}\rho_{ij}, \quad j=0,\ldots,m. \tag{A7.127}$$

Every solution of Eq. (A7.127) with $P_j\geq 0$, $j=0,\ldots,m$, is a *stationary initial distribution* of the Markov process involved.

A Markov process is *irreducible* if for every pair $i,j\in\{0,\ldots,m\}$ there exists a t such that $\mathrm{P}_{ij}(t)>0$, i.e. every state can be reached from every other state (it can be shown that if $\mathrm{P}_{ij}(t_0)>0$ for some $t_0>0$, then $\mathrm{P}_{ij}(t)>0$ for any $t>0$). A Markov process is irreducible if and only if its *embedded Markov chain* is irreducible. For an irreducible Markov process, there exist quantities $P_j>0$, $j=0,\ldots,m$, with $P_0+\ldots+P_m=1$, such that *independently* of the initial condition $\mathrm{P}_i(0)$ the following holds (Markov theorem, see e.g. [A6.6 (Vol. I)])

$$\lim_{t\to\infty}\mathrm{P}_j(t)=P_j>0, \qquad j=0,\ldots,m. \tag{A7.128}$$

For any $i=0,\ldots,m$ it follows then that

$$\lim_{t\to\infty}\mathrm{P}_{ij}(t)=P_j>0, \qquad j=0,\ldots,m. \tag{A7.129}$$

The set of values $P_0,\ldots,P_m$ from Eq. (A7.128) is the *limiting distribution* of the Markov process. From Eqs. (A7.74) and (A7.129) it follows that for an *irreducible Markov process* the limiting distribution is the only stationary distribution, i.e. the only solution of Eq. (A7.127) with $P_j>0$, $j=0,\ldots,m$.

The *asymptotic and stationary value* (*steady-state value*) of the point availability PA_S is then given by

$$\lim_{t\to\infty}\mathrm{PA}_{Si}(t)=PA_S=\sum_{Z_j\in U}P_j \qquad i=0,\ldots,m. \tag{A7.130}$$

If K is a subset of $\{Z_0,\ldots,Z_m\}$, the Markov process is irreducible, and $P_0,\ldots,P_m$ are the limiting probabilities obtained from Eq. (A7.127) then,

$$\Pr\{\lim_{t\to\infty} \frac{\text{total sojourn time in states } Z_j \in K \text{ in } (0,\ t]}{t} = \sum_{Z_j \in K} P_j\} = 1 \qquad \text{(A7.131)}$$

irrespective of the initial distribution $\mathrm{P}_i(0)$, $i = 0, \ldots, m$. From Eq. (A7.131) it follows that

$$\Pr\{\lim_{t\to\infty} \frac{\text{total operating time in } (0,t]}{t} = \sum_{Z_j \in U} P_j = PA_S\} = 1.$$

Defining the *average availability* of the system as

$$\mathrm{AA}_{Si}(t) = \frac{1}{t}\mathrm{E}[\text{total operating time in } (0,\ t] \mid \xi(0) = Z_i] = \frac{1}{t}\int_0^t \mathrm{PA}_{Si}(x)\,dx. \qquad \text{(A7.132)}$$

Equations (A7.130) and (A7.132) lead to (for any $Z_i \in U$)

$$\lim_{t\to\infty} \mathrm{AA}_{Si}(t) = AA_S = PA_S = \sum_{Z_j \in U} P_j. \qquad \text{(A7.133)}$$

Expressions of the form $\sum k P_k$ can be used to calculate the expected number of elements in repair or in standby, of repair crews used, etc., as well as for *cost optimizations*.

A7.5.4 Birth and Death Process

An important time-homogeneous Markov process used in reliability theory, for example to investigate a *k-out-of-n redundancy* with identical elements and constant failure and repair rates, is the birth and death process. The diagram of transition probabilities in $(t, t+\delta t]$ is a generalization of those in Fig. A7.5, and is given in Fig. A7.9. ν_i and θ_i are the transition rates from state Z_i to Z_{i+1} and from Z_i to Z_{i-1}, respectively. Transitions outside neighboring states can only occur in $(t, t+\delta t]$ with probability $o(\delta t)$. The system of differential equations describing the *birth and death process* given in Fig. A7.9 is (Eq. (A7.108))

$$\dot{\mathrm{P}}_j(t) = -(\nu_j + \theta_j)\mathrm{P}_j(t) + \nu_{j-1}\mathrm{P}_{j-1}(t) + \theta_{j+1}\mathrm{P}_{j+1}(t)$$
$$\text{with } \theta_0 = \nu_{-1} = \nu_n = \theta_{n+1} = 0, \qquad j = 0, \ldots, n. \qquad \text{(A7.134)}$$

The conditions $\nu_j > 0$ ($j = 0, \ldots, n-1$) and $\theta_j > 0$ ($j = 1, \ldots, n$) are sufficient for the existence of the *limiting probabilities*

$$\lim_{t\to\infty} \mathrm{P}_j(t) = P_j, \qquad \text{with} \quad \mathrm{P}_j > 0 \quad \text{and} \quad \sum_{j=0}^{n} P_j = 1. \qquad \text{(A7.135)}$$

It is easy to show (Example A7.8) that the limiting probabilities P_j, $j = 0, \ldots, n$, are given by

$$P_j = \pi_j P_0 = \frac{\pi_j}{\sum_{i=0}^{n} \pi_i}, \qquad \text{with} \quad \pi_i = \frac{\nu_0 \cdots \nu_{i-1}}{\theta_1 \ldots \theta_i} \quad \text{and} \quad \pi_0 = 1. \qquad \text{(A7.136)}$$

From Eq. (A7.136) one recognizes that $P_k \nu_k = P_{k+1}\theta_{k+1}$ ($k = 0, \ldots, n-1$) holds, expressing the basic property of a *stady-state transition diagram*.

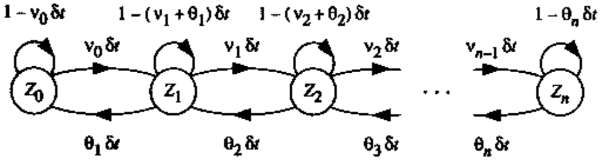


Figure A7.9 Diagram of transition probabilities in $(t, t+dt]$ for a birth and death process with $n+1$ states (t arbitrary, $\delta t \downarrow 0$, Markpv process)

Example A7.8

Assuming Eq. (A7.135) prove Eq. (A7.136).

Solution

Considering Eqs. (A7.134) and (A7.135), P_j are the solution of the following system of algebraic equations

$$\begin{aligned} 0 &= -\nu_0 P_0 + \theta_1 P_1 \\ &\vdots \\ 0 &= -(\nu_j + \theta_j)P_j + \nu_{j-1}P_{j-1} + \theta_{j+1}P_{j+1}, \quad j = 2, \ldots, n-1, \\ &\vdots \\ 0 &= -\theta_n P_n + \nu_{n-1}P_{n-1}. \end{aligned} \qquad \text{(7.137)}$$

From the first equation it follows that

$$P_1 = P_0 \nu_0 / \theta_1.$$

With this value for P_1, the second equation ($i = 1$) leads to

$$P_2 = \frac{\nu_1 + \theta_1}{\theta_2} P_1 - \frac{\nu_0}{\theta_2} P_0 = (\frac{\nu_1 + \theta_1}{\theta_2} \cdot \frac{\nu_0}{\theta_1} - \frac{\nu_0}{\theta_2}) P_0 = \frac{\nu_0 \nu_1}{\theta_1 \theta_2} P_0.$$

Recursively one obtains

$$P_j = \frac{\nu_0 \ldots \nu_{j-1}}{\theta_1 \ldots \theta_j} P_0 = \Pi_j P_0, \qquad j = 0, \ldots, n, \quad \Pi_0 = 1.$$

Taking into account that $P_0 + \ldots + P_n = 1$, P_0 follows and then Eq. (A7.136).

The values of P_j given by Eq. (A7.136) can be used in Eq. (A7.130) to determine the steady-state value of the *point availability*. The system *mean time to failure* follows from Eq. (A7.126) with $\rho_i = \nu_i + \theta_i$, $\rho_{ij} = \nu_i$ for $j = i+1$, and $\rho_{ij} = \theta_i$ for $j = i-1$, provided that the state Z_{i+1} still belong to U (if not, $\rho_{ij} = 0$ for $j = i-1$). Examples A7.9 and A7.10 are applications of the birth and death process.

Example A7.9

For the 1-out-of-2 active redundancy with only one repair crew (Examples A7.6 and A7.7) i.e. for $\nu_0 = 2\lambda$, $\nu_1 = \lambda$, $\theta_1 = \theta_2 = \mu$, $U = \{Z_0, Z_1\}$ and $\overline{U} = \{Z_2\}$ determine the asymptotic value of the point availability PA_S and the system's mean time to failure $MTTF_{S0}$.

Solution

The asymptotic value of point availability is given by Eqs. (A7.130) and (A7.136)

$$PA_s = P_0 + P_1 = \frac{1 + \frac{2\lambda}{\mu}}{1 + \frac{2\lambda}{\mu} + \frac{2\lambda^2}{\mu^2}} = \frac{\mu(2\lambda + \mu)}{2\lambda(\lambda + \mu) + \mu^2}. \qquad \text{(A7.138)}$$

For the system's mean time to failure it follows from Eq. (A7.126) that (considering $\rho_0 = \nu_0$, $\rho_{01} = \nu_0$, $\rho_1 = \nu_1 + \theta_1$, and $\rho_{10} = \theta_1$)

$$MTTF_{S0} = \frac{1}{2\lambda} + MTTF_{S1}$$

$$MTTF_{S1} = \frac{1}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} MTTF_{S0},$$

and thus

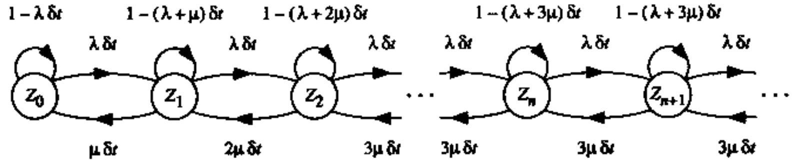
$$MTTF_{S0} = \frac{3\lambda + \mu}{2\lambda^2}. \qquad \text{(A7.139)}$$

Example A7.10

A computer system consists of 3 identical CPUs. Jobs arrive independently and the arrival times form a *Poisson process* with *intensity* λ. The duration of each individual job is distributed exponentially with parameter μ. All jobs have the same memory requirements D. Determine for $\lambda = 2\mu$ the minimum size n of the memory required in units of D, so that in the *steady-state* a new job can immediately find storage space with a probability γ of at least 95%. When overflow occurs, jobs are queued.

Solution

The problem can be solved using the following birth and death process



In state Z_i, exactly i memory units are occupied. n is the smallest integer such that in the steady-state, $P_0 + \ldots + P_{n-1} = \gamma \geq 0.95$ (if the assumption were made that jobs are lost if overflow occurs, then the process would stop at state Z_n). For steady-state, Eq. (A7.127) yields

$$\begin{aligned}
0 &= -\lambda P_0 + \mu P_1 \\
0 &= \lambda P_0 - (\lambda + \mu) P_1 + 2\mu P_2 \\
0 &= \lambda P_1 - (\lambda + 2\mu) P_2 + 3\mu P_3 \\
0 &= \lambda P_2 - (\lambda + 3\mu) P_3 + 3\mu P_4 \\
&\vdots \\
0 &= \lambda P_i - (\lambda + 3\mu) P_{i+1} + 3\mu P_{i+2}, \qquad i > 2.
\end{aligned} \qquad \text{(A7.140)}$$

The solution leads to

$$P_1 = \frac{\lambda}{\mu} P_0 \qquad \text{and} \qquad P_i = \frac{\lambda^i P_0}{2 \cdot 3^{i-2} \mu^i} = \frac{9}{2} (\frac{\lambda / \mu}{3})^i P_0 \quad \text{for} \quad i \geq 2.$$

Assuming $\lim_{n \to \infty} \sum_{i=0}^{n} P_i = 1$ and considering $\frac{\lambda / \mu}{3} < 1$ it follows that

$$P_0 [1 + \frac{\lambda}{\mu} + \sum_{i=2}^{\infty} \frac{9}{2} (\frac{\lambda / \mu}{3})^i] = P_0 [1 + \frac{\lambda}{\mu} + \frac{3(\lambda / \mu)^2}{2(3 - \lambda / \mu)}] = 1,$$

from which

$$P_0 = \frac{2(3 - \lambda / \mu)}{6 + \frac{4\lambda}{\mu} + (\lambda / \mu)^2}.$$

The size of the memory n can now be determined from

$$\frac{2(3 - \lambda / \mu)}{6 + \frac{4\lambda}{\mu} + (\lambda / \mu)^2} [1 + \frac{\lambda}{\mu} + \sum_{i=2}^{n-1} \frac{9}{2} (\frac{\lambda / \mu}{3})^i] \geq \gamma.$$

For $\lambda / \mu = 2$ and $\gamma = 0.95$, the smallest n satisfying the above equation is $n = 9$ ($P_0 = 1/9$, $P_1 = 2/9$, $P_i = 2^{i-1} / 3^i$ for $i \geq 2$).

As shown by Examples A7.9 and A7.10, reliability applications of birth and death processes identify the quantities ν_i as failure rates and θ_i as repair rates. In this case,

$$\nu_j << \theta_{j+1}, \qquad j = 0, \ldots, n-1,$$

with ν_j and θ_{j+1} as in Fig. A7.9. Assuming

$$\nu_j \leq r \theta_{j+1}, \qquad j = 0, \ldots, n-1, \qquad \text{(A7.141)}$$

the following relationships for the steady-state probability p_j can be obtained (Example A7.11)

$$P_j \ge \frac{1-r}{r(1-r^{n-i})} \sum_{i=j+1}^{n} P_i, \qquad 0<r<1, \quad j=0, \ldots, n-1, \quad n>j. \qquad \text{(A7.142)}$$

For $r \le 1/2$ it follows that

$$P_j \ge \sum_{i=j+1}^{n} P_i, \qquad j=0, \ldots, n-1. \qquad \text{(A7.143)}$$

Equation (A7.143) states that for $2\nu_j \le \theta_{j+1}$, the *steady-state probability* in a state Z_j of a birth and death process described by Fig. A7.9 is greater or equal the sum of the steady-state probabilities in all states following Z_j, $j=0, \ldots, n-1$. This relationship is useful in developing *approximate expressions* for system availability [2.51 (1992)].

Example A7.11

Assuming Eq.(A7.141), prove Eqs. (A7.142) and (A7.143).

Solution

Using Eq. (A7.136),

$$\frac{\sum_{i=j+1}^{n} P_i}{P_j} = \frac{\sum_{i=j+1}^{n} \pi_i}{\pi_j} = \sum_{i=j+1}^{n} \frac{\pi_i}{\pi_j} = \frac{\nu_j}{\theta_{j+1}} + \frac{\nu_j \nu_{j+1}}{\theta_{j+1}\theta_{j+2}} + \ldots + \frac{\nu_j \ldots \nu_{n-1}}{\theta_{j+1} \ldots \theta_n}.$$

Setting

$$\frac{\nu_i}{\theta_{i+1}} \le r, \qquad 0<r<1, \qquad i=j, j+1, \ldots, n-1,$$

it follows that

$$\frac{\sum_{i=j+1}^{n} P_i}{P_j} \le r + r^2 + \ldots + r^{n-j} = \frac{r(1-r^{n-j})}{1-r}$$

and thus Eq. (A7.142). Furthermore, for

$r \le 1/2$

it follows that

$$\frac{\sum_{i=j+1}^{n} P_i}{P_j} \le 1-(1/2)^{n-j} \le 1,$$

and hence Eq. (A7.143).

A7.6 Semi-Markov Processes with Finitely Many States

The description of Markov processes given in Appendix A7.5.2 allows a straightforward generalization to *semi-Markov processes*. In a semi-Markov process, the sequence of consecutively occurring states forms a time-homogeneous *Markov chain*, just as with Markov processes. However, the sojourn time in a given state Z_i is a positive random variable τ_{ij} whose distribution depends on Z_i and on the following state Z_j, but, in contrast to Markov processes, this sojourn time is *arbitrarily distributed*, i.e. not exponentially distributed as for Markov processes.

To define semi-Markov processes, let $\xi_0, \xi_1, \ldots$ be the sequence of consecutively occurring states, i.e. a sequence of random variables taking values in $\{Z_0, \ldots, Z_m\}$, and $\eta_0, \eta_1, \ldots$ the sojourn times between consecutive states, i.e. a sequence of positive random variables. A stochastic process $\xi(t)$ with the state space $\{Z_0, \ldots, Z_m\}$ is a *semi-Markov process* if for $n=1, 2, \ldots$, arbitrary i, j, $i_0, \ldots, i_{n-1} \in \{0, \ldots, m\}$ and arbitrary positive numbers $x_0, \ldots, x_{n-1}$,

$$\Pr\{(\xi_{n+1}=Z_j \cap \eta_n \le x) \,\big|\, (\xi_n=Z_i \cap \eta_{n-1}=x_{n-1} \cap \ldots \xi_1=Z_{i_1} \cap \eta_0=x_0 \cap \xi_0=Z_{i_0})\}$$
$$= \Pr\{(\xi_{n+1}=Z_j \cap \eta_n \le x) \,\big|\, \xi_n=Z_i\} = Q_{ij}(x). \qquad \text{(A7.144)}$$

The functions $Q_{ij}(x)$ in Eq. (A7.144), *defined only for* $j \ne i$, are the *semi-Markov transition probabilities* (see the remarks with Eqs. (A7.95) and (A7.106)). Setting

$$Q_{ij}(\infty) = p_{ij}, \qquad \text{(A7.145)}$$

and, for $p_{ij} > 0$,

$$F_{ij}(x) = \frac{Q_{ij}(x)}{p_{ij}}, \qquad \text{(A7.146)}$$

leads to

$$Q_{ij}(x) = p_{ij} F_{ij}(x), \qquad \text{(A7.147)}$$

with (Example A7.2)

$$p_{ij} = \Pr\{\xi_{n+1}=Z_j \mid \xi_n = Z_i\}, \qquad \text{with} \quad p_{ii}=0, \qquad \text{(A7.148)}$$

and

$$F_{ij}(x) = \Pr\{\eta_n \le x \,\big|\, (\xi_n = Z_i \cap \xi_{n+1} = Z_j)\}. \qquad \text{(A7.149)}$$

For $p_{ij} = 0$ $F_{ij}(x)$ can be arbitrary. From Eq. (A7.144), the consecutive jump points at which the process enters Z_i are *regeneration points*. This holds for any $i \in \{0, ..., m\}$. Thus, *all states* of a semi-Markov process are *regeneration states*. The renewal density of the *embedded renewal process* of consecutive jumps in Z_i (i-renewals) will be denoted as $h_i(t)$.

The *initial distribution*, i.e. the distribution of the vector $(\xi(0), \xi_1, \eta_0)$ is given for the general case by

$$\begin{aligned} A_{ij}(x) &= \Pr\{\xi(0) = Z_i \cap \xi_1 = Z_j \cap \text{residual sojourn time in } Z_i \le x\} \\ &= P_i(0) p_{ij} \, F^{\circ}_{ij}(x), \end{aligned} \tag{A7.150}$$

with $P_i(0) = \Pr\{\xi(0) = Z_i\}$, p_{ij} according to Eq. (A7.148), and $F^{\circ}_{ij}(x) = \Pr\{\text{residual sojourn time in } Z_j \le x \mid (\xi(0) = Z_i \cap \xi_1 = Z_j)\}$. $\xi(0)$ is used here for clarity instead of ξ_0. As pointed out above, the semi-Markov process is Markovian, i.e. *memoryless* (in general) *only at the transition points from one state to the other*. To have the time $t = 0$ as a regeneration point, the initial condition $\xi(0) = Z_i$, sufficient for Markov processes, *must* be reinforced for semi-Markov processes by

Z_i *is entered at* $t = 0$.

The sequence $\xi_0, \xi_1, \ldots$ forms a *Markov chain, embedded Markov chain*, with *transition probabilities* p_{ij} according to Eq. (A7.148), with $p_{ii} = 0$ and *initial probabilities* $P_i(0)$, $i = 0, \ldots, m$. $F_{ij}(x)$ is the *conditional distribution function* of the sojourn time in Z_i with consequent jump in Z_j (as next state to be visited).

A semi-Markov process is a *Markov process* if and only if $F_{ij}(x) = 1 - e^{-\rho_i x}$, for $i, j \in \{0, \ldots, m\}$. As an example of a two state semi-Markov process, consider the *alternating renewal process* introduced in Appendix A7.3 ($Z_0 = up$, $Z_1 = down$, $p_{01} = p_{10} = 1$, $F_{01}(x) = F(x)$, $F_{10}(x) = G(x)$, $F^{\circ}_0(x) = F_A(x)$, $F^{\circ}_1(x) = G_A(x)$, $P_0(0) = p$, $P_1(0) = 1 - p$).

In most applications, the quantities $Q_{ij}(x)$, or p_{ij}, and $F_{ij}(x)$, can be calculated using Eqs. (A7.99) to (A7.101), as illustrated in Sections 6.3 – 6.6.

For the *unconditional sojourn time* in Z_i, the distribution function is given by (considering $p_{ii} = 0$)

$$Q_i(x) = \Pr\{\eta_n \le x \mid \xi_n = Z_i\} = \sum_{\substack{j=0 \\ j \ne i}}^{m} p_{ij} \, F_{ij}(x) = \sum_{\substack{j=0 \\ j \ne i}}^{m} Q_{ij}(x), \tag{A7.151}$$

and the *mean* by

$$T_i = \int_0^{\infty} (1 - Q_i(x)) \, dx. \tag{A7.152}$$

In the following it will be assumed that

$$q_{ij}(x) = \frac{d Q_{ij}(x)}{dx} \tag{A7.153}$$

exists for all $i, j \in \{0, ..., m\}$.

Consider first the case in which *the process enters the state* Z_i *at* $t = 0$, i.e. that

$$P_i(0) = 1 \qquad \text{and} \qquad F^{\circ}_{ij}(x) = F_{ij}(x). \tag{A7.154}$$

The *transition probabilities*

$$P_{ij}(t) = \Pr\{\xi(t) = Z_j \mid Z_i \text{ is entered at } t = 0\} \tag{A7.155}$$

can be obtained by generalizing Eq. (A7.120), however considering that the condition Z_i *is entered at* $t = 0$ is mandatory for semi-Markov processes,

$$P_{ij}(t) = \delta_{ij} (1 - Q_i(t)) + \sum_{\substack{k=0 \\ k \ne i}}^{m} \int_0^t q_{ik}(x) P_{kj}(t - x) \, dx, \tag{A7.156}$$

with $\delta_{ij} = 0$ for $j \ne i$, $\delta_{ii} = 1$, $Q_i(t)$ as in Eq. (A7.151), and taking care of $p_{ii} = 0$ or $Q_{ii}(x) = q_{ii}(x) \equiv 0$ (Eq. (A7.148)). The *state probabilities* follow then as

$$P_j(t) = \Pr\{\xi(t) = Z_j\} = \sum_{i=0}^{m} \Pr\{Z_i \text{ is entered at } t = 0\} P_{ij}(t), \tag{A7.157}$$

with $P_j(t) \ge 0$ and $P_0(t) + \ldots + P_m(t) = 1$. If the state space is divided into the complementary sets U for the *up states* and $\overline{U}$ for the *down states*, as in Eq. (A7.107), the *point availability* follows again from Eq. (A7.112)

$$PA_{Si}(t) = \Pr\{\xi(t) \in U \mid Z_i \text{ is entered at } t = 0\} = \sum_{Z_j \in U} P_{ij}(t), \qquad i = 0, ..., m,$$

with $P_{ij}(t)$ as in Eq. (A7.156). The probability that the *first transition* from a state in U into a state in $\overline{U}$ occurs after time t, i.e. the *reliability function*, can be obtained by generalizing the system of equations (A7.122).

$$R_{Si}(t) = 1 - Q_i(t) + \sum_{\substack{Z_j \in U \\ j \ne i}} \int_0^t q_{ij}(x) R_{Sj}(t - x) \, dx, \qquad Z_i \in U, \tag{A7.158}$$

with $Q_i(t)$ as in Eq. (A7.151) and considering $p_{ii} = 0$, i.e. $q_{ii}(x) \equiv 0$. The *mean of the sojourn times in U*, i.e. the *system mean time to failure*, follows then from Eq. (A7.158) as solution of the following system of algebraic equations

$$MTTF_{S\,i} = T_i + \sum_{\substack{Z_j \in U \\ j \ne i}} p_{ij} \, MTTF_{Sj}, \qquad Z_i \in U, \tag{A7.159}$$

with T_i according to Eq. (A7.152).

Consider now the case of a *stationary semi-Markov process.* Under the assumption that the embedded Markov chain is *irreducible* (each state can be reached from every other state with probability > 0), the semi-Markov process is *stationary* if and only if the *initial distribution* (Eq. (A7.150)) is given by [A7.22, A7.23, A7.28]

$$\mathrm{A}_{ij}(x) = \frac{p_i\, p_{ij}}{\sum\limits_{k=0}^{m} p_k\, T_k} \int\limits_0^x (1 - \mathrm{F}_{ij}(y))\, dy. \tag{A7.160}$$

In Eq. (A7.160), p_{ij} are the *transition probabilities* (Eq. (A7.148) and p_j the *stationary distribution of the embedded Markov chain,* i.e. p_j are the unique solutions of (see also Eq. (A7.74) but with $p_{ii} \equiv 0$)

$$p_j = \sum_{i=0}^{m} p_i\, p_{ij}, \qquad \text{with} \quad p_{ii} \equiv 0, \quad p_{ij} = \mathrm{Q}_{ij}(\infty), \quad p_j > 0, \quad \sum_{j=0}^{m} p_j = 1, \quad j = 0, \ldots, m. \tag{A7.161}$$

For the stationary semi-Markov process, the *state probabilities* are independent of time and, for $t \ge 0$, given by

$$\mathrm{P}_i(t) = P_i = \frac{T_i}{T_{ii}} = \frac{p_i\, T_i}{\sum\limits_{j=0}^{m} p_j\, T_j}, \qquad i = 0, \ldots, m. \tag{A7.162}$$

T_{ii} is the mean of the time interval between two consecutive occurrences of Z_i. These time points form a *stationary renewal process* with renewal density

$$h_i(t) = \frac{1}{T_{ii}} = \frac{p_i}{\sum\limits_{j=0}^{m} p_j\, T_j}, \qquad i = 0, \ldots, m. \tag{A7.163}$$

It can be shown that Eq. (A7.163) is equivalent to Eq. (A7.160). The steady-state value of the *point availability* follows then from Eq. (A7.162)

$$PA_S = \sum_{Z_i \in U} P_i = \sum_{Z_i \in U} \frac{p_i\, T_i}{\sum\limits_{j=0}^{m} p_j\, T_j}. \tag{A7.164}$$

Under the assumptions made above, i.e. continuous sojourn times with finite means and an irreducible embedded Markov chain, the following applies for $i = 0, \ldots, m$ regardless of the initial distribution at $t = 0$

$$\lim_{t \to \infty} \Pr\{\xi(t) = Z_i \cap \text{next transition in } Z_j$$
$$\cap \text{ residual sojourn time in } Z_i \le x\} = \frac{p_i\, p_{ij}}{\sum\limits_{k=0}^{m} p_k\, T_k} \int\limits_0^x (1 - \mathrm{F}_{ij}(y))\, dy = \mathrm{A}_{ij}(x), \tag{A7.165}$$

and thus

$$\lim_{t \to \infty} \Pr\{\xi(t) = Z_i\} = P_i = \frac{T_i}{T_{ii}} = \frac{p_i\, T_i}{\sum\limits_{j=0}^{m} p_j\, T_j}, \quad \text{and} \quad \lim_{t \to \infty} \mathrm{PA}_S(t) = PA_S = \sum_{Z_i \in U} P_i. \tag{A7.166}$$

For the alternating renewal process (Sec. A7.3 with $Z_0 = up$, $Z_1 = down$, $T_0 = MTTF$, and $T_1 = MTTR$) it holds that $p_0 = p_1 = 1/2$ (embed. Markov chain) and $T_{00} = T_{11} = T_0 + T_1$. Eq.(A7.164), or (A7.166), leads to $PA_S = P_0 = T_0 / T_{00} = T_0 / (T_0 + T_1) = p_0 T_0 / (p_0 T_0 + p_1 T_1)$. This example shows best the *basic difference* between p_i as a stationary distribution of the embedded Markov chain and the limiting state probability P_i in state Z_i of the original process in continuous time ($p_0 \ne P_0 = PA_S$ for $T_0 \ne T_1$).

A7.7 Semi-regenerative Processes

As pointed out in Appendix A7.5.2, the time behavior of a repairable system can be *described by a Markov process* only if the failure-free operating times and repair times of all elements are *exponentially distributed* (the constant failure and repair rates can however depend upon the current state, see Figs. A7.4, 6.8, 6.13, and 2.12 for some examples). Beside the special case of the *Erlang distribution* (Eq. (A6.102) and Section 6.3.3), non-exponentially distributed repair times and/or failure-free operating times lead only in some few cases to semi-Markov processes, but more generally to processes *with only few regeneration states* or even to *nonregenerative processes*. To make sure that the time behavior of a system can be described by a *semi-Markov process*, there must be *no* "running" failure-free operating time or repair time *at any state transition* (state change) which is *not exponentially distributed,* otherwise the sojourn time to the next state transition would depend on *how long* these non-exponentially distributed times have already run. Example A7.12 shows the case of a process with states Z_0, Z_1, Z_2 in which only the states Z_0 and Z_1 are regenerative. Z_0 and Z_1 form a *semi-Markov process embedded* in the original process, on which the investigation can be based. Processes with an *embedded semi-Markov process* are called *semi-regenerative processes.* Their investigation can become time consuming and has to be performed, in general, on a *case-by-case basis*, see Sections 6.4.2 and 6.4.3 for some examples.

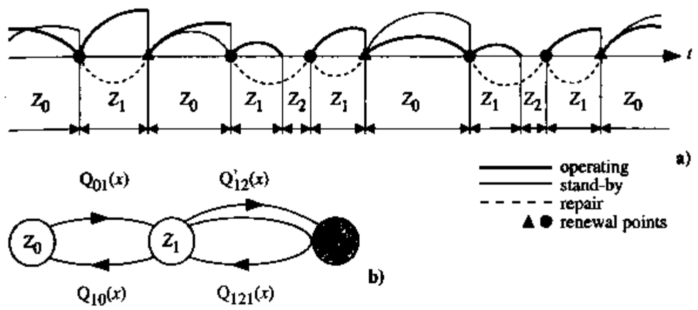


Figure A7.10 **a)** Time schedule for a 1-out-of-2 warm redundancy with constant failure rates (λ and λ_r), arbitrary repair rate, and only one repair crew (repair times are shown greatly exaggerated); **b)** *State transition diagram* (embedded semi-Markov process with regenerative states Z_0 and Z_1)

Example A7.12

Consider a *warm redundancy* as in Fig. A7.4a with one repair crew, constant failure rates λ in the operating and λ_r in the reserve state, and an arbitrarily distributed repair time with distribution function $G(x)$ and density $g(x)$. Determine the transition probabilities of the *embedded semi-Markov process*.

Solution

As the time schedule in Fig. A7.10 shows, only the states Z_0 and Z_1 are regenerative. Z_2 is not regenerative because at the transition points into Z_2 a repair with *arbitrary* repair rate is running. Thus, the process involved is *not* a semi-Markov process. However, states Z_0 and Z_1 form an ***embedded semi-Markov process*** on which investigations can be based. The transition probabilities of the embedded semi-Markov process are obtained, using Eq. (A7.99) and Fig. A7.10 as,

$$Q_{01}(x) = \Pr\{\tau_{01} \le x\} = 1 - e^{-(\lambda+\lambda_r)x}$$

$$Q_{10}(x) = \Pr\{\tau_{10} \le x \cap \tau_{12} > \tau_{10}\} = \int_0^x g(y) e^{-\lambda y} dy = G(x) e^{-\lambda x} + \int_0^x \lambda e^{-\lambda y} G(y)\, dy$$

$$Q_{121}(x) = \Pr\{\tau_{121} \le x\} = \int_0^x g(y)(1 - e^{-\lambda y})\, dy. \qquad \text{(A7.167)}$$

$Q_{121}(x)$ is used to compute the *point availability* (Section 6.4.2). It accounts for the process returning from state Z_2 to state Z_1 (Fig. A7.10a) and that Z_2 is a *not a regeneration state* (probability for the transition $Z_1 \to Z_2 \to Z_1$). $Q'_{12}(x)$ as given in Fig A7.10 is *not* a semi-Markov transition probability (Z_2 is not a regeneration state). However, $Q'_{12}(x)$ is useful in this example for the computation of an equivalent $Q_1(x) = Q_{10}(x) + Q'_{12}(x)$, see Section 6.4.2,

$$Q'_{12}(x) = \Pr\{\tau_{12} \le x \cap \tau_{10} > \tau_{12}\} = \int_0^x \lambda e^{-\lambda y}(1 - G(y))\, dy = 1 - e^{-\lambda x} - \int_0^x \lambda e^{-\lambda y} G(y)\, dy.$$

In the following, some general considerations on *semi-regenerative processes* are given. A pure jump process $\xi(t)$ with state space $Z_0, \ldots, Z_m$ is semi-regenerative, with regeneration states $Z_0, \ldots, Z_k$, $k < m$, if the following holds: Let $\zeta_0, \zeta_1, \ldots$ be the sequence of successively occurring regeneration states and $\varphi_0, \varphi_1, \ldots$ the random time intervals between consecutive occurrence of regenerative states (positive continuous random variables) then Eq.(A7.144) must be fulfilled for $n = 1, 2, \ldots$, arbitrary $i, j, i_0, \ldots, i_{n-1} \in \{0, \ldots, k\}$, and arbitrary positive values $x_0, \ldots, x_{n-1}$ (where ξ_n, η_n have been changed in ζ_n, φ_n). In other words, $\xi(t)$ as given by $\xi(t) = \zeta_n$ for $\varphi_0 + \ldots + \varphi_{n-1} \le t < \varphi_0 + \ldots + \varphi_n$ is a *semi-Markov process* with state space $Z_0, \ldots, Z_k$ and transition probabilities $Q_{ij}(x)$, $i, j \in \{0, \ldots, k\}$, *embedded in the original process* $\xi(t)$.

The piece $\xi(t)$, $\varphi_0 + \ldots + \varphi_{n-1} \le t < \varphi_0 + \ldots + \varphi_n$, $n = 1, 2, \ldots$, of the original process is a *cycle*, as defined in Appendix A7.4. Its distribution depends on ζ_n, i.e. on the regeneration state involved, and its probabilistic structure can be very complicated. The epochs at which a fixed state Z_i, $0 \le i \le k$ occurs are *regeneration points* and constitute a *renewal process.*

Often the set of system up states U is a subset of the regeneration states $Z_0, \ldots, Z_k$. Equations (A7.158) and (A7.159) can then be used to calculate the *reliability function* $R_{Si}(t)$ and the *system mean time to failure* $MTTF_{Si}$ (sojourn time in U). For the *point availability* $PA_{Si}(t)$, integral equations similar to Eq. (A7.156) can be given (Sections 6.4.2, 6.4.3).

If the embedded semi-Markov process has an *irreducible embedded Markov chain* and *continuous* conditional distribution functions $F_{ij}(x) = \Pr\{\varphi_n \le x \mid (\zeta_n = Z_i \cap \zeta_{n+1} = Z_j)\}$, $i, j \in \{0, \ldots, k\}$, then

$$\lim_{t\to\infty} \Pr\{\xi(t) = Z_i\}, \qquad i = 0, \ldots, k, \qquad \text{(A7.168)}$$

exist and do not depend on the initial distribution at $t = 0$ [A6.6 (Vol. II)]. The proof is based on the key renewal theorem (Eq. (A7.29)). Denoting by T_i the mean sojourn time in the state Z_i and by T^*_{ii} the mean of the time interval between *two consecutive occurrences* of Z_i (cycle length), it holds for $i = 0, \ldots, k$ that

$$\lim_{t\to\infty} \Pr\{\xi(t) = Z_i\} = P_i = \frac{T_i}{T^*_{ii}} \qquad \text{and} \qquad PA_S = \sum_{Z_i \in U} P_i. \qquad \text{(A7.169)}$$

For Example A7.12 it holds that $p_0 = p_1 = 1/2$ (embed. Markov chain), $T_0 = 1/(\lambda+\lambda_r)$, $T_1 = (1-\tilde{g}(\lambda))/\lambda$, $T^*_{00} = 1/(\lambda+\lambda_r) + MTTR + [(1-\tilde{g}(\lambda))/\tilde{g}(\lambda)]\, MTTR$, $T^*_{11} = \tilde{g}(\lambda)\,[1/(\lambda+\lambda_r) + MTTR] + (1-\tilde{g}(\lambda))\, MTTR$, $P_0 = T_0 / T^*_{00}$, and $P_1 = T_1 / T^*_{11}$, see Eq. (6.109) for $PA_S = P_0 + P_1$.

A7.8 Nonregenerative Stochastic Processes

The assumption of arbitrarily (i.e. not exponentially) distributed failure-free operating times and repair times for the elements of a system already leads to *nonregenerative stochastic processes* for *simple* series or parallel structures.

Solutions are often problem-oriented and the aim of this section is merly to present some basic considerations. A general method of analysis consists in transforming the given stochastic process into a semi-Markov or a Markov process by a suitable *state space extension*. The following possibilities are often used:

1. *Approximation of distribution functions*: Approximating the distribution functions (for failure-free operating times and for repair times) by *Erlang distributions* (Eq. (A6.102)) allows a transformation of the original process into a time-homogeneous Markov process through introduction of additional states.
2. *Introduction of supplementary variables*: Introducing for every element of a system as supplementary variables the failure-free operating time since the last repair and the repair time since the last failure, the original process can be transformed into a Markov process with state space consisting of discrete and continuous parameters. Investigations usually lead to partial differential equations which can be solved with the corresponding boundary conditions.

The first method is best used when the failure and/or repair rates involved are *monotonically increasing* from zero to a final value, its application is simple and easy to understand (Fig. 6.6). The second method is very general [A7.4 (1955)]. However, difficulties with partial differential equations can limit its use.

A further method is based on the geneal concept of *point processes*. Considering the sequence of jump times S_n and states ξ_n entered at these points, an equivalent description of the process $\xi(t)$ is obtained by a marked point process (S_n, ξ_n), $n=0, 1, \ldots$. Analysis of the system's *steady-state behavior* follows using *Korolyuk's theorem* ($\Pr\{\text{jump into } Z_i \text{ during } (t, t+\delta t]\} = \overset{\circ}{\lambda}_i \delta t + o(\delta t)$, with $\overset{\circ}{\lambda}_i = E$[Number of jumps in Z_i during the unit time interval]), see e.g. [A7.11, A7.12]. As an example, consider a coherent system (Section 2.3.4) with n independent elements. Let $\zeta_1(t), \ldots, \zeta_n(t)$ and $\zeta(t)$ be the binary processes with states 0 (*down*) and 1 (*up*) describing the elements and the system, respectively. If the steady-state *point availability* of any element E_i of the system exists

$$\lim_{t \to \infty} \mathrm{PA}_i(t) = \lim_{t \to \infty} \Pr\{\zeta_i(t)=1\} = PA_i = \frac{MTTF_i}{MTTF_i + MTTR_i}, \qquad i=1, \ldots, n, \qquad \text{(A7.170)}$$

then the steady-state point availability of the system is given by Eq. (2.48) and can be expressed as $PA_S = MTTF_S / (MTTF_S + MTTR_S)$, see e.g. [6.3, A7.10]. Investigations of the time behavior of systems with arbitrary failure and/or repair rates can become difficult. In these cases, *approximate expressions* can be used efficiently, see Sections 6.7 and 6.8.

A8 Basic Mathematical Statistics

Mathematical statistics deals basically with situations which can be described as follows: Given a population of *statistically identical, independent elements* with unknown (statistical) properties, measurements regarding these properties are made on a (random) sample of this population and, on the basis of the collected data, conclusions are made for the *remaining elements* of the population. Examples are the parameter estimation for the distribution function of an item's failure-free operating time τ, or the decision whether the expected value (mean) of τ is greater than a given value. Mathematical statistics thus goes from *observations* (realizations) of a given (random) event in a series of *independent trials* and searches for a suitable *probabilistic model* for the event considered (inductive approach). Methods used are based on probability theory and results obtained can only be formulated in a *probabilistic language*. *Minimization of the risk for a false conclusion* is an important objective in mathematical statistics. This Appendix introduces the basic concepts of mathematical statistics used in planning and evaluating quality and reliability tests, as given in Chapter 7. Emphased are *empirical methods*, (statistical) *parameter estimation*, and (statistical) *testing of hypotheses*. To simplify the notation, the terms *random* and *statistical* (in brackets) will often be omitted, and *mean* stands for *expected value*. This appendix is a compendium of mathematical statistics, consistent from a mathematical point of view but still with engineering applications in mind.

A8.1 Empirical Methods

Empirical methods allow a quick and easy estimation of the distribution function as well as of the mean, variance, and other moments characterizing a random variable. These *estimates* are generally based on the empirical distribution function and have thus great intuitive appeal.

A8.1.1 Empirical Distribution Function

A *sample* of size n of a random variable τ with the distribution function $F(t)$ is a random vector $\vec{\tau} = (\tau_1, \ldots, \tau_n)$ whose components τ_i are tacitly assumed to be *independent* and *identically distributed* random variables with $F(t) = \Pr\{\tau_i \le t\}$, $i = 1, \ldots, n$. For example, $\tau_1, \ldots, \tau_n$ can be the failure-free operating times of n items randomly selected from a lot of statistically identical items, with a distribution function $F(t)$ of the failure-free operating time τ. The *observed* failure-free operating times, i.e. the *realization* of the random vector $\vec{t} = (\tau_1, \ldots, \tau_n)$, is a set $t_1, \ldots, t_n$ of positive real values. The distinction between random variables $\tau_1, \ldots, \tau_n$ and their observations $t_1, \ldots, t_n$ is important from a mathematical point of view.*)

When the sample elements are ordered by increasing magnitude, an *ordered sample* $\tau_{(1)}, \ldots, \tau_{(n)}$ is obtained. The corresponding *ordered observations* are $t_{(1)}, \ldots, t_{(n)}$. For a set of ordered observations $t_{(1)}, \ldots, t_{(n)}$, the right continuous function

$$\hat{F}_n(t) = \begin{cases} 0 & \text{for } t < t_{(1)} \\ \dfrac{i}{n} & \text{for } t_{(i)} \le t < t_{(i+1)} \\ 1 & \text{for } t \ge t_{(n)} \end{cases} \qquad \text{(A8.1)}$$

is the *empirical distribution function* of the random variable τ, see Fig. A8.1 for a graphical representation. $\hat{F}_n(t)$ expresses the relative frequency of the event $\{\tau \le t\}$ in n independent trial repetitions and provides a well defined estimate of the distribution function $F(t) = \Pr\{\tau \le t\}$. In the following, the symbol ^ is used to denote an estimate of an unknown quantity.

As mentioned in the footnote below, when investigating the properties of the empirical distribution function $\hat{F}_n(t)$ it is necessary in Eq. (A8.1) to replace the observations $t_{(1)}, \ldots, t_{(n)}$ by the sample elements $\tau_{(1)}, \ldots, \tau_{(n)}$. For any given value of t, $n\hat{F}_n(t)$ is a *binomially-distributed* random variable (Eq. (A6.120)) with parameter $p = F(t)$. Thus $\hat{F}_n(t)$ has mean

$$E[\hat{F}_n(t)] = F(t), \qquad \text{(A8.2)}$$

and variance

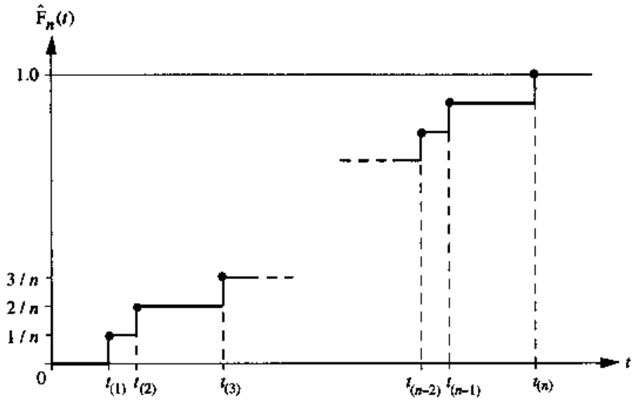


Figure A8.1 Example of an empirical distribution function

$$\operatorname{Var}[\hat{F}_n(t)] = \frac{F(t)(1 - F(t))}{n}. \qquad \text{(A8.3)}$$

Moreover, a direct application of the *strong law of large numbers* (Eq. (A6.146)) shows that for any given value of t, $\hat{F}_n(t)$ converges to $F(t)$ with *probability one* for $n \to \infty$. This convergence is *uniform* in t and thus holds for the whole distribution function $F(t)$. Proof of this important result is given in the *Glivenko-Cantelli theorem* [A8.5, A8.14, A8.16], which states that the largest absolute deviation between $\hat{F}_n(t)$ and $F(t)$ over all t, i.e.

$$D_n = \sup_{-\infty < t < \infty} \left| \hat{F}_n(t) - F(t) \right|, \qquad \text{(A8.4)}$$

converges with probability one toward 0

$$\Pr\{\lim_{n \to \infty} D_n = 0\} = 1. \qquad \text{(A8.5)}$$

A8.1.2 Empirical Moments and Quantiles

The moments of a random variable τ are completely determined by the distribution function $F(t) = \Pr\{\tau \le t\}$. The empirical distribution $\hat{F}_n(t)$ introduced in Appendix A8.1.1 can be used to estimate the unknown moments of τ.

The values $t_{(1)}, \ldots, t_{(n)}$ having been fixed, $\hat{F}_n(t)$ can be regarded as the distribution function of a discrete random variable with probability $p_k = 1/n$ at the

---

*) The investigation of statistical methods and the discussion of their properties can only be based on the (random) *sample* $\tau_1, \ldots, \tau_n$. However in applying the methods for a numerical evaluation (statistical decision), the *observations* $t_1, \ldots, t_n$ have to be used. For this reason, the same equation (or procedure) can be applied to τ_i or t_i according to the situation.

points $t_{(k)}$, $k=1,\dots,n$. Using Eq. (A6.35), the corresponding mean is the *empirical mean* (empirical expectation) of τ and is given by

$$\hat{\mathrm{E}}[\tau]=\frac{1}{n}\sum_{i=1}^{n}t_i. \qquad \text{(A8.6)}$$

Taking into account the footnote on p. 436, $\hat{\mathrm{E}}[\tau]$ is a random variable with mean

$$\mathrm{E}[\hat{\mathrm{E}}[\tau]]=\mathrm{E}[\frac{1}{n}\sum_{i=1}^{n}\tau_i]=\frac{1}{n}n\,\mathrm{E}[\tau]=\mathrm{E}[\tau], \qquad \text{(A8.7)}$$

and variance

$$\mathrm{Var}[\hat{\mathrm{E}}[\tau]]=\mathrm{Var}[\frac{1}{n}\sum_{i=1}^{n}\tau_i]=\frac{1}{n^2}n\,\mathrm{Var}[\tau]=\frac{\mathrm{Var}[\tau]}{n}. \qquad \text{(A8.8)}$$

Equation (A8.7) shows that $\hat{\mathrm{E}}[\tau]$ is an *unbiased estimate* of $\mathrm{E}[\tau]$, see Eq. (A8.18). Furthermore, from the strong law of large numbers (Eq. (A6.147)) it follows that $\hat{\mathrm{E}}[\tau]\to\mathrm{E}[\tau]$ as $n\to\infty$

$$\Pr\{\lim_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}\tau_i)=\mathrm{E}[\tau]\}=1. \qquad \text{(A8.9)}$$

The exact distribution function of $\hat{\mathrm{E}}[\tau]$ can only be given for some particular cases, e.g. normal, exponential, or gamma distributions. However, the *central limit theorem* (Eq. (A6.148)) shows that for large values of n *the* distribution of $\hat{\mathrm{E}}[\tau]$ can be approximated by a *normal distribution* with mean $\mathrm{E}[\tau]$ and variance $\mathrm{Var}[\tau]/n$.

Based on $\hat{\mathrm{F}}_n(t)$, Eqs. (A6.43) and (A8.6) provide an estimate of the variance as

$$\frac{1}{n}\sum_{i=1}^{n}(t_i-\frac{1}{n}\sum_{i=1}^{n}t_i)^2.$$

However, the expectation of this estimate is given by $\mathrm{Var}[\tau](n-1)/n$. For this reason, the *empirical variance* of τ is usually defined as

$$\hat{\mathrm{Var}}[\tau]=\frac{1}{n-1}\sum_{i=1}^{n}(t_i-\hat{\mathrm{E}}[\tau])^2=\frac{1}{n-1}[\sum_{i=1}^{n}t_i^2-\frac{1}{n}(\sum_{i=1}^{n}t_i)^2], \qquad \text{(A8.10)}$$

for which it follows that

$$\mathrm{E}[\hat{\mathrm{Var}}[\tau]]=\mathrm{Var}[\tau]. \qquad \text{(A8.11)}$$

The *higher-order moments* (Eqs. (A6.41) and (A6.50)) can be similarly estimated with

$$\frac{1}{n}\sum_{i=1}^{n}t_i^k \qquad \text{and} \qquad \frac{1}{n-1}\sum_{i=1}^{n}(t_i-\hat{\mathrm{E}}[\tau])^k. \qquad \text{(A8.12)}$$

The *empirical quantile* $\hat{t}_q$ is defined as the *q quantile* of the empirical distribution function $\hat{\mathrm{F}}_n(t)$

$$\hat{t}_q=\inf\{t\colon \hat{\mathrm{F}}_n(t)\ge q\}. \qquad \text{(A8.13)}$$

A8.1.3 Further Applications of the Empirical Distribution Function..

Comparison of the empirical distribution function $\hat{\mathrm{F}}_n(t)$ with a given distribution function $\mathrm{F}(t)$ is the basis for several *non-parametric* statistical methods. These include goodness-of-fit tests, confidence bands for distribution functions, and graphical methods using probability charts.

A quantity often used in this context is the largest absolute deviation D_n between $\hat{\mathrm{F}}_n(t)$ and $\mathrm{F}(t)$, defined by Eq. (A8.4). If the distribution function $\mathrm{F}(t)$ of the random variable τ is continuous, then the random variable $\mathrm{F}(\tau)$ is uniformly distributed in $[0, 1]$. It follows that D_n has a distribution independent of $\mathrm{F}(t)$. A.N. Kolmogorov showed [A8.20] that for $\mathrm{F}(t)$ continuous and $x>0$,

$$\lim_{n\to\infty}\Pr\{\sqrt{n}\,D_n\le x\mid \mathrm{F}(t)\}=1+2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2x^2}.$$

The series converges rapidly, so that for $x>1/\sqrt{n}$,

$$\lim_{n\to\infty}\Pr\{D_n\le y\mid \mathrm{F}(t)\}\approx 1-2e^{-2ny^2}. \qquad \text{(A8.14)}$$

The distribution function of D_n has been tabulated for small values of n [A8.25], see Table A9.5 and Table A8.1. From the above it follows that:

For a given continuous distribution function $\mathrm{F}(t)$, *the band* $\mathrm{F}(t)\pm y_{1-\alpha}$ *overlaps the empirical distribution function* $\hat{\mathrm{F}}_n(t)$ *with probability* $1-\alpha_n$ *where* $\alpha_n\to\alpha$ *as* $n\to\infty$, *with* $y_{1-\alpha}$ *defined by*

$$\Pr\{D_n\le y_{1-\alpha}\mid \mathrm{F}(t)\}=1-\alpha \qquad \text{(A8.15)}$$

and given in Table A9.5 or Table A8.1.

If $\mathrm{F}(t)$ is not continuous, it can be shown that with $y_{1-\alpha}$ from Eq. (A8.15), the band $\mathrm{F}(t)\pm y_{1-\alpha}$ overlaps $\hat{\mathrm{F}}_n(t)$ with a probability $1-\alpha_n'$, where $\alpha_n'\to\alpha'\le\alpha$ as $n\to\infty$.

Table A8.1 Values of $y_{1-\alpha}$, for which $\Pr\{D_n \le y_{1-\alpha} \mid F(t)\} = 1-\alpha$

n	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.05$
2	0.684	0.776	0.842
5	0.447	0.509	0.563
10	0.323	0.369	0.409
20	0.232	0.265	0.294
30	0.190	0.218	0.242
40	0.165	0.189	0.210
50	0.148	0.170	0.188
for $n > 50$	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$

The role of $F(t)$ and $\hat{F}_n(t)$ can be *reversed* in the above considerations, i.e.:

> *The random band* $\hat{F}_n(t) \pm y_{1-\alpha}$ *overlaps the true (unknown) distribution function* $F(t)$ *with probability* $1-\alpha_n$, *where* $\alpha_n \to \alpha$ *as* $n \to \infty$.

This last aspect is one of *mathematical statistics* (while the former one in relation to Eq. (A8.15) was a problem of *probability theory*), allowing an *estimation* of an unknown continuous distribution function $F(t)$ on the basis of the empirical distribution function $\hat{F}_n(t)$.

Example A8.1

How wide is the confidence band around $\hat{F}_n(t)$ for $n = 30$ and $n = 100$ if $\alpha = 0.2$?

Solution

From Table A8.1, $y_{0.8} = 0.19$ for $n = 30$ and $y_{0.8} = 0.107$ for $n = 100$. This leads to the band $\hat{F}_n(t) \pm 0.19$ for $n = 30$ and $\hat{F}_n(t) \pm 0.107$ for $n = 100$.

To simplify investigations, it is often useful to draw $\hat{F}_n(t)$ on a so-called *probability chart*. The method is as follows:

> *The empirical distribution function* $\hat{F}_n(t)$ *is drawn in a system of coordinates in which a postulated type of continuous distribution function is represented by a straight line; if the underlying distribution* $F(t)$ *belongs to this type of distribution function then, for a sufficiently large value of n the points* $(t_{(i)}, \hat{F}_n(t_{(i)}))$ *will approximate to a straight line (a systematic deviation from a straight line, particularly in the domain* $0.1 < \hat{F}_n(t) < 0.9$, *leads to rejection of the type of distribution function assumed).*

In many cases it is possible to obtain estimates for unknown parameters of the underlying distribution function $\hat{F}_n(t)$ from the estimated straight line for $\hat{F}_n(t)$.

This holds in particular for the exponential, Weibull, lognormal and normal distribution functions. Corresponding *probability charts* are given in Appendix A9.8. The following is a derivation of the *Weibull probability chart*. The function

$$F(t) = 1 - e^{-(\lambda t)^\beta}$$

can be transformed to $\log_{10}(\frac{1}{1-F(t)}) = (\lambda t)^\beta \log_{10}(e)$ and finally to

$$\log_{10}\log_{10}(\frac{1}{1-F(t)}) = \beta\log_{10}(t) + \beta\log_{10}(\lambda) + \log_{10}\log_{10}(e). \qquad \text{(A8.16)}$$

In the system of coordinates $\log_{10}(t)$ and $\log_{10}\log_{10}(1/(1-F(t))$, the *Weibull distribution* function given by $F(t) = 1 - e^{-(\lambda t)^\beta}$ appears as a straight line. Fig. A8.2 shows this for $\beta = 1.5$ and $\lambda = 1/800\,h$. As illustrated by Fig. A8.2, the parameters β and λ can be obtained graphically

- β is the slope of the straight line, it appears on the scale $\log_{10}\log_{10}(1/(1-F(t))$ if t is changed by one decade,
- for $\log_{10}\log_{10}(1/(1-F(t)) = \log_{10}\log_{10}(e)$, i.e. on the dashed line in Fig. A8.2, one has $\log_{10}(\lambda t) = 0$ and thus $\lambda = 1/t$.

Natural logarithms (ln) can be used instead of decimal logarithms ($\log_{10}$). The Weibull probability chart also applies to the *exponential distribution* ($\beta = 1$).

Practical examples for the application of graphical estimation methods are given in Section 7.5 (Figs. 7.12 – 7.14).

For a three *parameter Weibull distribution* ($F(t) = 1 - e^{-(\lambda(t-\psi))^\beta}$, $t \ge \psi$) one can operate with the time axis $t' = t - \psi$, giving a straight line as before, or consider the concave curve obtained when using t (see Fig. A8.2 for an example). Conversely, from a concave curve describing a Weibull distribution (e.g. in the case of an *empirical evaluation of data*) it is possible to find ψ using the relationship $\psi = (t_1 t_2 - t_m^2)/(t_1 + t_2 - 2t_m)$ existing between two arbitrary points t_1, t_2 and t_m obtained from the mean of $F(t_1)$ and $F(t_2)$ on the scale $\log_{10}\log_{10}(1/(1-F(t)))$, see Example A6.14 for a derivation and Fig. A8.2 for an application with $t_1 = 400h$ and $t_2 = 1000h$, giving $t_m = 600h$ and $\psi = 200h$.

A8.2 Parameter Estimation

In many applications it can be assumed that the *type* of distribution function $F(t)$ of the underlying random variable τ is known. This means that $F(t) = F(t, \theta_1, \ldots, \theta_r)$ is known in its functional form, the real-valued parameters $\theta_1, \ldots, \theta_r$ having to be estimated. The unknown parameters of $F(t)$ must be estimated on the basis of the observations $t_1, \ldots, t_n$. A distinction is made between *point* and *interval estimation*.

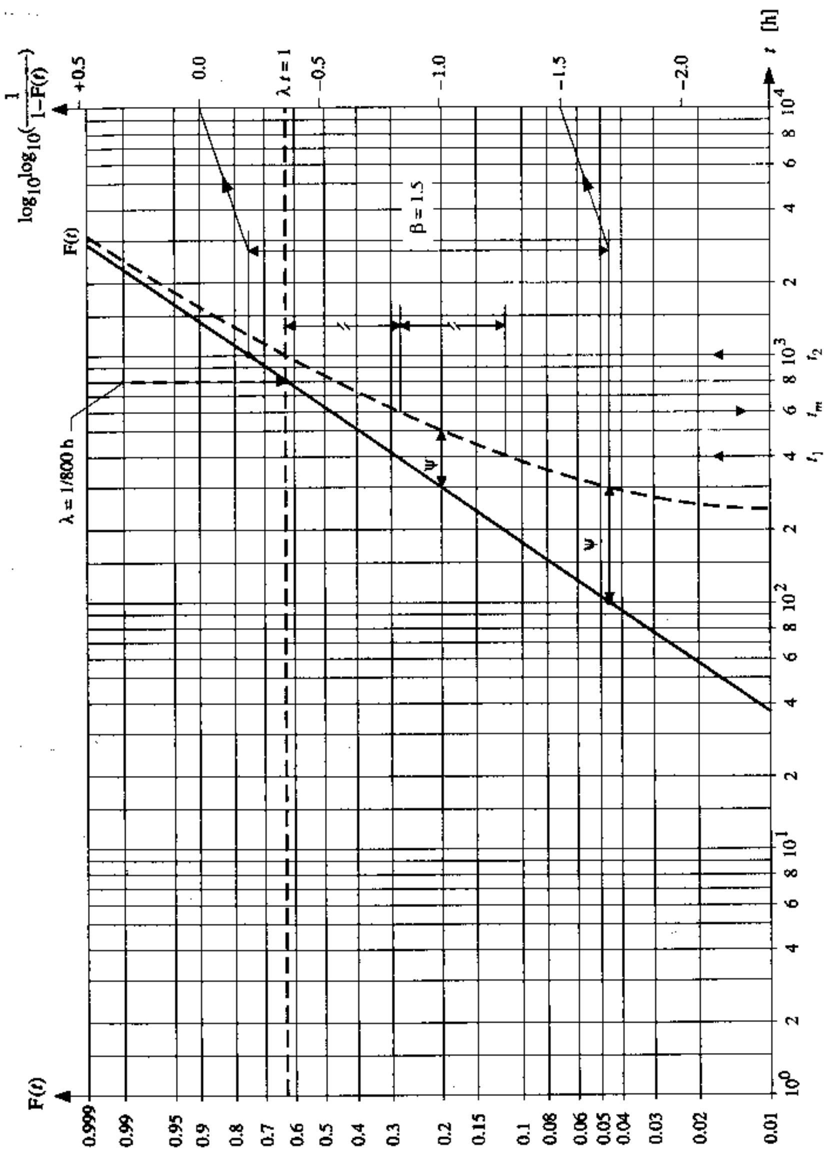


Figure A8.2 Weibull probability chart: The distribution function $F(t)=1-e^{-(\lambda t)^{\beta}}$ appears as a straight line (in the example $\lambda = 1/800\,\text{h}$ and $\beta = 1.5$); for a three parameter distribution $F(t)=1-e^{-(\lambda(t-\psi))^{\beta}}$, $t \ge \psi$, one can use $t'=t-\psi$ or operate with a concave curve and determine (as necessary) ψ, λ, and β graphically (dashed curve for $\lambda = 1/800\,\text{h}$, $\beta = 1.5$, and $\psi = 200\,\text{h}$ as an example)

A8.2.1 Point Estimation

Consider first the case where the given distribution function F(t) only depends of one (unknown) parameter θ. A *point estimate* for θ is a function

$$\hat{\theta}_n = u(t_1, \ldots, t_n) \tag{A8.17}$$

of the observations $t_1, \ldots, t_n$ of the random variable τ (not of the unknown parameter θ itself).[*] The estimate $\hat{\theta}_n$ is

- *unbiased*, if

  $$E[\hat{\theta}_n] = \theta, \tag{A8.18}$$

- *consistent*, if $\hat{\theta}_n$ converges to θ in probability, i.e. if for any $\varepsilon > 0$

  $$\lim_{n\to\infty} \Pr\{|\hat{\theta}_n - \theta| > \varepsilon\} = 0, \tag{A8.19}$$

- *strongly consistent*, if $\hat{\theta}_n$ converges to θ with probability one, i.e.

  $$\Pr\{\lim_{n\to\infty} \hat{\theta}_n = \theta\} = 1, \tag{A8.20}$$

- *efficient*, if

  $$E[(\hat{\theta}_n - \theta)^2] \tag{A8.21}$$

  is a minimum for every value of n,

- *sufficient*, if $\hat{\theta}_n$ contains all information about θ available in the observations t_1, ..., t_n, i.e. if the conditional distribution of $\vec{\tau}$ for given $\hat{\theta}_n$ does not depend on θ.

For an unbiased estimate, Eq. (A8.21) becomes

$$E[(\hat{\theta}_n - \theta)^2] = E[(\hat{\theta}_n - E[\hat{\theta}_n])^2] = \text{Var}[\hat{\theta}_n]. \tag{A8.22}$$

An unbiased estimate is thus efficient if $\text{Var}[\hat{\theta}_n]$ is a minimum, and consistent if $\text{Var}[\hat{\theta}_n] \to 0$ for $n \to \infty$. This last statement is a consequence of Chebyschev's inequality (Eq. (A6.49)). Other useful properties of estimates are asymptotic unbiasedness and asymptotic efficiency.

Several methods are known for estimating θ. To these belong the methods of moments, quantiles, least squares, and maximum likelihood. The *maximum likelihood method* [A8.6, A8.10, A8.18, A8.27] is commonly used in engineering applications. It provides point estimates which, under relatively general conditions,

---

[*] Bayesian estimation theory, which considers θ as a random variable and assigns to it an a priori distribution function, will not be considered here. As a function of the random sample $\vec{\tau} = (\tau_1, \ldots, \tau_n)$), $\hat{\theta}_n$ is a random variable (see footnote on p. 436), θ is an unknown constant.

are consistent, asymptotically unbiased, asymptotically efficient, and asymptotically normal-distributed. It can be shown that if an efficient estimate exists, then the *likelihood equation* (Eqs. (A8.25) or (A8.26)) has this estimate as a unique solution. Furthermore, an estimate $\hat{\theta}_n$ is *sufficient* if and only if the *likelihood function* (Eqs. (A8.23) or (A8.24)) can be written in two factors, one depending on $t_1, \dots, t_n$ only, the other on θ and $\hat{\theta}_n = \mathrm{u}(t_1, \dots, t_n)$, see Examples A8.2 to A8.4.

The *maximum likelihood method* was developed by R.A. Fisher [A8.15 (1921)] and is based on the following idea:

> *If $t_1, \dots, t_n$ are n independent observations of a discrete random variable τ, the probability of obtaining exactly these observations within a sample of size n of τ is given by the likelihood function*
>
> $$\mathrm{L}(t_1, \dots, t_n, \theta) = \prod_{i=1}^{n} \mathrm{p}_i(\theta), \qquad \text{with} \quad \mathrm{p}_i(\theta) = \Pr\{\tau = t_i\}; \tag{A8.23}$$
>
> *as an estimate of the unknown parameter θ, the value $\hat{\theta}_n$ has to be selected which maximizes the likelihood function L.*

In the case of a continuous random variable τ, the probabilities $\mathrm{p}_i(\theta)$ can be replaced by the density functions $\mathrm{f}(t_i, \theta)$ and Eq. (A8.23) becomes

$$\mathrm{L}(t_1, \dots, t_n, \theta) = \prod_{i=1}^{n} \mathrm{f}(t_i, \theta). \tag{A8.24}$$

Since the logarithmic function is monotonically increasing, the use of ln(L) instead of L leads to the same result. If $\mathrm{L}(t_1, \dots, t_n, \theta)$ is derivable and the *maximum likelihood estimate* $\hat{\theta}_n$ exists, then it will satisfy the equation

$$\left. \frac{\partial \mathrm{L}(t_1, \dots, t_n, \theta)}{\partial \theta} \right|_{\theta = \hat{\theta}_n} = 0, \tag{A8.25}$$

or

$$\left. \frac{\partial \ln(\mathrm{L}(t_1, \dots, t_n, \theta))}{\partial \theta} \right|_{\theta = \hat{\theta}_n} = 0. \tag{A8.26}$$

The maximum likelihood method can be generalized to the case of a distribution function with a finite number of unknown parameters $\theta_1, \dots, \theta_r$. Instead of Eq. (A8.26) for example, the following system of r algebraic equations must be solved

$$\left. \frac{\partial \ln(\mathrm{L}(t_1, \dots, t_n, \theta_1, \dots, \theta_r))}{\partial \theta_i} \right|_{\theta_i = \hat{\theta}_{in}} = 0, \qquad i = 1, \dots, r. \tag{A8.27}$$

The existence and uniqueness of a maximum likelihood estimate is satisfied in most practical applications. To simplify the notation, the index n will be omitted for the estimated parameters in the following.

Example A8.2

Let $t_1, \dots, t_n$ be statistically independent observations of an exponentially distributed failure-free operating time τ. Determine the maximum likelihood estimate for the unknown parameter λ of the *exponential distribution.*

Solution

With $\mathrm{f}(t, \lambda) = \lambda e^{-\lambda t}$, Eq. (A8.24) yields $\mathrm{L}(t_1, \dots, t_n, \lambda) = \lambda^n e^{-\lambda(t_1 + \dots + t_n)}$, from which

$$\hat{\lambda} = \frac{n}{t_1 + \dots + t_n}. \tag{A8.28}$$

This case corresponds to a sampling plan with n elements without replacement, terminated at the occurrence of the n-th failure (*Type II censoring*). $\hat{\lambda}$ depends only on the sum $t_1 + \dots + t_n$, not on the individual values of t_i; $t_1 + \dots + t_n$ is a *sufficient statistic* and $\hat{\lambda}$ is a *sufficient estimate* $(\mathrm{L} = 1 \cdot \lambda^n e^{-n\lambda/\hat{\lambda}})$. It must be noted that $\hat{\lambda} = n/(t_1 + \dots + t_n)$ is a biased estimate, unbiased is $\hat{\lambda} = (n-1)/(t_1 + \dots + t_n)$; unbiased is also $\hat{\mathrm{E}}[\tau] = (t_1 + \dots + t_n)/n$, see Eq. (A8.7).

Example A8.3

Assuming that an event A has occurred exactly k times in n Bernoulli trials, compute the max. likelihood estimate for the unknown *probability p* for event A to occur (binomial distribution).

Solution

Using Eq. (A6.120), the likelihood function (Eq. (A8.23)) becomes

$$L = p_k = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{or} \qquad \ln L = \ln\binom{n}{k} + k \ln p + (n-k) \ln(1-p).$$

This leads to

$$\hat{p} = \frac{k}{n}. \tag{A8.29}$$

$\hat{p}$ is unbiased. It depends only on k, i.e. on the number of the event occurrences in n independent trials; k is a *sufficient statistic* and $\hat{p}$ is a *sufficient estimate* $(L = \binom{n}{k} [p^{\hat{p}} (1-p)^{(1-\hat{p})}]^n)$.

Example A8.4

Let $k_1, \dots, k_n$ be independent observations of a random variable ζ distributed according to the Poisson distribution defined by Eq. (A6.125). Determine the maximum likelihood estimate for the unknown parameter m of the *Poisson distribution.*

Solution

The likelihood function becomes

$$L = \frac{m^{k_1 + \dots + k_n}}{k_1! \dots k_n!} e^{-nm} \qquad \text{or} \qquad \ln L = (k_1 + \dots + k_n) \ln m - mn - \ln(k_1! \dots k_n!)$$

and thus

$$\hat{m} = \frac{k_1 + \dots + k_n}{n}. \tag{A8.30}$$

$\hat{m}$ is unbiased. It depends only on the sum $k_1 + \dots + k_n$, not on the individual k_i; $k_1 + \dots + k_n$ is a *sufficient statistic and* $\hat{m}$ is a *sufficient estimate* $(L = (1/k_1! \dots k_n!) \cdot (m^{n\hat{m}} e^{-nm}))$.

Example A8.5

Let $t_1, \ldots, t_n$ be statistically independent observations of a Weibull distributed failure-free operating time τ. Determine the maximum likelihood estimate for the unknown parameters λ and β of the *Weibull distribution.*

Solution

With $f(t, \lambda, \beta) = \beta\lambda(\lambda t)^{\beta-1} e^{-(\lambda t)^\beta}$ it follows from Eq. (A8.24) that

$$L(t_1, \ldots, t_n, \lambda, \beta) = (\beta\lambda^\beta)^n e^{-\lambda^\beta (t_1^\beta + \ldots + t_n^\beta)} \prod_{i=1}^{n} t_i^{\beta-1}.$$

This leads to

$$\hat\beta = \left[\frac{\sum_{i=1}^{n} t_i^{\hat\beta} \ln t_i}{\sum_{i=1}^{n} t_i^{\hat\beta}} - \frac{1}{n}\sum_{i=1}^{n} \ln t_i\right]^{-1} \qquad \text{and} \qquad \hat\lambda = \Big[\frac{n}{\sum_{i=1}^{n} t_i^{\hat\beta}}\Big]^{1/\hat\beta}. \qquad \text{(A8.31)}$$

The solution for $\hat\beta$ is unique and can be found, for instance, using Newton's approximation method (the value obtained from the empirical distribution function can give a good initial value, see Fig. 7.12).

Due to cost and time limitations, the situation often arises in reliability applications when the items under test are run in parallel and the test is stopped before all items have failed. If there are n items, and at the end of the test k have failed (at the individual failure times $t_1 < t_2 < \ldots < t_k$) and $n-k$ are still working, then the operating times $T_1, \ldots, T_{n-k}$ of the items still working at the end of the test should also be accounted for in the evaluation. Considering a *Weibull distribution* as in Example A8.5, and assuming that the operating times $T_1, \ldots, T_{n-k}$ have been observed in addition to the failure-free operating times $t_1, \ldots, t_k$, then

$$L(t_1, \ldots, t_k, \lambda, \beta) \sim (\beta\lambda^\beta)^k e^{-\lambda^\beta (t_1^\beta + \ldots + t_k^\beta)} \prod_{i=1}^{k} t_i^{\beta-1} \prod_{j=1}^{n-k} e^{-(\lambda T_j)^\beta},$$

yielding

$$\hat\beta = \left[\frac{\sum_{i=1}^{k} t_i^{\hat\beta} \ln t_i + \sum_{j=1}^{n-k} T_j^{\hat\beta} \ln T_j}{\sum_{i=1}^{k} t_i^{\hat\beta} + \sum_{j=1}^{n-k} T_j^{\hat\beta}} - \frac{1}{k}\sum_{i=1}^{k} \ln t_i\right]^{-1} \quad \text{and} \quad \hat\lambda = \Big[\frac{k}{\sum_{i=1}^{k} t_i^{\hat\beta} + \sum_{j=1}^{n-k} T_j^{\hat\beta}}\Big]^{1/\hat\beta}. \quad \text{(A8.32)}$$

Often, $T_1 = \ldots = T_{n-k} = T_{test}$ is the fixed (given) test duration (*Type I (time) censoring*) or, alternatively, $T_1 = \ldots = T_{n-k} = t_k$, i.e. the test is stopped at the (random) occurrence of the k-th failure (*Type II (failure) censoring*). The two situations are different from a statistical point of view and this has to be considered in the analysis of data.

For $\beta = 1$, i.e. for the *exponential distribution*, Eq. (A8.31) reduces to Eq. (A8.28) and Eq. (A8.32) to

$$\hat\lambda = \frac{k}{\sum_{i=1}^{k} t_i + \sum_{j=1}^{n-k} T_j}. \qquad \text{(A8.33)}$$

If the test is stopped on the occurrence of the k-th failure, then (in general) $T_1 = \ldots = T_{n-k} = t_k$ and the quantity

$$T_R = t_1 + \ldots + t_k + (n-k)t_k \qquad \text{(A8.34)}$$

is the *random cumulative operating time* over all items during the test. Considering that $T_R = n\,t_1 + (n-1)(t_2 - t_1) + \ldots + (n-k+1)(t_k - t_{k-1})$, this situation corresponds to a sampling plan with n elements *without replacement* (renewal), censored at the occurrence of the k-th failure (*Type II censoring*), see Example A8.2 ($k = n$). It can be shown that the estimate $\hat\lambda = k / T_R$ is biased. An unbiased estimate is given by

$$\hat\lambda = \frac{k-1}{T_R}. \qquad \text{(A8.35)}$$

If the test is stopped at the fixed time T_{test}, then $T_R = t_1 + \ldots + t_k + (n-k)T_{test}$. In this case, $t_1, \ldots, t_k$ and k are *random*; k / T_R is a biased estimate for λ. This situation corresponds to a sampling plan with n elements *without replacement*, censored at a fixed (given) test time T_{test} (*Type I censoring*).

The case of n elements ($n \geq 1$) *with replacement* is important for practical applications and is considered in Sections A8.2.2.2 and 7.2.2.1 to 7.2.2.3.

A8.2.2 Interval Estimation

As shown in Appendix A8.2.1, a point estimation has the advantage of providing an estimate quickly. However, it does not give any indication as to the deviation of the estimate from the true parameter. More information can be obtained from an interval estimation. With an *interval estimation*, a (random) interval $[\hat\theta_l, \hat\theta_u]$ is sought such that it *overlaps* the true value of the unknown parameter θ with a given probability γ. $[\hat\theta_l, \hat\theta_u]$ is the *confidence interval*, $\hat\theta_l$ and $\hat\theta_u$ are the *lower* and *upper confidence limits*, and γ is the *confidence level*. γ has the following interpretation:

> *In an increasing number of independent samples of size n (used to obtain confidence intervals), the relative frequency of the cases in which the confidence intervals $[\hat\theta_l, \hat\theta_u]$ overlap the unknown parameter θ converges to the confidence level γ.*

In the discrete case it is often impossible to reach a given confidence level γ exactly. The true overlap probability should be here near to (but not less than) γ.

The confidence interval can also be *one-sided*, i.e. $(-\infty, \hat\theta_u]$ or $[\hat\theta_l, \infty)$, with $(0, \hat\theta_u]$ if $\theta \geq 0$. Figure A8.3 shows some examples of confidence intervals.

The concept of confidence intervals was introduced independently by J. Neyman and R. A. Fisher around 1930. In the following, three important cases for quality control and reliability tests are considered, see e.g. [A8.1] for more general cases.

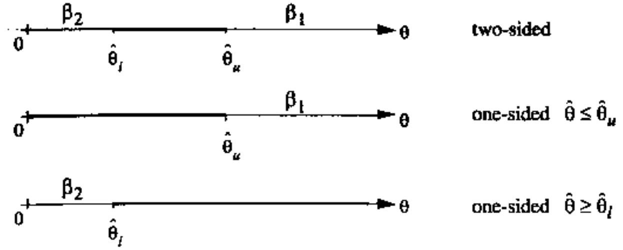


Figure A8.3 Examples of confidence intervals for $\theta \geq 0$

A8.2.2.1 Estimation of an Unknown Probability p

Consider a sequence of *Bernoulli trials* (Appendix A6.10.7) where a given event A can occur with constant probability p at each trial. The *binomial distribution*

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}$$

gives the probability that the event A will occur exactly k times in n independent trials. From the expression for p_k, it follows that

$$\Pr\{k_1 \leq \text{observations of } A \text{ in } n \text{ trials} \leq k_2 \mid p\} = \sum_{i=k_1}^{k_2} \binom{n}{i} p^i (1-p)^{n-i}. \qquad \text{(A8.36)}$$

In mathematical statistics, the parameter p is unknown. A *confidence interval* for p is sought, based on the observed number of occurrences of the event A in n Bernoulli trials. A first solution to this problem has been presented by Clopper and Pearson in [A8.7]. For given $\gamma = 1 - \beta_1 - \beta_2$ $(0 < \beta_1 < 1 - \beta_2 < 1)$ the following holds:

> ***If in n trials the event A has occurred k times, there is a probability nearly equal to (but not smaller than)*** $\gamma = 1 - \beta_1 - \beta_2$ ***that the confidence interval*** $[\hat{p}_l, \hat{p}_u]$ ***overlaps the true (unknown) probability p, with*** $\hat{p}_l$ ***and*** $\hat{p}_u$ ***given by***
>
> $$\sum_{i=k}^{n} \binom{n}{i} \hat{p}_l^i (1-\hat{p}_l)^{n-i} = \beta_2, \qquad \text{for } 0 < k < n, \qquad \text{(A8.37)}$$
>
> ***and***
>
> $$\sum_{i=0}^{k} \binom{n}{i} \hat{p}_u^i (1-\hat{p}_u)^{n-i} = \beta_1, \qquad \text{for } 0 < k < n; \qquad \text{(A8.38)}$$
>
> *for* $k = 0$ *take*
>
> $$\hat{p}_l = 0 \quad \text{and} \quad \hat{p}_u = 1 - \sqrt[n]{\beta_1}, \qquad \text{with} \quad \gamma = 1 - \beta_1, \qquad \text{(A8.39)}$$
>
> *and for* $k = n$ *take*
>
> $$\hat{p}_l = \sqrt[n]{\beta_2} \quad \text{and} \quad \hat{p}_u = 1, \qquad \text{with} \quad \gamma = 1 - \beta_2. \qquad \text{(A8.40)}$$

Considering that k is a random variable, $\hat{p}_l$ and $\hat{p}_u$ are *random variables*. According to the footnote on p. 436, it would be more correct (from a mathematical point of view) to compute from Eqs. (A8.37) and (A8.38) the quantities p_{kl} and p_{ku}, and then to set $\hat{p}_l = p_{kl}$ and $\hat{p}_u = p_{ku}$. For simplicity, this has been omitted here. Assuming p as a random variable, β_1 and β_2 would be the probabilities for p to be greater than $\hat{p}_u$ and smaller than $\hat{p}_l$, respectively (Fig. A8.3). The proof of Eqs. (A8.37) to (A8.40) is based on the monotonic property of the function

$$B_n(k, p) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}.$$

$B_n(k, p)$ decreases in p for fixed k and increases in k for fixed p (Fig. A8.4). Thus, for any $p > \hat{p}_u$ it follows that

$$B_n(k, p) < B_n(k, \hat{p}_u) = \beta_1.$$

For $p > \hat{p}_u$, the probability that the (random) number of observations in n trials will take one of the values $0, 1, \ldots, k$ is thus $< \beta_1$ (for $p > p'$ in Fig. A8.4, the statement would also be true for a $K > k$). This holds in particular for a number of observations equal to k and proves Eq. (A8.38). The proof of Eq. (A8.37) is similar.

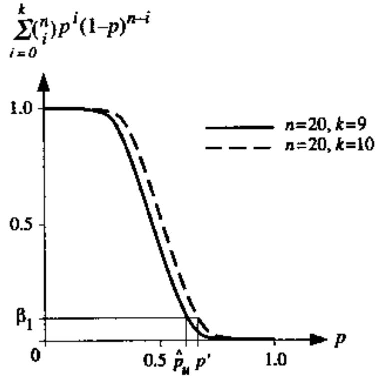


Figure A8.4 Graphs of the binomial distribution for n fixed and two values of k

To determine $\hat{p}_l$ and $\hat{p}_u$ as in Eqs. (A8.37) and (A8.38), a Table of the *Fisher distribution* (Appendix A9.4) or of the *Beta function* can be used. In practice however, for $\beta_1 = \beta_2 = (1-\gamma)/2$ and n sufficiently large one of the following *approximate solutions* is sufficient:

1. For $\min(np, n(1-p)) > 5$, a good estimate for $\hat{p}_l$ and $\hat{p}_u$ can be found using the *integral Laplace theorem (De Moivre-Laplace theorem)*. Rearranging Eq. (A6.149) and considering $\sum_{i=1}^{n} \delta_i = k$ yields

$$\lim_{n\to\infty} \Pr\{(\frac{k}{n}-p)^2 \le \frac{b^2 p(1-p)}{n}\} = \frac{2}{\sqrt{2\pi}}\int_0^b e^{-\frac{x^2}{2}} dx. \qquad (A8.41)$$

The right-hand side of Eq. (A8.41) is equal to the confidence level γ, i.e.

$$\frac{2}{\sqrt{2\pi}}\int_0^b e^{-\frac{x^2}{2}} dx = \gamma.$$

Thus, for a given γ the value of b can be obtained from a table of the normal distribution. b is the $(1+\gamma)/2$ quantile of the standard normal distribution $\Phi(t)$, i.e., $b = t_{(1+\gamma)/2}$ giving e.g. $b = 1.64$ for $\gamma = 0.9$ (Table A9.1). On the left-hand side of Eq. (A8.41), the expression

$$(\frac{k}{n}-p)^2 = \frac{b^2 p(1-p)}{n} \qquad (A8.42)$$

is the equation of the *confidence ellipse*. For given values of k, n, and b, confidence limits $\hat{p}_l$ and $\hat{p}_u$ can thus be determined as roots of Eq. (A8.42)

$$\hat{p}_{u,l} = \frac{k + 0.5b^2 \pm b\sqrt{k(1-k/n) + b^2/4}}{n + b^2}. \qquad (A8.43)$$

For the confidence limits obtained using Eq. (A8.43), $\beta_1 = \beta_2 = (1-\gamma)/2$. Figure A8.5 shows confidence ellipses for $\gamma = 0.9$ and different values of n.

2. For small values of n, confidence limits can be determined graphically from the envelopes of Eqs. (A8.37) and (A8.38) for $\beta_1 = \beta_2 = (1-\gamma)/2$. These curves are given in Fig. 7.1 for $\gamma = 0.8$ and $\gamma = 0.9$. For $n > 50$, the curves of Fig. 7.1 coincide with the confidence ellipses given by Eq. (A8.42).

One-sided confidence intervals can also be determined from the above values for $\hat{p}_l$ and $\hat{p}_u$. Figure A8.3 shows that

$$0 \le p \le \hat{p}_u, \text{ with } \gamma = 1-\beta_1 \quad \text{and} \quad \hat{p}_l \le p \le 1, \text{ with } \gamma = 1-\beta_2. \qquad (A8.44)$$

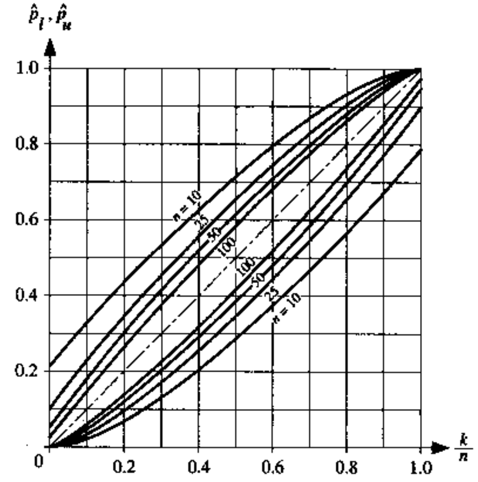


Figure A8.5 Confidence ellipses for a confidence level $\gamma = 0.9$ and for $n = 10, 25, 50$, and 100

Example A8.6

Using confidence ellipses, determine the confidence interval $[\hat{p}_l, \hat{p}_u]$ of an unknown probability p for the case $n = 50$, $k = 5$, and $\gamma = 0.9$.

Solution

Setting $n = 50$, $k = 5$, and $b = 1.64$ in Eq. (A8.43) yields the confidence interval [0.05, 0.19]. The corresponding one-sided confidence intervals would be given by $p \le 0.19$ or $p \ge 0.05$ with $\gamma = 0.95$.

The role of k/n and p in Eq. (A8.42) can be *reversed*, and Eq. (A8.42) can be used to solve a problem of *probability theory*, i.e. to compute, for a *given* probability γ ($\gamma = 1-\beta_1-\beta_2$ with $\beta_1 = \beta_2$), the limits k_1 and k_2 of the number of observations k in n *independent trials* (number k of defective items in a sample of size n for example) for *given* values of p and n

$$k_{2,1} = np \pm b\sqrt{np(1-p)}. \qquad (A8.45)$$

As in Eq. (A8.43), the quantity b in Eq. (A8.45) is the $(1+\gamma)/2$ quantile of the normal distribution (e.g. $b = 1.64$ for $\gamma = 0.9$ from Table A9.1). For a graphical solution, Fig. A8.5 can be used by taking the ordinate p as known and by reading k_1/n and k_2/n from the abscissa.

A8.2.2.2 Estimation of the Parameter λ of an Exponential Distribution for Fixed Test Duration T *(Type I censoring)* with Replacement

Consider now an item having a constant failure rate λ and assume that at each failure it will be *immediately replaced* by a new, statistically equivalent item, in a negligible replacement time (Appendix A7.2). Because of the *memoryless property* (constant failure rate), the number of failures in $(0, T]$ is Poisson-distributed and is given by (Eq. (A7.39)) $\Pr\{k \text{ failures in } (0,T] \mid \lambda\} = (\lambda T)^k e^{-\lambda T}/k!$. A maximum likelihood point estimate for λ follows then from Eq. (A8.30), with $n = 1$ and $m = \lambda T$, as

$$\hat{\lambda} = \frac{k}{T}. \qquad \text{(A8.46)}$$

A confidence interval estimation for the failure rate λ can thus be reduced to *estimating the confidence interval for the parameter $m = \lambda T$ of a Poisson distribution*. Considering Eqs. (A8.37) and (A8.38) and the similarity between the binomial and the Poisson distribution, the confidence limits $\hat{\lambda}_l$ and $\hat{\lambda}_u$ can be determined for a given $\gamma = 1 - \beta_1 - \beta_2$ $(0 < \beta_1 < 1 - \beta_2 < 1)$, from

$$\sum_{i=k}^{\infty} \frac{(\hat{\lambda}_l T)^i}{i!} e^{-\hat{\lambda}_l T} = \beta_2, \qquad \text{for } k > 0, \qquad \text{(A8.47)}$$

and

$$\sum_{i=0}^{k} \frac{(\hat{\lambda}_u T)^i}{i!} e^{-\hat{\lambda}_u T} = \beta_1, \qquad \text{for } k > 0; \qquad \text{(A8.48)}$$

for $k = 0$ takes

$$\hat{\lambda}_l = 0 \qquad \text{and} \qquad \hat{\lambda}_u = \frac{\ln(1/\beta_1)}{T}, \qquad \text{with} \quad \gamma = 1 - \beta_1, \qquad \text{(A8.49)}$$

On the basis of the known relationship to the chi-square (χ^2) distribution, the values $\hat{\lambda}_l$ and $\hat{\lambda}_u$ from Eqs. (A8.47) and (A8.48) can be obtained from of the *quantiles of the chi-square distribution* (Appendix A9.2). Thus,

$$\hat{\lambda}_l = \frac{\chi^2_{2k,\beta_2}}{2T}, \qquad \text{for} \quad k > 0, \qquad \text{(A8.50)}$$

and

$$\hat{\lambda}_u = \frac{\chi^2_{2(k+1),1-\beta_1}}{2T}, \qquad \text{for} \quad k \ge 0. \qquad \text{(A8.51)}$$

$\beta_1 = \beta_2 = (1-\gamma)/2$ is frequently used in practical applications, see Fig. 7.6 for a graphical solution of Eqs. (A8.47) and (A8.48) for this case.

One-sided confidence intervals are given as in the previous section by

$$0 \le \lambda \le \hat{\lambda}_u, \quad \text{with } \gamma = 1 - \beta_1 \quad \text{and} \quad \hat{\lambda}_l \le \lambda \le \infty, \quad \text{with } \gamma = 1 - \beta_2. \qquad \text{(A8.52)}$$

The situation considered in this section corresponds to that of a sampling plan with n elements *with replacement*, each of them with falure rate $\lambda' = \lambda/n$, terminated at a fixed test time T *(Type I censoring)*. This situation is statistically different from that presented in Section A8.2.2.3.

A8.2.2.3 Estimation of the Parameter λ of an Exponential Distribution for Fixed Number n of Failures *(Type II censoring)* without Replacement

Let $\tau_1, \ldots, \tau_n$ be independent random variables distributed according to a common distribution function $F(t) = \Pr\{\tau_i \le t\} = 1 - e^{-\lambda t}$, $i = 1, \ldots, n$. From Eq. (A7.38),

$$\Pr\{\tau_1 + \ldots + \tau_n \le t\} = 1 - \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t} = \frac{1}{(n-1)!} \int_0^{\lambda t} x^{n-1} e^{-x}\, dx \qquad \text{(A8.53)}$$

and thus

$$\Pr\{a < \tau_1 + \ldots + \tau_n \le b\} = \frac{1}{(n-1)!} \int_{a\lambda}^{b\lambda} x^{n-1} e^{-x}\, dx.$$

Setting $a = n(1-\varepsilon_2)/\lambda$ and $b = n(1+\varepsilon_1)/\lambda$ it follows that

$$\Pr\{\frac{1-\varepsilon_2}{\lambda} < \frac{\tau_1 + \ldots + \tau_n}{n} \le \frac{1+\varepsilon_1}{\lambda}\} = \frac{1}{(n-1)!} \int_{n(1-\varepsilon_2)}^{n(1+\varepsilon_1)} x^{n-1} e^{-x}\, dx. \qquad \text{(A8.54)}$$

Considering now $\tau_1, \ldots, \tau_n$ as a random sample of τ with $t_1, \ldots, t_n$ as observations, Eq. (A8.54) can be used to compute confidence limits $\hat{\lambda}_l$ and $\hat{\lambda}_u$ for the parameter λ. For a given confidence level $\gamma = 1 - \beta_1 - \beta_2$ $(0 < \beta_1 < 1 - \beta_2 < 1)$, this leads to

$$\hat{\lambda}_l = (1-\varepsilon_2)\hat{\lambda} \qquad \text{and} \qquad \hat{\lambda}_u = (1+\varepsilon_1)\hat{\lambda}, \qquad \text{(A8.55)}$$

with

$$\hat{\lambda} = \frac{n}{t_1 + \ldots + t_n} \qquad \text{(A8.56)}$$

and ε_1, ε_2 given by

$$\frac{1}{(n-1)!} \int_{n(1+\varepsilon_1)}^{\infty} x^{n-1} e^{-x}\, dx = \beta_1 \qquad \text{and} \qquad \frac{1}{(n-1)!} \int_0^{n(1-\varepsilon_2)} x^{n-1} e^{-x}\, dx = \beta_2. \qquad \text{(A8.57)}$$

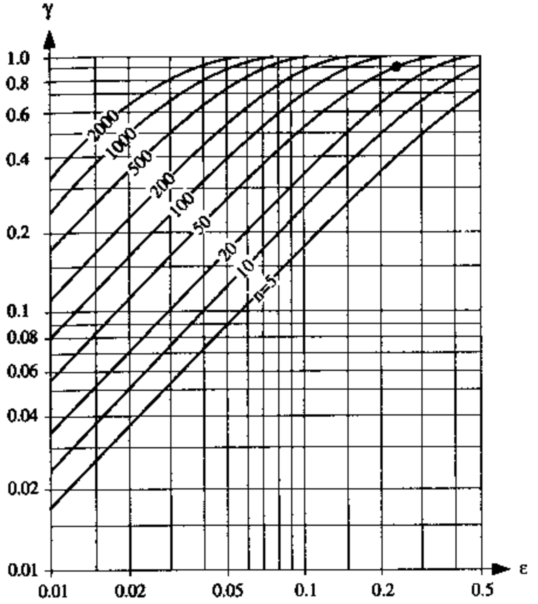


Figure A8.6 Probability γ that the interval $(1 \pm \varepsilon)\hat{\lambda}$ overlaps the true value of λ for the case of a fixed number n of failures $(\hat{\lambda} = n/(t_1 + \ldots + t_n)$, $\Pr\{\tau \le t\} = 1 - e^{-\lambda t})$

From this (Appendix A9.2), $\varepsilon_1 = (\chi^2_{2n,\,1-\beta_1})/2n - 1$ and $\varepsilon_2 = 1 - (\chi^2_{2n,\,\beta_2})/2n$, and thus

$$\hat{\lambda}_l = \frac{\chi^2_{2n,\,\beta_2}}{2(t_1 + \ldots + t_n)} \qquad \text{and} \qquad \hat{\lambda}_u = \frac{\chi^2_{2n,\,1-\beta_1}}{2(t_1 + \ldots + t_n)}. \tag{A8.58}$$

$\varepsilon_2 = 1$ or $\varepsilon_1 = \infty$ lead to *one-sided confidence intervals* $[0, \hat{\lambda}_u]$ or $[\hat{\lambda}_l, \infty)$. Figure A8.6 gives the graphical relationship between n, γ, and ε for the case $\varepsilon_1 = \varepsilon_2 = \varepsilon$.

The case considered here corresponds to the situation described in Example A8.2, i.e a sampling plan with n elements *without replacement*, terminated at the *occurrence of the n-th failure* (*Type II censoring*), which is statistically different from that presented in Section A8.2.2.2.

Example A8.7

For the case of a fixed number $n = 50$ of failures and $\gamma = 0.9$, determine the two-sided confidence interval for the parameter λ of an exponential distribution as a function of $\hat{\lambda}$.

Solution

Figure A8.6 yields $\varepsilon = 0.24$ and thus leads to the confidence interval $[0.76\hat{\lambda}, 1.24\hat{\lambda}]$.

A8.3 Testing Statistical Hypotheses

When testing a statistical hypothesis, the objective is to solve the following problem:

> *From one's own experience, the nature of the problem, or simply as a basic hypothesis, a specific null hypothesis H_0 is formulated for the statistical properties of the observed random variable; a rule is then sought which allows rejection or acceptance of H_0 on the basis of the observations made from a sample of the random variable under consideration.*

If R is the unknown reliability of an item, the following null hypotheses H_0 are for instance possible:

1a) H_0: $R = R_0$
1b) H_0: $R > R_0$
1c) H_0: $R < R_0$.

To test whether the failure-free operating time of an item is distributed according to an exponential distribution $F_0(t) = 1 - e^{-\lambda t}$ with unknown λ, or $F_0(t) = 1 - e^{-\lambda_0 t}$ with known λ_0, the following null hypotheses H_0 can be for instance formulated:

2a) H_0: the distribution function is $F_0(t)$
2b) H_0: the distribution function is different from $F_0(t)$
2c) H_0: $\lambda = \lambda_0$, provided the distribution is exponential
2d) H_0: $\lambda < \lambda_0$, provided the distribution is exponential
2e) H_0: the distribution function is $1 - e^{-\lambda t}$, parameter λ unknown.

It is usual to subdivide hypotheses into *parametric* (1a, 1b, 1c, 2c, 2d) and *non-parametric* ones (2a, 2b, and 2e). For each of these types, a distinction is also made between *simple hypotheses* (1a, 2a, 2c) and *composite hypotheses* (1b,1c,2b,2d,2e).

When testing a hypothesis, *two kinds of errors can occur* (Table A8.2):

- *type I error, i.e. the error of rejecting a true hypothesis* H_0, the probability of this error is denoted by α
- *type II error, i.e. the error of accepting a false hypothesis* H_0, the probability of this error is denoted by β (to find β, an *alternative hypothesis* H_1 is necessary, β is then the probability of accepting H_0 assuming H_1 is true).

If the sample space is divided into two complementary sets, $\mathcal{A}$ for acceptance and $\bar{\mathcal{A}}$ for rejection, the type I and type II errors are given by

$$\alpha = \Pr\{\text{sample in } \bar{\mathcal{A}} \mid H_0 \text{ true}\}, \tag{A8.59}$$

$$\beta = \Pr\{\text{sample in } \mathcal{A} \mid H_0 \text{ false } (H_1 \text{ true})\}. \tag{A8.60}$$

Both kinds of error are possible and cannot be minimized simultaneously (nondeterministic case). Often α is selected and a test is sought so that, for a given H_1, β will be minimized. It can be shown that such a test always exists if H_0 and H_1 are simple hypotheses [A8.22]. The following sections consider important procedures which apply in quality control and reliability tests, see Chapter 7 for applications.

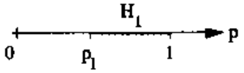
A8.3.1 Testing an Unknown Probability *p*

Let *A* be an event which can occur at every independent trial with the constant, unknown probability *p*. A rule (test plan) is sought which allows testing of the hypothesis

$$H_0: \ p < p_0 \qquad \text{(A8.61)}$$

against the alternative hypothesis

$$H_1: \ p > p_1 \qquad (p_1 \geq p_0) \qquad \text{(A8.62)}$$



The *type I error* should be nearly equal to (but not smaller than) α for $p = p_0$ (if possible less than α for $p < p_0$). The *type II error* should be nearly equal to (but not smaller than) β for $p = p_1$ (if possible less than β for $p > p_1$). Such a situation often occurs in practical applications, in particular in:

- *quality control*, where *p* refers to the fraction of defective items (defective probability),
- *reliability tests*, where it is usual to set $p = 1 - R$ (*R* = reliability) or $p = 1 - PA$ (*PA* = steady-state availability).

In both cases, α is the *producer's risk* and β the *consumer's risk*. The two most frequently used procedures for testing hypotheses defined by the relations (A8.61) and (A8.62), with $p_1 > p_0$, are the simple two-sided sampling plan and the sequential test.

Table A8.2 Possible errors when testing a hypothesis

	H_0 is rejected	H_0 is accepted
H_0 is true	false → type I error (α)	correct
H_0 is false (H_1 is true)	correct	false → type II error (β)

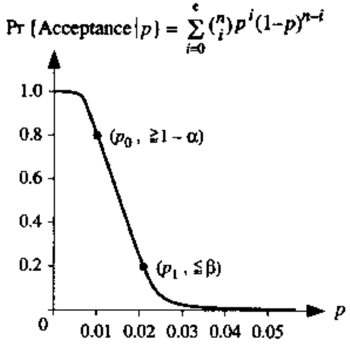


Figure A8.7 Operating characteristic (probability of acceptance) as a function of *p* for fixed *n* and *c* ($p_0 = 1\%$, $p_1 = 2\%$, $\alpha \approx \beta \approx 0.185$, $n = 462$, $c = 6$)

A8.3.1.1 Simple Two-sided Sampling Plan

The rule for the *simple two-sided sampling plan* (simple two-sided test) is:

1. For given p_0, p_1, α, and β ($0 < \alpha < 1 - \beta < 1$), compute the smallest integers *c* and *n* which satisfy

$$\sum_{i=0}^{c}\binom{n}{i}p_0^i(1-p_0)^{n-i} \geq 1-\alpha \qquad \text{(A8.63)}$$

and

$$\sum_{i=0}^{c}\binom{n}{i}p_1^i(1-p_1)^{n-i} \leq \beta. \qquad \text{(A8.64)}$$

2. Perform *n* independent trials (Bernoulli trials), determine the number *k* in which the event *A* (component defective for example) has occurred, and

   - reject H_0, if $k > c$
   - accept H_0, if $k \leq c$. (A8.65)

As in the case of Eqs. (A8.37) and (A8.38), the proof of the above rule is based on the monotonic property of

$$B_n(c,p) = \sum_{i=0}^{c}\binom{n}{i}p^i(1-p)^{n-i}.$$

For known n, c, and p, $B_n(c,p)$ gives the probability of having *up to c defectives in a sample of size n*. Thus, assuming H_0 true, it follows that the probability of rejecting H_0 (i.e. the probability of having more than c defectives in a sample of size n) is smaller than α, then

$$\Pr\{\text{rejection of } H_0 \mid H_0 \text{ true}\} = \sum_{i=c+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \Big|_{p<p_0} < \alpha .$$

Similarly, if H_1 is true ($p > p_1$), it follows that the probability of accepting H_0 is smaller than β, then

$$\Pr\{\text{acceptance of } H_0 \mid H_1 \text{ true}\} = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \Big|_{p>p_1} < \beta .$$

The assumptions made with Eqs. (A8.63) and (A8.64) are thus satisfied. As shown by the above inequalities, the type I error and the type II error are in this case $< \alpha$ for $p < p_0$ and $< \beta$ for $p > p_1$, respectively. Figure A8.7 presents the results for $p_0 = 1\%$, $p_1 = 2\%$, and $\alpha = \beta \lessapprox 20\%$. The curve of Fig. A8.7 is known as the *operating characteristic* (OC). If p_0 and p_1 are small (up to a few %) or close to 1, the *Poisson approximation* (Eq. (A6.129)) can be used.

A8.3.1.2 Sequential Test

Assume that in a two-sided sampling plan with $n = 50$ and $c = 2$, $k = 3$ occurs at the 12th trial. Since $k > c$, the hypothesis H_0 will be rejected as per procedure (A8.65), independent of how often the event A will occur during the remaining 38 trials. This example brings up the question of whether a plan can be established for testing H_0 in which *no unnecessary trials* (the remaining 38 in the above example) have to be performed. To solve this problem, A. Wald proposed in 1947 the *sequential test* [A8.30]. For this test, one element after another is taken from the lot and tested. Depending upon the actual frequency of the observed event, the decision is made to either

- reject H_0
- accept H_0
- perform a further trial.

The testing procedure can be described as follows (Fig. A8.8):

> *In a system of cartesian coordinates, the number n of trials is recorded on the abscissa and the number k of trials in which the event A occurred on the ordinate; the test is stopped with acceptance or rejection as soon as the resulting staircase curve $k = f(n)$ crosses the acceptance or rejection line given in the cartesian coordinates for specified values of p_0, p_1, α and β.*

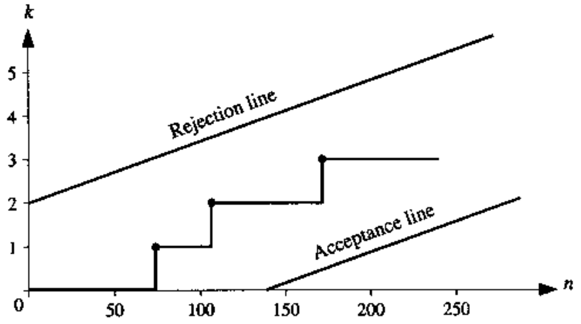


Figure A8.8 Sequential test for $p_0 = 1\%$, $p_1 = 2\%$, and $\alpha = \beta = 20\%$

The acceptance and rejection lines can be determined from:

$$\text{Acceptance line:} \quad k = an - b_1, \tag{A8.66}$$

$$\text{Rejection line:} \quad k = an + b_2, \tag{A8.67}$$

with [A8.30]

$$a = \frac{\ln\frac{1-p_0}{1-p_1}}{\ln\frac{p_1}{p_0} + \ln\frac{1-p_0}{1-p_1}}, \qquad b_1 = \frac{\ln\frac{1-\alpha}{\beta}}{\ln\frac{p_1}{p_0} + \ln\frac{1-p_0}{1-p_1}}, \qquad b_2 = \frac{\ln\frac{1-\beta}{\alpha}}{\ln\frac{p_1}{p_0} + \ln\frac{1-p_0}{1-p_1}}. \tag{A8.68}$$

Figure A8.8 shows the acceptance and the rejection line for $p_0 = 1\%$, $p_1 = 2\%$, and $\alpha = \beta = 20\%$. Practical remarks related to sequential tests are given in Sections 7.1.2.2 and 7.2.2.2.

A8.3.1.3 Simple One-sided Sampling Plan

One-sided sampling plans (one sided tests) are used in practical applications when $H_0 : p < p_0$ and $H_1 : p > p_0$, or $H_0 : p < p_1$ and $H_1 : p > p_1$ are assumed. These plans are introduced here and discussed in Section 7.1.3.

Setting $p_1 = p_0$ in the relationship (A8.62), i.e. testing

$$H_0 : \quad p < p_0 \tag{A8.69}$$

against

$$H_1 : \quad p > p_0 \tag{A8.70}$$

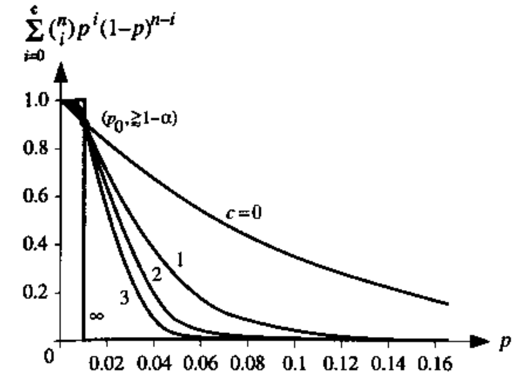


Figure A8.9 Operating characteristics for $p_0 = 1\%$, $\alpha \approx 0.1$ and $c = 0$ ($n = 10$), $c = 1$ ($n = 53$), $c = 2$ ($n = 110$), $c = 3$ ($n = 174$) und $c = \infty$

with one c, n pair (for $c = 0, 1, \ldots$) from Eq. (A8.63) and the test procedure (A8.65), the *type II error* can become very large and reach approximately the value $1-\alpha$ for $p = p_0$. Depending upon the value selected for $c = 0, 1, \ldots$ and that computed for n (the smallest n which satisfies Eq. (A8.63)), different plans (pairs of c, n) are possible. Each of these plans yields different type II errors. Figure A8.9 shows this for some values of c (the type II error is equal to the ordinate of the operating characteristic for $p > p_0$). It is common usage in such cases to define

$$p_0 = \text{AQL}, \tag{A8.71}$$

where AQL stands for *Acceptable Quality Level.* The above considerations show that with the choice of only p_0 and α (instead of p_0, p_1, α, and β) the *producer can realize an advantage*, particularly if small values of c are used.

On the other hand, setting $p_0 = p_1$ in the relationship (A8.62), i.e. testing

$$H_0: \quad p < p_1 \tag{A8.72}$$

against

$$H_1: \quad p > p_1 \tag{A8.73}$$

with one c, n pair (for $c = 0, 1, \ldots$) from Eq. (A8.64) and the test procedure (A8.65), the *type I error* can become very large and reach approximately the value $1-\beta$ for $p = p_1$. Depending upon the value selected for $c = 0, 1, \ldots$ and that computed for n (the largest n which satisfies Eq.(A8.64)), different plans (pairs of c, n) are possible.

Considerations here are similar to those of the previous case, where only p_0 and α were selected. For small values of c the *consumer can realize an advantage.* It is common usage to define

$$p_1 = \text{LTPD}, \tag{A8.74}$$

where LTPD stands for *Lot Tolerance Percent Defective.*

A8.3.2 Goodness-of-fit Tests for Completely Defined $\mathrm{F}_0(t)$

Goodness-of-fit tests belong to a particular class of hypothesis testing [A8.10]. Let $t_1, \ldots, t_n$ be n independent observations of a random variable τ, a rule is sought to test the null hypothesis

$$H_0: \quad \text{the distribution function of } \tau \text{ is } \mathrm{F}_0(t), \tag{A8.75}$$

against the alternative hypothesis

$$H_1: \quad \text{the distribution function of } \tau \text{ is not } \mathrm{F}_0(t), \tag{A8.76}$$

where $\mathrm{F}_0(t)$ is, in this section, *completely defined.* Among other possibilities, the Kolmogorov-Smirnov and chi-square (χ^2) tests are frequently used to solve this problem. Both tests are based upon comparison of the *empirical distribution function* $\hat{\mathrm{F}}_n(t)$ with the postulated distribution function $\mathrm{F}_0(t)$.

1. The *Kolmogorov-Smirnov test* uses the statistics

   $$D_n = \sup_{-\infty<t<\infty} |\hat{\mathrm{F}}_n(t) - \mathrm{F}_0(t)|$$

   defined in Appendix A8.1.3. If τ is continuous, the distribution of $\hat{\mathrm{F}}_n(t)$ under the hypothesis H_0 is independent of the type of $\mathrm{F}_0(t)$. For a given *type I error* α, the hypothesis H_0 must be rejected for

   $$D_n > y_{1-\alpha},$$

where $y_{1-\alpha}$ is defined by

$$\Pr\{D_n > y_{1-\alpha} \mid H_0 \text{ is true}\} = \alpha. \tag{A8.77}$$

Values for $y_{1-\alpha}$ are given in Tables A8.1 and A9.5. However, in general nothing can be said about the risk of accepting a false hypothesis H_0 (to compute the type II error β, an alternative hypothesis H_1 must be assumed). Figure A8.10 illustrates the Kolmogorov-Smirnov test with hypothesis H_0 not rejected.

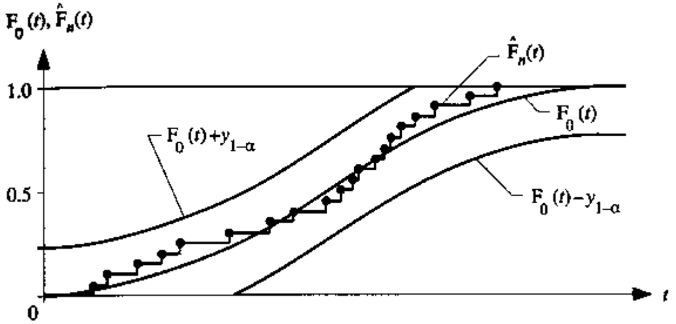


Figure A8.10 Kolmogorov-Smirnov test ($n = 20$, $\alpha = 20\%$)

2. The *chi-square (χ^2) goodness-of-fit test* starts from a selected partition $(a_1, a_2]$, $(a_2, a_3]$, ..., $(a_k, a_{k+1}]$ of the set of possible values of τ and uses the statistics

$$X_n^2 = \sum_{i=1}^{k} \frac{(k_i - n\,p_i)^2}{n\,p_i} = \sum_{i=1}^{k} \frac{k_i^2}{n\,p_i} - n, \tag{A8.78}$$

where

$$k_i = n(\hat{F}_n(a_{i+1}) - \hat{F}_n(a_i)) \tag{A8.79}$$

are the number of observations (realizations of τ) in $(a_i, a_{i+1}]$ and

$$n\,p_i = n(F_0(a_{i+1}) - F_0(a_i)) \tag{A8.80}$$

are the expected number of observations in $(a_i, a_{i+1}]$ under the hypothesis H_0. K. Pearson [A8.26] has shown that *the asymptotic distribution of X_n^2 under H_0 for $n \to \infty$ is a χ^2 distribution with $k-1$ degrees of freedom.* Thus for given type I error α,

$$\lim_{n\to\infty} \Pr\{X_n^2 > \chi^2_{k-1,1-\alpha} \mid H_0 \text{ true}\} = \alpha \tag{A8.81}$$

holds, and the hypothesis H_0 must be *rejected* if

$$X_n^2 > \chi^2_{k-1,1-\alpha}. \tag{A8.82}$$

$\chi^2_{k-1,1-\alpha}$ is the $(1-\alpha)$ *quantile* of the χ^2 distribution with $k-1$ degrees of freedom. The classes $(a_1, a_2]$, $(a_2, a_3]$, ..., $(a_k, a_{k+1}]$ are to be chosen *before* the test is performed in such a way that all p_i are approximately equal. Convergence is generally good, even for relatively small values of n ($n\,p_i \ge 5$).

A8.3.3 Goodness-of-fit Tests for a Distribution $F_0(t)$ with Unknown Parameters

The Kolmogorov-Smirnov test can be used in some cases where the underlying distribution function $F_0(t)$ is not completely known (parameter unknown). However, the quantities $y_{1-\alpha}$ must be calculated for each type of distribution.

The chi-square (χ^2) test offers a more general approach. Let $F_0(t)$ be the assumed distribution function, known up to the parameters $\theta_1, \dots, \theta_r$. If

- the unknown parameters $\theta_1, \dots, \theta_r$ are estimated according to the maximum likelihood method on the basis of the observed frequencies k_i (using the multinomial distribution given by Eq. (A6.124)), i.e. from the following system of r algebraic equations (see Example A8.8)

$$\sum_{i=1}^{k} \frac{k_i}{p_i(\theta_1, \dots, \theta_r)} \cdot \frac{\partial p_i(\theta_1, \dots, \theta_r)}{\partial \theta_j}\bigg|_{\theta_j = \hat{\theta}_j} = 0, \qquad j = 1, \dots, r, \tag{A8.83}$$

  with $p_i = F_0(a_{i+1}, \theta_1, \dots, \theta_r) - F_0(a_i, \theta_1, \dots, \theta_r) > 0$, $p_1 + \dots + p_k = 1$, and $k_1 + \dots + k_k = n$,

- $\dfrac{\partial p_i}{\partial \theta_j}$ and $\dfrac{\partial^2 p_i}{\partial \theta_j \partial \theta_m}$ exist ($i = 1, \dots, k$; $j, m = 1, \dots, r < k-1$),

- the matrix with elements $\dfrac{\partial p_i}{\partial \theta_j}$ is of rank r,

then the statistics

$$\hat{X}_n^2 = \sum_{i=1}^{k} \frac{(k_i - n\,\hat{p}_i)^2}{n\,\hat{p}_i} = \sum_{i=1}^{k} \frac{k_i^2}{n\,\hat{p}_i} - n, \tag{A8.84}$$

computed with $\hat{p}_i = F_0(a_{i+1}, \hat{\theta}_1, \dots, \hat{\theta}_r) - F_0(a_i, \hat{\theta}_1, \dots, \hat{\theta}_r)$, has under H_0 asymptotically for $n \to \infty$ a *χ^2 distribution with $k-1-r$ degrees of freedom* [A8.15 (1924)]. Thus, for a given *type I error* α,

$$\lim_{n\to\infty} \Pr\{\hat{X}_n^2 > \chi^2_{k-1-r,1-\alpha} \mid H_0 \text{ true}\} = \alpha, \tag{A8.85}$$

holds, and the hypothesis H_0 must be *rejected* if

$$\hat{X}_n^2 > \chi^2_{k-1-r,1-\alpha}. \tag{A8.86}$$

$\chi^2_{k-1-r,1-\alpha}$ is the $(1-\alpha)$ *quantile* of the χ^2 distribution with $k-1-r$ *degrees of freedom*. It should be noted that the computation of the parameters $\theta_1, \dots, \theta_r$ directly from the observations $t_1, \dots, t_n$ is not allowed [A8.8, A8.10].

Example A8.8
Prove Eq. (A8.83).

Solution

The observed frequencies $k_1, \ldots, k_k$ in the classes $(a_1, a_2]$, $(a_2, a_3]$, …, $(a_k, a_{k+1}]$ result from n trials, where each observation falls into one of the classes $(a_i, a_{i+1}]$ with probability $p_i = \mathrm{F}_0(a_{i+1}, \theta_1, \ldots, \theta_r) - \mathrm{F}_0(a_i, \theta_1, \ldots, \theta_r), i = 1, \ldots, k$. The *multinomial distribution* applies. Taking into account Eq. (A6.124), the likelihood function (Eq. (A8.23)) becomes

$$\mathrm{L}(p_1, \ldots, p_k) = \frac{n!}{k_1! \ldots k_k!} p_1^{k_1} \ldots p_k^{k_k} \tag{A8.87}$$

or

$$\ln \mathrm{L}(p_1, \ldots, p_k) = \ln \frac{n!}{k_1! \ldots k_k!} + k_1 \ln p_1 + \ldots + k_k \ln p_k, \tag{A8.88}$$

with $p_i = p_i(\theta_1, \ldots, \theta_r)$, $p_1 + \ldots + p_k = 1$ and $k_1 + \ldots + k_k = n$. Equation (A8.83) is then obtained from $\dfrac{\partial \ln \mathrm{L}}{\partial \theta_j} = 0$ for $\theta_j = \hat{\theta}_j$ and $j = 1, \ldots, r$.

A9 Tables and Charts

A9.1 Standard Normal Distribution

Definition: $\Phi(t) = \Pr\{\tau \le t\} = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx, \qquad -\infty < t < \infty$

Parameters: $\mathrm{E}[\tau] = 0, \quad \mathrm{Var}[\tau] = 1, \quad$ Modal value $= \mathrm{E}[\tau]$

Properties:
- $\Phi(0) = 0.5, \quad \Phi(-t) = 1 - \Phi(t)$
- for $\mathrm{E}[\tau] = m$ and $\mathrm{Var}[\tau] = \sigma^2$ it is

$$\mathrm{F}(t) = \Pr\{\tau \le t\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{(y-m)^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t-m}{\sigma}} e^{-\frac{x^2}{2}} dx = \Phi(\tfrac{t-m}{\sigma})$$

Table A9.1 Standard normal distribution $\Phi(t)$ for $t = 0.00 - 2.99$ $(\mathrm{E}[\tau] = 0, \mathrm{Var}[\tau] = 1)$

t	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9820	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

Examples: $\Pr\{\tau \le 2.33\} = 0.9901$; $\Pr\{\tau \le -1\} = 1 - \Pr\{\tau \le 1\} = 1 - 0.8413 = 0.1587$;
$\Pr\{-1 < \tau \le 1\} = 1 - 2\Pr\{\tau \le -1\} = 1 - 2(1 - \Pr\{\tau \le 1\}) = 2\Pr\{\tau \le 1\} - 1 = 0.6826$

A9.2 χ^2-Distribution (Chi-Square Distribution)

Definition: $F(t) = \Pr\{\chi_\nu^2 \le t\} = \frac{1}{2^{\frac{\nu}{2}}\,\Gamma(\frac{\nu}{2})}\int_0^t x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}\,dx,$

$t \ge 0, \quad \nu = 1, 2, \ldots$ (degrees of freedom)

Parameters: $E[\chi_\nu^2] = \nu, \quad \mathrm{Var}[\chi_\nu^2] = 2\nu, \quad$ Modal value $= \nu - 2 \quad (\nu > 2)$

Relationsships:

- Normal distribution: $\chi_\nu^2 = \frac{1}{\sigma^2}\sum_{i=1}^{\nu}(\xi_i - m)^2,$

  $\xi_1, \ldots, \xi_n$ independent, normal distrib.

  with $E[\xi_i] = m$ and $\mathrm{Var}[\xi_i] = \sigma^2$

- Poisson disdtribution: $\sum_{i=0}^{\frac{\nu}{2}-1} \frac{(\frac{t}{2})^i}{i!} e^{-\frac{t}{2}} = 1 - F(t), \qquad \nu = 2, 4, \ldots$

- Incomplete Gamma function: $\gamma(\frac{\nu}{2}, \frac{t}{2}) = F(t)\,\Gamma(\frac{\nu}{2})$

Table A9.2 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, and 0.975 quantile of the χ^2 distribution ($t_{\nu,q}$ for which $F(t_{\nu,q}) = q$; $t_{\nu,q} \approx (x + \sqrt{2\nu - 1})^2/2$ for $\nu > 100$)

ν \ q	0.05	0.10	0.20	0.40	0.60	0.80	0.90	0.95	0.975
1	0.0039	0.0158	0.0642	0.275	0.708	1.642	2.706	3.841	5.024
2	0.103	0.211	0.446	1.022	1.833	3.219	4.605	5.991	7.378
3	0.352	0.584	1.005	1.869	2.946	4.642	6.251	7.815	9.348
4	0.711	1.064	1.649	2.753	4.045	5.989	7.779	9.488	11.143
5	1.145	1.610	2.343	3.655	5.132	7.289	9.236	11.070	12.833
6	1.635	2.204	3.070	4.570	6.211	8.558	10.645	12.592	14.449
7	2.167	2.833	3.822	5.493	7.283	9.803	12.017	14.067	16.013
8	2.733	3.490	4.594	6.423	8.351	11.030	13.362	15.507	17.535
9	3.325	4.168	5.380	7.357	9.414	12.242	14.684	16.919	19.023
10	3.940	4.865	6.179	8.295	10.473	13.442	15.987	18.307	20.483
11	4.575	5.578	6.989	9.237	11.530	14.631	17.275	19.675	21.920
12	5.226	6.304	7.807	10.182	12.584	15.812	18.549	21.026	23.337
13	5.892	7.042	8.634	11.129	13.636	16.985	19.812	22.362	24.736
14	6.571	7.790	9.467	12.078	14.685	18.151	21.064	23.685	26.119
15	7.261	8.547	10.307	13.030	15.733	19.311	22.307	24.996	27.488
16	7.962	9.312	11.152	13.983	16.780	20.465	23.542	26.296	28.845
17	8.672	10.085	12.002	14.937	17.824	21.615	24.769	27.587	30.191
18	9.390	10.865	12.857	15.893	18.868	22.760	25.989	28.869	31.526
19	10.117	11.651	13.716	16.850	19.910	23.900	27.204	30.144	32.852
20	10.851	12.443	14.578	17.809	20.951	25.038	28.412	31.410	34.170
22	12.338	14.041	16.314	19.729	23.031	27.301	30.813	33.924	36.781
24	13.848	15.659	18.062	21.652	25.106	29.553	33.196	36.415	39.364
26	15.379	17.292	19.820	23.579	27.179	31.795	35.563	38.885	41.923
28	16.928	18.939	21.588	25.509	29.249	34.027	37.916	41.337	44.461
30	18.493	20.599	23.364	27.442	31.316	36.250	40.256	43.773	46.979
40	26.509	29.051	32.345	37.134	41.622	47.269	51.805	55.758	59.342
60	43.188	46.459	50.641	56.620	62.135	68.972	74.397	79.082	83.298
80	60.391	64.278	69.207	76.188	82.566	90.405	96.578	101.879	106.629
100	77.929	82.358	87.945	95.808	102.946	111.667	118.498	124.342	129.561
x	-1.645	-1.282	-0.841	-0.253	0.253	0.841	1.282	1.645	1.960

Examples: $F(t_{16,0.9}) = 0.9 \rightarrow t_{16,0.9} = 23.542; \quad \sum_{i=0}^{8}\frac{13^i}{i!}e^{-13} = 1 - F(26)$ for $\nu = 18 \approx 0.0998$

A9.3 *t*-Distribution (Student distribution)

Definition: $F(t) = \Pr\{t \le t\} = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\int_{-\infty}^{t}(1 + \frac{x^2}{\nu})^{-\frac{\nu+1}{2}}\,dx,$

$-\infty < t < \infty, \qquad \nu = 1, 2, \ldots$ (degrees of freedom)

Parameters: $E[t] = 0, \qquad \mathrm{Var}[t] = \frac{\nu}{\nu - 2} \quad (\nu > 2), \qquad$ Modal value $= 0$

Properties: $F(0) = 0.5, \qquad F(-t) = 1 - F(t)$

Relationsships:

- Normal distribution and χ_ν^2 distribution: $t = \frac{\xi}{\sqrt{\chi_\nu^2/\nu}}$

  ξ is normal distributed with $E[\xi] = 0$ and $\mathrm{Var}[\xi] = 1$; χ_ν^2 is χ^2 distributed with ν degrees of freedom, ξ and χ_ν^2 indep.

- Cauchy distribution ($\alpha = 0$ and $\beta = 1$): $F(t)$ with $\nu = 1$

Table A9.3 0.7, 0.8, 0.9, 0.95, 0.975, 0.99, 0.995, 0.999 quantile of the *t* distribution ($t_{\nu,q}$ for which $F(t_{\nu,q}) = q$)

ν \ q	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999
1	0.7265	1.3764	3.0777	6.3138	12.7062	31.8207	63.6574	318.3088
2	0.6172	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271
3	0.5844	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145
4	0.5686	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732
5	0.5594	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934
6	0.5534	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076
7	0.5491	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853
8	0.5459	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008
9	0.5435	0.8834	1.3839	1.8331	2.2622	2.8214	3.2498	4.2968
10	0.5415	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437
11	0.5399	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247
12	0.5386	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296
13	0.5375	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520
14	0.5366	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874
15	0.5357	0.8662	1.3406	1.7531	2.1315	2.6025	2.9467	3.7328
16	0.5350	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862
17	0.5344	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458
18	0.5338	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105
19	0.5333	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794
20	0.5329	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518
22	0.5321	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050
24	0.5314	0.8569	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668
26	0.5309	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350
28	0.5304	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082
30	0.5300	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852
40	0.5286	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069
60	0.5272	0.8477	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317
80	0.5265	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953
100	0.5261	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737
∞	0.5240	0.8410	1.2820	1.6450	1.9600	2.3260	2.5760	3.0900

Examples: $F(t_{16,0.9}) = 0.9 \rightarrow t_{16,0.9} = 1.3368; \quad F(t_{16,0.1}) = 0.1 \rightarrow t_{16,0.1} = -1.3368$

A9.4 $\boldsymbol{F}$-Distribution (Fisher distribution)

Definition: $$\mathrm{F}(t) = \Pr\{\boldsymbol{F} \le t\} = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}\,\nu_1^{\frac{\nu_1}{2}}\,\nu_2^{\frac{\nu_2}{2}}\int_0^t \frac{x^{\frac{\nu_1-2}{2}}}{(\nu_1 x+\nu_2)^{\frac{\nu_1+\nu_2}{2}}}\,dx,$$

$t \ge 0, \quad \nu_1, \nu_2 = 1, 2, \ldots$ (degrees of freedom)

Parameters: $$\mathrm{E}[\boldsymbol{F}] = \frac{\nu_2}{\nu_2-2} \quad (\nu_2 > 2), \quad \mathrm{Var}[\boldsymbol{F}] = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} \quad (\nu_2 > 4),$$

$$\text{Modal value} = \frac{(\nu_1-2)\nu_2}{2\nu_1^2+\nu_2} \qquad (\nu_1 > 2)$$

Relationships: • χ^2 distribution: $\boldsymbol{F} = \dfrac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$,

for $\chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$ see also the first relationship of Section A9.2

• Binomial distribution: $\sum_{i=0}^{k}\binom{n}{i}p^i(1-p)^{n-i} = 1-\mathrm{F}(\frac{n-k}{k+1}\cdot\frac{p}{1-p})$

with $\nu_1 = 2(k+1)$ and $\nu_2 = 2(n-k)$

Table A9.4a 0.90 quantile of the $\boldsymbol{F}$ distribution
($t_{\nu_1,\nu_2,0.9}$ for which $\mathrm{F}(t_{\nu_1,\nu_2,0.9}) = 0.9$)

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10	20	50	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	59.44	60.19	61.74	62.69	63.33
2	8.526	9.000	9.162	9.243	9.293	9.325	9.367	9.392	9.441	9.471	9.491
3	5.538	5.462	5.391	5.343	5.309	5.285	5.252	5.230	5.184	5.155	5.134
4	4.545	4.325	4.191	4.107	4.051	4.010	3.955	3.920	3.844	3.795	3.761
5	4.060	3.780	3.619	3.520	3.453	3.404	3.339	3.297	3.207	3.147	3.105
6	3.776	3.463	3.289	3.181	3.107	3.055	2.983	2.937	2.836	2.770	2.722
7	3.589	3.257	3.074	2.960	2.883	2.827	2.752	2.702	2.595	2.523	2.471
8	3.458	3.113	2.924	2.806	2.726	2.668	2.589	2.538	2.425	2.348	2.293
9	3.360	3.006	2.813	2.693	2.611	2.551	2.469	2.416	2.298	2.218	2.159
10	3.285	2.924	2.728	2.605	2.522	2.461	2.377	2.323	2.201	2.117	2.055
12	3.176	2.807	2.605	2.480	2.394	2.331	2.245	2.188	2.060	1.970	1.904
14	3.102	2.726	2.522	2.395	2.307	2.243	2.154	2.095	1.962	1.869	1.797
16	3.048	2.668	2.462	2.333	2.244	2.178	2.088	2.028	1.891	1.793	1.718
18	3.007	2.624	2.416	2.286	2.196	2.130	2.038	1.977	1.837	1.736	1.657
20	2.975	2.589	2.380	2.249	2.158	2.091	1.998	1.937	1.794	1.690	1.607
30	2.881	2.489	2.276	2.142	2.049	1.980	1.884	1.819	1.667	1.552	1.456
50	2.809	2.412	2.197	2.061	1.966	1.895	1.796	1.729	1.568	1.441	1.327
100	2.756	2.356	2.139	2.002	1.906	1.834	1.732	1.663	1.494	1.355	1.214
1000	2.711	2.308	2.089	1.950	1.853	1.780	1.676	1.605	1.428	1.273	1.060
∞	2.705	2.303	2.084	1.945	1.847	1.774	1.670	1.599	1.421	1.263	1.000

Example: $\nu_1 = 10,\ \nu_2 = 16 \rightarrow t_{10,16,0.9} = 2.028$

Table A9.4b 0.95 quantile of the $\boldsymbol{F}$ distribution (see Table A9.4a)

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	8	10	20	50	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	241.9	248.0	251.8	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.40	19.45	19.48	19.50
3	10.13	9.552	9.277	9.117	9.013	8.941	8.845	8.785	8.660	8.581	8.526
4	7.709	6.944	6.591	6.388	6.256	6.163	6.041	5.964	5.802	5.699	5.628
5	6.608	5.786	5.409	5.192	5.050	4.950	4.818	4.735	4.558	4.444	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.147	4.060	3.874	3.754	3.669
7	5.591	4.737	4.347	4.120	3.971	3.866	3.726	3.636	3.444	3.319	3.230
8	5.318	4.459	4.066	3.838	3.687	3.580	3.438	3.347	3.150	3.020	2.928
9	5.117	4.256	3.863	3.633	3.482	3.374	3.230	3.137	2.936	2.803	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	3.072	2.978	2.774	2.637	2.538
12	4.747	3.885	3.490	3.259	3.106	2.996	2.849	2.753	2.544	2.401	2.296
14	4.600	3.739	3.344	3.112	2.958	2.848	2.699	2.602	2.388	2.240	2.131
16	4.494	3.634	3.239	3.007	2.852	2.741	2.591	2.493	2.276	2.124	2.010
18	4.414	3.555	3.160	2.928	2.773	2.661	2.510	2.412	2.191	2.035	1.917
20	4.351	3.493	3.098	2.866	2.711	2.599	2.447	2.348	2.124	1.966	1.843
30	4.171	3.316	2.922	2.690	2.534	2.420	2.266	2.165	1.932	1.761	1.622
50	4.034	3.183	2.790	2.557	2.400	2.286	2.130	2.026	1.784	1.599	1.438
100	3.936	3.087	2.695	2.463	2.305	2.191	2.032	1.927	1.676	1.477	1.283
1000	3.851	3.005	2.614	2.381	2.223	2.108	1.948	1.840	1.581	1.363	1.078
∞	3.841	2.996	2.605	2.372	2.214	2.099	1.938	1.831	1.570	1.350	1.000

A9.5 Table for the Kolmogorov-Smirnov Test

$$D_n = \sup_{-\infty<t<\infty} \left| \hat{\mathrm{F}}_n(t) - \mathrm{F}_0(t) \right|, \qquad \hat{\mathrm{F}}_n(t) = \text{empirical distribution function}$$
$$\mathrm{F}_0(t) = \text{postulated continuous distrib. function}$$

Table A9.5 $1-\alpha$ quantile of the distribution function of D_n $(\Pr\{D_n \le y_{1-\alpha} \mid H_0 \text{ true}\} = 1-\alpha)$

n	$\alpha = 0.20$	0.10	0.05	0.02	0.01	n	$\alpha = 0.20$	0.10	0.05	0.02	0.01
1	0.900	0.950	0.975	0.900	0.993	21	0.226	0.259	0.287	0.321	0.344
2	684	776	842	900	929	22	221	253	281	314	337
3	565	636	708	785	829	23	216	247	275	307	330
4	493	565	624	689	734	24	212	242	269	301	323
5	447	509	563	627	669	25	208	238	264	295	317
6	410	468	519	577	617	26	204	233	259	290	311
7	381	436	483	538	576	27	200	229	254	284	305
8	358	410	454	507	542	28	197	225	250	279	300
9	339	387	430	480	513	29	193	221	246	275	295
10	323	369	409	457	489	30	190	218	242	270	290
11	309	352	291	437	468	32	184	211	234	262	281
12	296	338	375	419	449	34	179	205	227	254	273
13	285	325	361	404	432	36	174	199	221	247	265
14	275	314	349	390	418	38	170	194	215	241	258
15	266	304	338	377	404	40	165	189	210	235	252
16	258	295	327	366	392	42	162	185	205	229	246
17	250	286	318	355	381	44	158	181	201	224	241
18	244	279	300	346	371	46	155	177	196	219	235
19	237	271	301	337	361	48	151	173	192	215	231
20	232	265	294	329	352	50	148	170	188	211	226
						for $n>50$	$\frac{1.070}{\sqrt{n}}$	$\frac{1.220}{\sqrt{n}}$	$\frac{1.360}{\sqrt{n}}$	$\frac{1.520}{\sqrt{n}}$	$\frac{1.630}{\sqrt{n}}$

Example: $n = 20,\ \alpha = 0.10 \rightarrow y_{1-\alpha} = 0.265$

A9.6 Gamma function

Definition: $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx,$

Re$(z) > 0$ (Euler's integral), solution of
$\Gamma(z+1) = z\Gamma(z)$ with $\Gamma(1) = 1$

Special values: $\Gamma(0) = \infty, \quad \Gamma(\tfrac{1}{2}) = \sqrt{\pi}, \quad \Gamma(1) = \Gamma(2) = 1, \quad \Gamma(\infty) = \infty$

Factorial: $n! = \Gamma(n+1) = \sqrt{2\pi}\, n^{n+1/2} e^{-n+\theta/12n}, \quad 0 < \theta < 1$

Relationships:

- Beta function: $B(z, w) = \int_0^1 x^{z-1}(1-x)^{w-1}\, dx = \dfrac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)}$

- Psi function: $\psi(z) = \dfrac{d(\ln \Gamma(z))}{dz}$

- Incomplete Gamma function:

$$\gamma(z, t) = \int_0^t x^{z-1} e^{-x}\, dx, \quad \text{Re}(z) > 0$$

Table A9.6 Gamma function for $1.00 \le t \le 1.99$ (t real), for other values use $\Gamma(z+1) = z\Gamma(z)$

t	0	1	2	3	4	5	6	7	8	9
1.00	1.0000	.9943	.9888	.9835	.9784	.9735	.9687	.9641	.9597	.9554
1.10	.9513	.9474	.9436	.9399	.9364	.9330	.9298	.9267	.9237	.9209
1.20	.9182	.9156	.9131	.9107	.9085	.9064	.9044	.9025	.9007	.8990
1.30	.8975	.8960	.8946	.8934	.8922	.8911	.8902	.8893	.8885	.8878
1.40	.8873	.8868	.8863	.8860	.8858	.8857	.8856	.8856	.8857	.8859
1.50	.8862	.8866	.8870	.8876	.8882	.8889	.8896	.8905	.8914	.8924
1.60	.8935	.8947	.8959	.8972	.8986	.9001	.9017	.9033	.9050	.9068
1.70	.9086	.9106	.9126	.9147	.9168	.9191	.9214	.9238	.9262	.9288
1.80	.9314	.9341	.9368	.9397	.9426	.9456	.9487	.9518	.9551	.9584
1.90	.9618	.9652	.9688	.9724	.9761	.9799	.9837	.9877	.9917	.9958

Examples: $\Gamma(1.25) = 0.9064$; $\Gamma(0.25) = \dfrac{\Gamma(1.25)}{0.25} = 3.6256$; $\Gamma(2.25) = 1.25 \cdot \Gamma(1.25) = 1.133$

A9.7 Laplace Transform

Definition: $\tilde{F}(s) = \int_0^\infty e^{-st}\, F(t)\, dt$

F(t) defined for $t \ge 0$, F(t) piecewise continuos,
$|\,F(t)\,| < Ae^{Bt} \quad (0 < A,\ B < \infty)$

Inverse transform: $F(t) = \dfrac{1}{2\pi j} \int_{C-j\infty}^{C+j\infty} \tilde{F}(s) e^{st}\, ds$

the integral exists in the halfplane defined by
Re$(s) = C > B, \quad j = \sqrt{-1}$

Table A9.7a Properties of the Laplace Transform

	Transform Domain	Time Domain
Linearity	$a_1 \tilde{F}_1(s) + a_2 \tilde{F}_2(s)$	$a_1 F_1(t) + a_2 F_2(t)$
Scale Change	$\tilde{F}(s/a)$	$a F(at), \quad a > 0$
Shift	$\tilde{F}(s-a)$	$e^{at} F(t)$
	$e^{-as} \tilde{F}(s)$	$F(t-a)\,u(t-a), \quad a > 0$
Differentiation	$s^n \tilde{F}(s) - s^{n-1} F(+0) - \ldots - F^{(n-1)}(+0)$	$\dfrac{d^n F(t)}{dt^n} = F^{(n)}(t)$
	$\dfrac{d^n \tilde{F}(s)}{ds^n}$	$(-1)^n t^n F(t)$
Integration	$\dfrac{1}{s} \tilde{F}(s)$	$\int_0^t F(x)\, dx$
	$\int_s^\infty \tilde{F}(z)\, dz$	$\dfrac{F(t)}{t}$
Convolution $(F_1 * F_2)$	$\tilde{F}_1(s) \tilde{F}_2(s)$	$\int_0^t F_1(x) F_2(t-x)\, dx$
Initial and Final * ValueTheorems	$\lim_{s \to \infty} s\tilde{F}(s)$	$\lim_{t \downarrow 0} F(t)$
	$\lim_{s \downarrow 0} s\tilde{F}(s)$	$\lim_{t \to \infty} F(t)$

* existence of the limits is assumed

Table A9.7b Important Laplace Transforms

Transform Domain	Time Domain
$\tilde{F}(s) = \int_0^{\infty} F(t) e^{-st}\, dt$	$F(t)$
1	Impulse $(\delta(0))$
$\frac{1}{s}$	Unit step $(u(t) = 0 \text{ for } t < 0,\ 1/2 \text{ for } t = 0,\ 1 \text{ for } t > 0)$
$\frac{1}{s^n}, \quad n = 1, 2, \ldots$	$\frac{t^{n-1}}{(n-1)!}, \qquad n! = 1 \cdot 2 \cdot \ldots \cdot n, \quad 0! = 1$
$\frac{1}{(s+a)^n}, \quad n = 1, 2 \ldots$	$\frac{t^{n-1}}{(n-1)!} e^{-at}$
$\sum_{i=0}^{n} \frac{a^i}{(s+a)^{i+1}}$	$\sum_{i=0}^{n} \frac{(at)^i e^{-at}}{i!}$
$\frac{1}{(s+a)^{\beta}}, \quad \beta > 0$	$\frac{t^{\beta-1} e^{-at}}{\Gamma(\beta)}, \qquad \beta = n \rightarrow \Gamma(\beta) = (n-1)!$
$\frac{1}{(s+a)(s+b)}, \quad a \neq b$	$\frac{e^{-bt} - e^{-at}}{a-b}$
$\frac{s}{(s+a)(s+b)}, \quad a \neq b$	$\frac{a e^{-at} - b e^{-bt}}{a-b}$
$\frac{1}{(s+a)(s+b)(s+c)}, \quad a \neq b \neq c$	$\frac{(c-b) e^{-at} + (a-c) e^{-bt} + (b-a) e^{-ct}}{(a-b)(b-c)(c-a)}$
$\frac{1}{s^2 + \alpha^2}$	$\frac{1}{\alpha} \sin(\alpha t)$
$\frac{s}{s^2 + \alpha^2}$	$\cos(\alpha t)$
$\frac{1}{(s+\beta)^2 + \alpha^2}$	$\frac{1}{\alpha} e^{-\beta t} \sin(\alpha t)$
$\frac{s+\beta}{(s+\beta)^2 + \alpha^2}$	$e^{-\beta t} \cos(\alpha t)$
$\frac{s}{(s^2 + \alpha^2)^2}$	$\frac{t}{2\alpha} \sin(\alpha t)$
$\frac{P(s)}{Q(s)}, \quad Q(s) = \prod_{k=1}^{n} (s - a_k)$	$\sum_{k=1}^{n} \frac{P(a_k)}{Q'(a_k)} e^{a_k t}, \qquad Q'(a_k) = \left. \frac{dQ(s)}{ds} \right\|_{s=a_k}$

A9.8 Probability Charts

A distribution function appears as a straight line when plotted on a probability chart belonging to its family. The use of probability charts simplifies the analysis and interpretation of data, in particular of life times or failure-free operating times. In the following the charts for *lognormal*, *Weibull*, and *normal* distributions are given (see Section A8.1.3 for the derivation of the Weibull probability chart).

A9.8.1 Lognormal Probability Chart

The distribution function (Eq. (A6.110))

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^t \frac{1}{y} e^{-\frac{(\ln y + \ln \lambda)^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\ln(\lambda t)}{\sigma}} e^{-\frac{x^2}{2}} dx, \qquad t \geq 0, \quad \lambda, \sigma > 0$$

appears as a straight line on the chart of Fig. A9.1 (λ in h^{-1}).

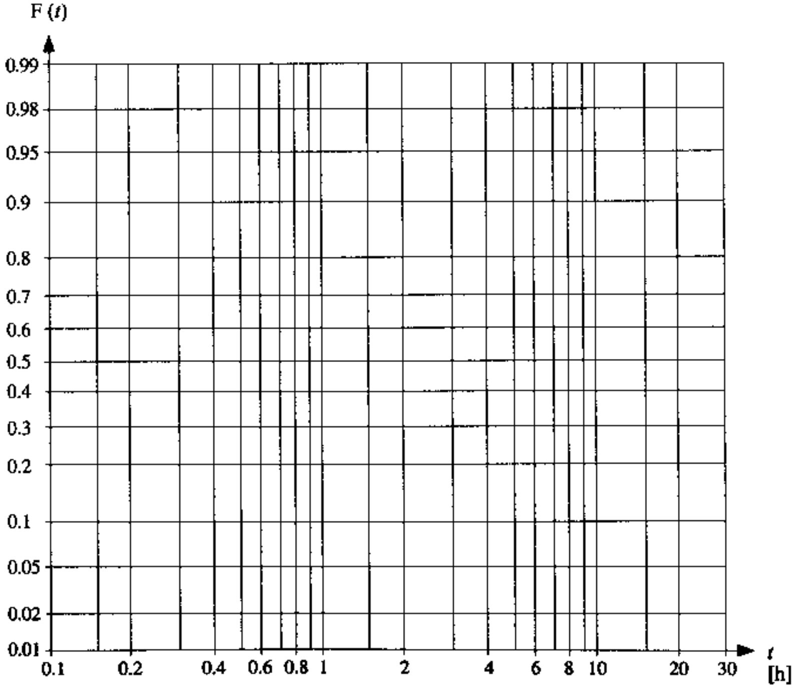


Figure A9.1 Lognormal probability chart

A9.8.2 Weibull Probability Chart

The distribution function $\mathrm{F}(t)=1-e^{-(\lambda t)^{\beta}}$, $t \geq 0$, $\lambda, \beta>0$ (Eq. (A6.89)) appears as a straight line on the chart of Fig. A9.2 (λ in h^{-1}), see Appendix A8.1.3 for a demonstration; on the dashed line $\lambda=1/t$, β appears on the scale $\log_{10}\log_{10}\frac{1}{1-\mathrm{F}(t)}$ when t is varied by one decade (see Figs. A8.2, 7.12, and 7.13).

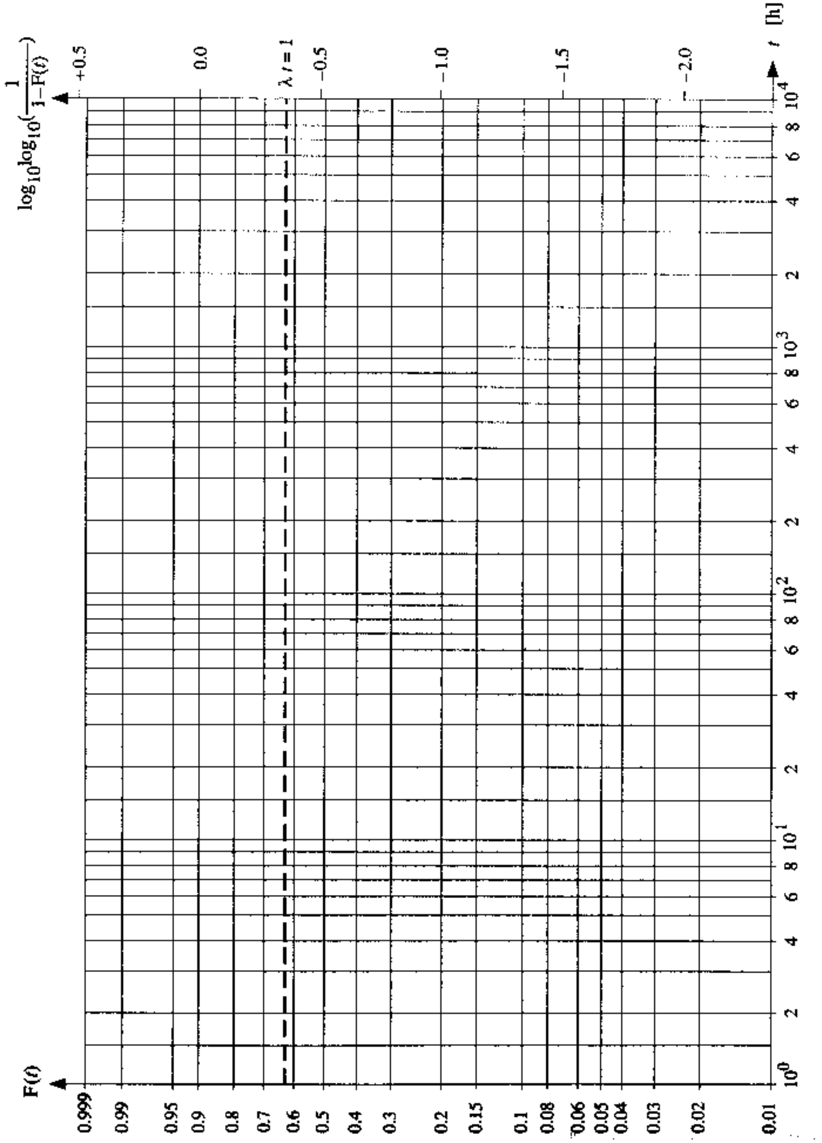


Figure A9.2 Weibull probability chart

A9.8.3 Normal Probability Chart

The distribution function (Eq. (A6.105))

$$\mathrm{F}(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{(y-m)^2}{2\sigma^2}}\,dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{t-m}{\sigma}} e^{-\frac{x^2}{2}}\,dx=\Phi(\tfrac{t-m}{\sigma}), \qquad -\infty<t,\ m<\infty, \quad \sigma>0$$

appears as a straight line on the chart of Fig. A9.3.

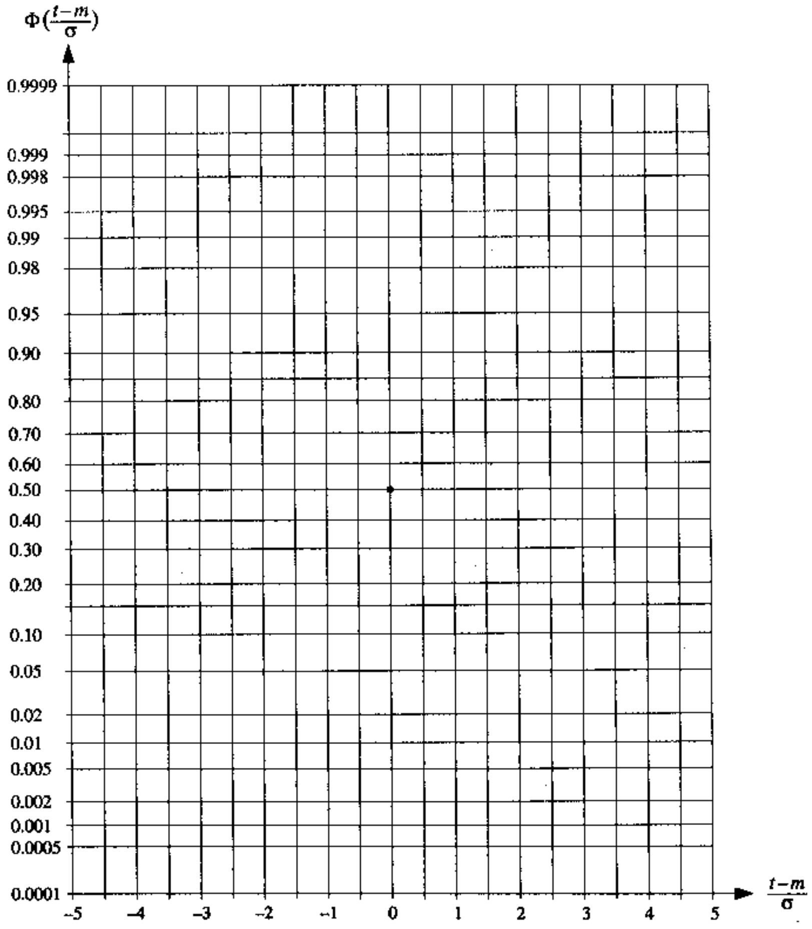


Figure A9.3 Normal probability chart (standard normal distribution)

Acronyms

ACM : Association for Computing Machinery, New York, NY 10036
AFCIQ : Association Française pour le Contrôle Industriel de la Qualité, F-92080 Paris
ANSI : American National Standards Institute, New York, NY 10036
AQAP : Allied Quality Assurance Publications (NATO-Countries)
ASQC : American Society for Quality Control, Milwaukee, WI 53203
BSI : British Standards Institution, London, WIA 2BS
BWB : Bundesamt für Wehrtechnik und Beschaffung, D-56000 Koblenz
CECC : Cenelec Electronic Components Committee, B-1050 Bruxelles
CENELEC : European Committee for Electrotechnical Standardization, B-1050 Bruxelles
CNET : Centre National d'Etudes des Telecommunications, F-22301 Lannion
CSA : Canadian Standards Association, Rexdale, Ontario M9W 1R3, CND
DGQ : Deutsche Gesellschaft für Qualität, D-60549 Frankfurt a. M.
DIN : Deutsches Institut für Normung, D-14129 Berlin 30
DOD : Departement of Defense, Washington, D.C. 20301
EOQC : European Organization for Quality Control, B-1000 Brussel
EOS/ESD : Electrical Overstress/Electrostatic Discharge Association, Rome, NY 13400
ESA : European Space Agency, NL-2200 AG Noordwijk
ESREF : European Symp. on Rel. of Electron. Devices, Failure Physics and Analysis
ETH : Swiss Federal Institute of Technology, CH-8092 Zürich
EXACT : Int. Exchange of Authentic. Electronic Comp. Perf. Test Data, London, NW4 4AP
GIDEP : Government-Industry Data Exchange Program, Corona, CA 91720
GPO : Government Printing Office, Washington, D.C. 20402
GRD : Gruppe Rüstung, CH-3000 Bern 25
IEC (CEI) : International Electrotechnical Commission, CH-1211 Genève 20, P.O.Box131
IECEE : IEC System for Conformity Testing and Certif. of Electrical Equip., CH-1211Genève20
IECQ : IEC Quality Assessment System for Electronic Components, CH-1211 Genève 20
IEEE : Institute of Electrical and Electronics Engineers, Piscataway, NJ 08855-0459 (-1331 for Std.)
IES : Institute of Environmental Sciences, Mount Prospect, IL 60056
IPC : Institute for Interconnecting and Packaging El. Circuits, Lincolnwood, IL 60646
IRPS : International Reliability Physics Symposium (IEEE), USA
ISO : International Organisation for Standardization, CH-1211 Genève 20, P.O.Box56
MIL-STD : Military (USA) Standard, Standardiz. Doc. Order Desk, Philadelphia, PA19111-5094
NASA : National Aeronautics and Space Administration, Washington, D.C. 20546
NTIS : National Technical Information Service, Springfield, VA 22161-2171
RAC : Reliability Analysis Center, Rome, NY 13440-6916
Rel. Lab. : Reliability Laboratory at the ETH (1998 transf. to EMPA S173, CH-8600 Dübendorf)
RL : Rome Laboratory, Griffiss AFB, NY 13441-4505
SAQ : Schweizerische Arbeitsgemeinschaft für Qualitätsförderung, CH-4600 Olten
SEV : Schweizerischer Elektrotechnischer Verein, CH-8320 Fehraltorf
SNV : Schweizerische Normen-Vereinigung, CH-8008 Zürich
SOLE : Society of Logistic Engineers, Huntsville, AL 35806
VDI/VDE : Verein Deutscher Ing./Verband Deut. Elektrotechniker, D-60549 Frankfurt a. M.

References

(see Acronyms on p. 476)

1 Basic Concepts, Historical Development, Quality and Reliability Assurance

[1.1] Birnbaum Z.W., Esary J.D., Saunders S.C., "Multi-component systems and structures and their reliability", *Technometrics*, 3(1961), pp. 55-77.

[1.2] Birolini A., "Product assurance tasks and organisation", *Proc. 21st EOQC Conf.*, Varna 1977, Vol.1, pp. 316-329; "Qualitäts- und Zuverlässigkeitssicherung komplexer Systeme: Teil 1 und 2", *Bull. SEV/VSE*, 70 (1979), pp. 142-148 and 237-243; *Quality and Reliability of Technical Systems*, 2nd Ed. 1997, Springer, Berlin; *Zuverlässigkeit von Geräten und Systemen, 4th Ed.* 1997, Springer, Berlin; "Reliability engineering: Cooperation between University and Industry at the ETH Zurich", *Quality Eng.*, 8(1996)4, pp. 659-674.

[1.3] Buckley F.J., *Conf. Manag.: Hardware, Software, Firmware*, 1993, IEEE Press, Piscataway NJ.

[1.4] Coppola A., "Reliability engineering of electronic equipment - a historical perspective", *IEEE Trans. Rel.*, 33(1984), pp. 29-35.

[1.5] Feigenbaum A.V., *Total Quality Control*, 3rd Ed. 1983, McGraw-Hill, New York.

[1.6] Frehr H., Hormann D. (Ed.), *Produktentwicklung und Qualitätsmanagement*, 1993,VDE, Berlin.

[1.7] Haug S. et al, "Impact of electronic components obsolescence on commercial aerospace", *Aerospace Mag.*, 1999, March , pp. 26-31.

[1.8] IEEE, Special issue on: 50th Anniversary IEEE Rel. Soc., *IEEE Trans. Rel.*, 47(1998)3-SP.

[1.9] Juran J.M., Gryna F.M. (Eds)., *Quality Control Handbook*, 4th Ed. 1988, McGraw-Hill, New York.

[1.10] Keys L.K., Rao R., Balakrishnan K., "Concurrent engineering for consumer, industrial products, and Government systems", *IEEE Trans. Comp., Manuf. Technol.*, 15(1992)3, pp.282-287.

[1.11] Kretschmer F. et al, *Produkthaftung in der Unternehmungspraxis*, 1992, Kohlhammer, Stuttgart.

[1.12] Kuehn R., "Four decades of reliability experience", *Proc. Ann. Rel. & Maint. Symp.*, 1991, pp. 76-8

[1.13] Kull T.M., "Vorbeugen ist besser als Haften", *TR Transfer*, (1994)5, pp. 4-7; "Technische Richtlinien in der EU", *TR Transfer*, (1994) 16, pp. 44-46.

[1.14] Kusiak A. (Ed.), *Concurrent Engineering: Automation, Tools and Techniques*, 1993, Wiley, N.Y.

[1.15] Masing W. (Ed), *Handbuch der Qualitätssicherung*, 2nd Ed. 1988, Hanser, Munich.

[1.16] Mattana G., *Qualità, Affidabilità, Certificazione*, 2nd Ed. 1988, Angeli, Milano.

[1.17] MIL-HDBK-338: *Electronic Reliability Design Handbook*, Vol. I Ed. A 1988, Vol. II 1984.

[1.18] Moore E.F., Shauman C.E., "Reliable circuits using less reliable relays", *J. of the Franklin Inst.*, 262(1956), pp. 191-208 and 281-297.

[1.19] Orgalime Legal Affairs WG: *Product Liability in Europe*, 2nd Ed. 1993, Orgalime, Bruxelles.

[1.20] Peters G.A., "Risk analysis", *Technology, Law and Insurance,* (1997) 2, pp. 97-110.

[1.21] RAC/RL, *Reliability Toolkit : Commercial Practices Edition*, 1995, RAC, Rome NY.

[1.22] Seghezzi H.D. (Ed.), *Top Management and Quality*, 1992, Hanser, Munich.

[1.23] Taguchi G., *System of Experimental Design-Engineering Methods to Optimize Quality and Minimize Costs*, Vol. 1 & 2., 1987, Unipub, White Plains NY.

[1.24] Umiker B. et al., "Wie lassen sich grosse Industriekatastrophen verhüten?", *Manag. Zeitschrift*, 1(1987), pp. 15-22; "The modern art of a discourse on risk", *4th Europ. Conf. on Safety Anal. & Risk Manag.*, Oct. 19, 1993, Roma; "Risk management: Concept and implementation", *ASCOM Tech. Mag.*, 3(1994), pp. 33-36; Kommunikationssicherheit in der Sicherheitskommunikation", *Amer. Soc. for Ind. Security Meeting*, Zürich ETH, June 4, 1997.

[1.25] Von Neumann J., "Probabilistic logics and the synthesis of reliable organisms from unreliable components", *Ann. of Math. Studies*, 34(1956), pp. 43-98.

see also [A1.1 to A5.6]

2 Reliability Analysis

Reliability Techniques

[2.1] Arlat J. et al, "Fault injection and dependability evaluation of fault tolerant systems", *IEEE Trans. Comp.*, 42(1993)8, pp. 913-923.

[2.2] Birolini A., *Zuverlässigkeit von Schaltungen und Systemen*, 4th Ed. 1982, Course at the ETH Zurich; *Modelle zur Berechnung der Rentabilität der Q.- und Z.-sicherung komplexer Waffensysteme*, 1986, GRD, Bern; "Zuverlässigkeitssicherung von Automatisierungssystemen", *ebi*, 107(1990), pp. 258-271; *Quality and Reliability of Technical Systems*, 2nd Ed. 1997, Springer, Berlin; *Zuverlässigkeit von Geräten und Systemen, 4th Ed.* 1997, Springer, Berlin;

[2.3] Catuneanu V.M., Mihalache A.N., *Reliability Fundamentals*, 1989, Elsevier, Amsterdam.

[2.4] Dhillon B.S., *Human Reliability*, 1986, Pergamon, New York.

[2.5] Friedman M.A., Tran P., "Reliability techniques for combined hardware/software systems", *Proc. Ann. Rel. & Maint, Symp.*, 1992, pp. 290-293.

[2.6] Henley E.J., Kummamoto H., *Probabilistic Risk Assessment*, 1992, IEEE Press, Piscataway NJ.

[2.7] IEC 60300-3-1: *Dep. Manag.- Analysis Tech. for Dependability*, 1991; -9: - *Risk Anal. of Technol. Systems*, 1995; see also 60812 (FMEA), 61025 (FTA), 61078 (Rel. Block Diagram Method), 61165 (Markov Tech.), 60863 (Pres. of Rel., Maint., and Availability Predictions), 61709 (Failure Rates).

[2.8] Jensen F., *Electronic Component Reliability*, 1995, Wiley, New York.

[2.9] Klaassen K., "Active red. in analogue el. systems", *Proc.Ann.Rel.&M.Symp.*, 1975, pp.573-578.

[2.10] Messerschmitt-Bölkow-Blohm (Ed.), *Technische Zuverlässigkeit*, 3rd Ed. 1986, Springer.

[2.11] MIL-HDBK-338: *Electronic Reliability Design Handbook*, Vol. I Ed. A 1988, Vol. II 1984.

[2.12] NASA CR -1126-1129: *Practical Reliability* (Vol. 1 to 4), 1968.

[2.13] O' Connor P.D.T., *Practical Reliability Engineering*, 3th Ed. 1991, Wiley, New York.

[2.14] Pecht M.G., Palmer M., Naft J., "Thermal reliability management in PCB design", *Proc. Ann. Rel. & Maint. Symp.*, 1987, pp. 312-315.

[2.15] RAC, WCCA: *Worst Case Circuit Analysis Appl. Guidelines*, 1993; RTMG: *Thermal Management Guidebook*, 1995; RADC-TR-90-109: *Integration of Sneak Analysis with Design, 1990.*

[2.16] Rooney J.P., "Storage reliability", *Proc. Ann. Rel. & Maint. Symp.*, 1989, pp.178-182.

[2.17] Siewiorek D.P., "Architecture of fault-tolerant computers, an historical perspective", *Proc. IEEE*, 79(1991)12, pp. 1710-1734; - and Swarz R.S., *Reliable Computer Systems Design and Evaluation*, 1992, Digital Press, Bedford MA.

[2.18] Suich R.C., Patterson R.L, "Minimize system cost by choosing optimal subsystem reliability and redundancy", *Proc. Ann. Rel. & Maint. Symp.*, 1993, pp. 293-297.

[2.19] Tomek L. et al, "Rel. model. of fife-critical real-time systems", *Proc IEEE*, 79(1994)1, pp.108-21.

[2.20] Villemeur A., *Sûreté de Fonctionnement des Systèmes Industriels*, 2nd Ed. 1993, Eyrolles, Paris.

see also [5.1 to 5.58 and 6.1 to 6.19]

Failure Rates/Models for Electronic Components

[2.21] AFCIQ: *Données de Fiabilité en Stockage des Composants Electroniques*, 1983.

[2.22] Bellcore, TR-332: *Reliability Prediction Procedure for Electronic Equipment*, 6th Ed. 1997, Bellcore, Livingston NJ.

[2.23] CNET *RDF 93*: *Recueil de Données de Fiabilité des Composants Electroniques*, 1993, CNET, Lannion (also as *British Telecom Rel. HDBK* and *Italtel Rel. Pred. HDBK*); *RDF 99* announced.

[2.24] IEC 61709: *Electronic Components Reliability - Reference Condition for Failure Rates and Stress Models for Conversion*, 1996.

[2.25] IEEE Std1413-1998: *IEEE Standard Methodology for Rel. Pred. and Assessm. for el. Syst.& Equip.*

[2.26] MIL-HDBK-217: *Reliability Prediction of Electronic Equipment*, Ed. F 1991, Not. 2 1995.

[2.27] NTT: *Standard Rel. Tables for Semicond. Dev.*, 1985, Nippon Telegraph and Tel., Tokyo.

[2.28] RAC, NONOP-1: *Nonoperating Rel. Data*, 1992; NPRD-95: *Nonelectronic Parts Rel. Data*, 1995; see also TR-89-177: *VHSIC/ VHSIC Rel. Modeling*, 1989 and TR-90-72: *Rel. Analysis Assessment of Advanced Technologies*, 1990; *Failure Rate Models*, announced with PRISM [2.57].

[2.29] Siemens, SN 29 500 Teil 1: *Ausfallraten Bauelemente*, 1991, Siemens, Munich (also in DIN 40039).

Reliability of Large/Complex Structures

[2.31] Arunkumar S., Lee S.H., "Enumeration of all minimal cut-sets for a node pair in a graph", *IEEE Trans. Rel.*, 28(1987)1, pp. 51-55.

[2.32] Bansal V.K., "Minimal pathsets and minimal cutsets using search techniques", *Microel. & Rel.*, 22(1982)6, pp. 1067-1075.

[2.33] Barlow R.E., Proschan F., *Mathematical Theory of Reliability*, 1965, Wiley, New York; *Statistical Theory of Reliability and Life Testing*, 1975, Holt Rinehart, New York.

[2.34] Bollinger R.C., Salvia A.A., "Consecutive-k-out-of-n: F networks", *IEEE Trans. Rel.*, 31(1982)1, pp. 53-56; "Strict consecutive-k-out-of-n: F systems", *IEEE Trans. Rel.*, 34(1985)1, pp. 50-52.

[2.35] Heneley E.J., Kumamoto H., *Probabilistic Risk Assessment*, 1992, IEEE Press, Piscataway NJ.

[2.36] Hura G.S., "A Petri net approach to enumerate all system success paths for rel. evaluation of complex systems"; "Petri net approach to the analysis of a structured program"; "Petri net as a mod. tool", *Microel. & Rel.*, 22 (1982)3, pp. 427-439, 23(1983), pp. 157-159, 467-475, 851-853.

[2.37] *IEEE Trans. Rel.*, Special issues on: Network Rel., 35(1986)3, Reliability of Parallel & Distributed Computing Networks, 38(1989)1, Experimental Evaluation of Computer Systems Rel., 39(1990)4, Design for Reliability of Telecomm. Systems and Services, 40(1991)4.

[2.38] Kossow A., Preuss W., "Failure probability of strict consecutive-k-out-of-n: F systems", *IEEE Trans. Rel.*, 36(1987)5, pp. 551-553; "Rel. of consecutive-k-out-of-n: F systems with nonidentical component rel.", *IEEE Trans. Rel.*, 38(1989), pp. 229-233; "Mean time to failure for linear-consec.-k-out-of-n: F systems", *IEEE Trans. Rel.*, 40(1991)3, pp. 271-272; "Rel. of linear consecutive connected systems with multistate comp.", *IEEE Trans. Rel.*, 44(1995)3, pp. 518-522.

[2.39] Luo T., Trivedi K.S., "An improved algorithm for coherent-system reliability", *IEEE Trans. Rel.*, 47(1998)1, pp. 73-78.

[2.40] Rai S. and Agrawal D.P. (Ed.), *Advances in Distributed Systems Reliability*, 1990 IEEE Press, Piscataway NJ; *Distributed Computing Network Reliability*, 1990, IEEE Press, Piscataway NJ.

[2.41] Râde L., "Reliability survival equivalence", *Microel. & Rel.*, 33(1993)6, pp. 881-894.

[2.42] Sanso B., Soumis F., "Communication and transportation network reliability using routing models", *IEEE Trans. Rel.*, 40(1991)1, pp. 29-38.

[2.43] Schneeweiss W., *Zuverlässigkeits-Systemtheorie*, 1980, Datakontex, Köln; *Boolean Functions with Engineering Applications and Computer Programs*, 1989, Springer, Berlin; "Usefulness of MTTF of *s*-independent case in other cases", *IEEE Trans. Rel.*, 41(1992)2, pp. 196-200; "Calculating mean system-failure freq. with prescribed accuracy", *IEEE Trans. Rel.*, 46(1997)2, pp. 201-207.

[2.44] Störmer H., *Mathematische Theorie der Zuverlässigkeit*, 2nd Ed., 1983, Oldenbourg, Munich.

see also [6.1 to 6.19]

Software Tools

[2.51] Bernet R., "CARP - A program to calculate the predicted reliability", *6th Int. Conf. on Rel. & Maint.*, Strasbourg 1988, pp. 306-310; *Modellierung reparierbarer Systeme durch Markoff- und Semiregenerative Prozesse*, 1992, Ph.D.Thesis 9682, ETH Zurich; Birolini A. et al, *CARAP ETH Technical Spec.*, 1995, Report S10, ETH Zürich, Rel. Lab.; Kovalenko I. and Kuznetov N., *Basis of the RS-Program/Guidance to the RS-Program*, 1997, Rep. S13/S14, ETH Zurich, Rel. Lab.

[2.52] Bowles J.B. and Klein L.A., "Comparison of commercial reliability-prediction programs", *Proc. Ann. Rel. & Maint. Symp.*, 1990, pp. 450-455.

[2.53] Gymayr J. and Ebeken N., "Fault-tree analysis: a knowledge-engineering approach", *IEEE Trans. Rel.*, 44(1995)1, pp. 37-45.

[2.54] Jaeger H., "Computerprogramm RAMTOOL", *Proc. ETH/IEEE Int. Symp. on Rel. Eng. 2'000*, ETH Zurich, Rel. Lab., Oct. 17, 1996 (20 pp); *Zuverlässigheit und Materialerhaltbarkeit*, Course at the Bundesakad. für Wehrverwaltung und Wehrtechnik, Mannheim, 1998.

[2.55] Johnson A.M. and Malek M., "Survey of software tools for evaluating reliability, availability, and serviceability", *ACM Comp. Surveys*, 20(1988)4, pp. 227-269.

[2.56] Locks M.O., "Recent develop. in computing system rel.", *IEEE Trans. Rel.*, 34(1985)5, pp. 425-36.

[2.57] RAC, *RMST: Rel. & Maint. Software Tools, 1993; PRISM: Software Package for Assessing Components and Systems Reliability*, announced.

[2.58] Sahner R., Trivedi K.,"Rel. modeling using SHARPE", *IEEE Trans. Rel. 36(1987)*, pp. 186-193.

Mechanical Reliability

[2.61] AFCIQ, *Guide d' Evaluation de la Fiabilité en Mécanique*, 1981.
[2.62] Barer R.D., *Why Metals Fail*, 3rd Ed. 1974, Gordon & Breach, New York.
[2.63] Bertsche B., Lechner G., *Zuverlässigkeit im Maschinenbau*, 1990, Springer, Berlin.
[2.64] Bogdanoff J.L., Kozin F., *Probabilistic Models for Cumulative Damage*, 1985, Wiley, New York.
[2.65] Carter A.D.S., *Mechanical Reliability*, 2nd Ed. 1986, Macmillan, London.
[2.66] Collins J.A., *Failure of Materials in Mechanical Design*, 1981, Wiley, New York.
[2.67] Engelmaier W., *Reliable Surface Mount Solder Attachements Through Design and Manuf. Quality*, 1993, Report L21, ETH Zurich, Rel. Lab.(also in *Proc.ETH/IEEE Workshop on SMT*, 1992).
[2.68] Freddi S., *Design of Experiment*, Course at the 15th Symp. Danubia-Adria, Bertinoro, 1998.
[2.69] Hutchings F., Unterweiser P. (Ed.), *Failure Analysis*, 1981, Am. Soc. Met., Metals Park OH.
[2.70] Kececioglu D., *Reliability Eng. Handbook* (Vol. 1 & 2), 1991, Prentice, Englewood Cliffs NJ.
[2.71] Kutz M. (Ed.), *Mechanical Engineers' Handbook*, 1986, Wiley, New York.
[2.72] Manson S.S., *Thermal Stress and Low-Cycle Fatigue*, 1981, Krieger, Malabar FL.
[2.73] Nelson J. et al, "Rel. models for mech. equip.", *Proc. Ann. Rel.&Maint. Symp.*, 1989, pp.146-153.
[2.74] RAC, *Mechanical Applications in Reliability Engineering*, 1993.

Failure (Fault) Mode and Effects Analysis (FMEA), Fault Tree Analysis (FTA)

[2.81] Bednarz S. et al, "Efficient analysis for FMEA", *Proc. Ann. Rel. & Maint. Symp.*, 1988, pp. 416-21.
[2.82] DIN 25419: *Störfallablaufanalyse*, 1977-79; 25424: *Fehlerbaumanalyse*, 1981; 25448: *Ausfalleffektanalyse*, 1980; 31000: *Allg. Leit. für das sicherheitsgerechte Gestalten tech. Erzeug.*, 1979.
[2.83] Feo T., "PAFT F77: Program for the anal. of fault trees", *IEEE Trans. Rel.*, 35(1986)1, pp.48-50.
[2.84] IEC 60812: *Procedure for FMEA*, 1985; 61025: *Fault Tree Analysis (FTA)*, 1990.
[2.85] Hall F.M., Paul R.A., and Snow W.E., "Hardware/Software FMECA", *Proc. Ann. Rel. & Maint. Symp.*, 1983, pp. 320-327.
[2.86] Jackson T., "Integration of sneak circuit analysis with FMEA", *Proc. Ann. Rel. & Maint. Symp.*, 1986, pp. 408-414.
[2.87] MIL-STD-1629: *Procedures for Performing a FMECA*, Ed. A 1980.
[2.88] Price C.J. et al, "Identifying design glitches through automated design analysis", *Proc. Ann. Rel.&Maint. Symp.*,1999, pp.277-282.
[2.89] RAC, *FMECA*, 1993; *FTA*, 1990; *WCCA* (Worst Case Circuit Analysis), 1992.
[2.90] Stamenkovic B. and Holovac S., "Failure modes, effects and criticality analysis: The basic concepts and applications", *Proc. Int. Summer Seminar*, Dubrovnik, 1987, pp. 21-25.

see also [4.4, 5.42, 5.43, 5.47, 5.54, 5.57]

3 Qualification of Components and Assemblies

Selection Criteria and Qualification Tests for Components

[3.1] Barber M.R., "Fundamental timing problems in testing MOS VLSI on modern ATE", *IEEE Design & Test*, (1984)8, pp. 90-97.
[3.2] Birolini A., "Mögl. und Grenzen der Qualifikation, Prüfung und Vorbehandlung von ICs", *QZ*, 27(1982)11, pp. 321-326; "Prüfung und Vorbehandlung von Bauelem. und bestück. Leiterplatten", *VDI/VDE Fachtagung*, Karlsruhe 1984, VDI Bericht Nr. 519, pp. 49-61; "Neue Ergebnisse aus der Qualif. grosser Halbleiterspeicher", *me*, 7(1993) 2, pp. 98-102; – , Büchel W., Heavner D., "Test and screening strategies for large memories", *1st European Test Conf.*, Paris: 1989, pp. 276-283.
[3.3] Brambilla P., Canali C., Fantini F., Magistrali F., Mattana G., "Rel. evaluation of plastic-packaged device for long life applications by THB test", *Microel. & Rel.*, 26(1986)2, pp. 365-384.
[3.4] Crook D.L., "Evolution in VLSI reliability engineering", *Int. Rel. Phys. Symp.*, 1990, pp. 2-11.
[3.5] CECC 00107: *Quality Assessment Procedures* - part I to III, 1979-1982 (see also 00100 to 00106, 00108, 00109, and 00111 to 00113); 00200: *Qualified Product List.*

[3.6] ESA PSS 01-603: *ESA Preferred parts List*, 3rd Ed. 1995.
[3.7] ETH Zurich Reliability Lab., *Reports Q2-Q12: Qualification Test for DRAMs* 256Kx1, SRAMS 32Kx8, EPROMs 32Kx8, SRAMs 8Kx8, DRAMs 1Mx1, EEPROMs 8Kx8, SRAMs 128Kx8, DRAMs 4Mx1, EEPROMs 32Kx8, EPROMs 64Kx16, and FLASH-EPROMs 128Kx8. 1989-92.
[3.8] Gerling W., "Modern reliability assurance of integr. circuits". *Proc. ESREF'90*, Bari, pp. 1-12.
[3.9] IEC 60068: *Environmental Testing* (Parts 1-5), 1969-97;
60721: *Classification of Environmental Conditions* (Parts 1-3), 1982-97;
60749: *Semicon. Devices: Mech. and Climatic Test Methods*, 1996.
61000: *Electromagnetic Compatibility (EMC)*, (Parts 1-6), 1990-98.
[3.10] IEEE, Special issues on: Reliability of Semiconductor Devices, *Proc. IEEE*, 62(1974)2; Micron and Submicron Circuit Engineering, *Proc. IEEE*, 71(1983)5; Integrated circuit technologies of the future, *Proc. IEEE*, 74(1986)12; VLSI Reliability, *Proc. IEEE*, 81(1993)5.
[3.11] MIL-STD-883: *Test Methods and Procedures for Microelectronics*, Ed. D 1991; see also -199, -202, -750, -810, -976, -I3 8535, -M 38510, -S 19500.
[3.12] Ousten Y., Danto Y., Xiong N., Birolini A., "Rel. eval. and failure diagnostic of Ta-cap. using freq. anal. and design of exp", *Proc.Int. Symp. for Testing & Failure Analysis* , 1992, pp. 189-196.
[3.13] Powell R.F., *Testing Active and Passive Electronic Components*, 1987, Dekker, New York.
[3.14] RAC, *PSAC: Parts Selection, Appl. and Control, 1993; CAP: Reliable Appl. of Components, 1993; PEM2: Reliable Appl. of Microcircuits, 1996; HYB: Reliable Appl. of Hybryds, 1993; MCM: Reliable Appl. of Multichip Modules, 1995.*
[3.15] Ratchev D., "Are NV-Mem. non-volatile?" *Proc. 1993 IEEE Workshop on Memory Test.*, pp.102-06.
[3.16] Sawada K. and Kagano S., "An evaluation of I_{DDQ} versus conventional testing for CMOS sea-of-gate ICs", *Int. Test Conf.*, 1992, pp. 158-167.
[3.17] Thomas R.W., "The US Department of Defense procurement strategy and the semiconductor industry in the 1990's", *Proc. 4th Int. Conf. Quality in El. Comp.*, Bordeaux 1989, pp. 1-3.
[3.18] van de Goor A.J, *Testing Semiconductor Memories*, 1991, Wiley, New York.
[3.19] Williams T.W. (Ed.), *VLSI - Testing*, 1986, North-Holland, Amsterdam.
[3.20] Wolfgang E., Görlich S., Kölzer J., "Electron beam testing", *Proc. ESREF'90*, Bari, pp.111-120.
[3.21] Zerbst M. (Ed), *Meß- und Prüftechnik*, 1986, Springer, Berlin.
[3.22] Zinke O., Seither H., *Widerstände, Kondensatoren Spulen u. ihre Werkstoffe*, 1982, Springer, Berlin.

Failure Mechansims, Failure Analysis

[3.31] Amerasekera E., Campbell D., *Failure Mechanisms in Semiconductor Devices*. 1987, Wiley, NY.
[3.32] Barbottin G., Vapaille A., (Eds.), *Instabilities in Silicon Dev*, 1986, North-Holland, Amsterdam.
[3.33] Brox M., Weber W., "Dynamic degradation in MOSFETs – Part 1: The physical effects", *IEEE Trans. El. Dev.*, 38(1991)8, pp. 1852-1858.
[3.34] Chen P., et al, "A unified compact scalable ΔI_d model for hot carrier reliability simulation", *Proc. Int. Rel. Phys. Symp.*, 1999, pp. 243-248.
[3.35] Ciappa M., *Ausfallmech. integrierter Schaltungen*, 1991, Reports F1 and F4, ETH Zurich, Rel. Lab.
[3.36] De Salvo B., et al, "A new physical model for NVM data-retention time to failure", *Proc. Int. Rel. Phys. Symp.*, 1999, pp. 19-23.
[3.37] Degraeve R., et al, "On the field depend. of intrinsic and extrinsic time-dep. dielectric breakdown", *Proc. Int. Rel. Phys. Symp.*, 1996, pp. 44-54.
[3.38] Fantini F., "Reliability and failure physics of integrated circuits", in *Dinemite II*, (Vol. IV), Leuven, B: Interuniversitair Micro-Elektronica Centrum, 1986, pp. 1-30.
[3.39] Fiegna C., Venturi F., Melanotte M., Sangiorgi E., Riccò B., "Simple and efficient modeling of EPROM writing", *IEEE Trans. El. Dev.*, 38(1991)3, pp. 603-610.
[3.40] Fung R.C.-Y., Moll J.L., "Latch-up model for the parasitic p-n-p-n path in bulk CMOS", *IEEE Trans. El. Devices*, 31(1984)1, pp. 113-120.
[3.41] Ghidini G. et al, "Charge trapping mechanism under dynamic stress and their effect on failure time", *Proc. Int. Rel. Phys. Symp.*, 1999, pp. 88-92.
[3.42] Glasstone S., Laidler K.J., Eyring H.E., *The Theory of Rate Processes*, 1941, McGraw-Hill, N.Y.
[3.43] Herrmann M., *Charge Loss Modeling of EPROMs with ONO Interpoly Dielectric*, 1994, Ph.D. Thesis 10817, ETH Zurich; - and Schenk A., "Field and high- temperature dependance of the longterm charge loss in EPROMs", *J. Appl. Phys.*, 77 (1995)9, pp. 4522-4540.

[3.44] Howes M.J., Morgan D.V. (Eds.), *Reliability and Degradation - Semiconductor Devices and Circuits*, 1981, Wiley, New York.
[3.45] Hu C. (Ed.), *Nonvolatile Semiconductor Memories: Technologies, Design, and Applications*. 1991, IEEE Press, Piscataway NJ.
[3.46] Hu C., Lu Q., "A unified gate oxide reliability model", *Proc. Int. Rel. Phys. Symp.*,1999, pp. 47-51.
[3.47] Kolesar S.C., "Principles of corrosion", *Proc. Int. Rel. Phys. Symp.*, 1974, pp. 155-167.
[3.48] Lall P.,"Temperature as input to microelectronics-rel. models", *IEEE Trans. Rel.*, 45(1996)1, pp.3-9.
[3.49] Lantz L., "Soft errors induced by α - particles", *IEEE Trans. Rel.*, 45 (1996)2, pp. 174-179.
[3.50] Lee J.H., et al, "Using erase self-detrapped effect to eliminate the flash cell program/erase cycling V_{th} window close", *Proc. Int. Rel. Phys. Symp.*, 1999, pp. 24-29.
[3.51] Li E. et al, "Hot carrier effects in nMOSFETs in 0.1µmm CMOS technology", *Proc. Int. Rel. Phys. Symp.*, 1999, pp. 253-258.
[3.52] Mann J.E., "Failure analysis of passive devices", *Proc. Int. Rel. Phys. Symp.*, 1978, pp. 89-92.
[3.53] Menzel E., Kubalek E., "Fundamentals of electron beam testing of integrated circuits", *Scanning*, (1983)5, pp. 136-142.
[3.54] Pecht M.G., Radjojcic R., *Guidebook for Managing Silicon Chip Reliability*, 1999, CRC Press, Boca Raton FL.
[3.55] Peck D.S., "Comprehensive model for humidity testing correlation", *Proc. Int. Rel. Phys. Symp.*, 1986, pp. 44-50; - and Thorpe W.R., "Highly accelerated stress (THB) - Test history, some problems and solutions", *Tutorial Notes at the Int. Rel. Phys. Symp.*, 1990, pp. 4.1-4.27.
[3.56] Quader K.N., Fang P., Yue J., Ko P.K., Hu C., "Simulation of CMOS circuit degradation due to hot-carrier effects", *Proc. Int. Rel. Phys. Symp.*, 1992, pp. 16-23.
[3.57] RAC, FMD: *Failure Mode/Mechanism Distribution*, 1991; MFAT-1: *Microelectronic Failure Analysis Techn. - a Procedual Guide*, 1981; MFAT-2: *GaAs Microcircuit Charact. & Failure Analysis Techn*, 1988.
[3.58] Reiner J., "Latent Gate Oxide Defects Caused by CDM-ESD", *Proc. EOS/ESD Symp*, Phoenix AZ, 1995, pp. 6.5.1-11, also in *Jour. of Electrostatic*, 38(1996) pp. 131-157; *Latent Gate Oxide Damage Induced by Ultrafast Electrostatic Discharge*, 1995, Ph.D. Thesis 11212, ETH Zurich.
[3.59] Schuegraf K.F., Hu C., "Reliability of thin SiO_2", *Semicond. Sci. Technol.*, 9(1994), pp. 989-1004.
[3.60] Srinivasan G.R., "Modeling the cosmic-ray-induced soft errors in IC's: An overview", *IBM J. of R&D*, 40 (1996)1,pp. 77-90.
[3.61] Troutmann R.R., "Latch-up in CMOS technol.", *IEEE Circuits and Dev. Mag.*, (1987)5, pp.15-21.
[3.62] Wang C.T. (Ed.), *Hot Carrier Design Considerations for MOS Devices and Circuits*, 1992, Van Nostrand Reinhold, New York.
[3.63] Weber W. et al, "Dynamic degradation in MOSFETs-Part2: Application in the circuit environment", *IEEE Trans. El. Dev.*, 38(1991)8, pp. 1859-1867.
[3.64] Wong K.L., "The common thread for operational reliability and failure physics", *Microel. & Rel.*, 22(1982)2, pp. 177-186; - , Kallis J.M., Burkhard A.H., "Culprits causing avionic equipment failures", *Proc. Ann. Rel. & Maint. Symp.*, 1987, pp. 416-421.
[3.65] Wu W., Held M., Umbricht N., Birolini A., "dv/dt induced latching failures in 1200V/400A halfbridge IGBT modules", *Int. Rel. Phys. Symp.*, 1994, pp.

Microconnections and Packaging

[3.71] ASM, *Packaging*, Vol.1, 1989, ASM Int., Material park OH.
[3.72] Barker D.B., Dasgupta A., Pecht M., "Printed-wiring-board solder-joint fatigue-life calculation under thermal and vibration loading", *Proc. Ann. Rel. & Maint. Symp.*, 1991, pp. 451-459.
[3.73] Ciappa M., Malberti P., "Reliability of laser-diode modules in temperature-uncontrolled env.", *Int. Rel. Phy. Symp.*, 1994, pp. 466-469; - and Scacco P., "Selective back-etch for silicon devices.", *Proc. ISTFA'95*, Santa Clara CA, 1995, pp. 257-261.
[3.74] Darveause R. and Banerji K., "Constitutive relations for tin-based solder joints.", *IEEE Trans. Compon., Pack., and Manuf. Technol.*, 15 (1992) 6, pp.1013-1024.
[3.75] Engelmaier, W., "Enviromental stress screening and use enviroments - their impact on solder joint and plated-through-hole reliability", *Proc. Int. Electronics Packaging Conf.*, Marlborough MA, 1990, pp. 388 - 393; - and Attarwala A.I., "Surface-mount attachment boards", *IEEE Trans., Compon., Pack., and Manuf. technol.*, 12(1989)2, pp. 284-289.
[3.76] ETH Zurich, Rel. Lab., *Reports P3-P16 and P18: Qualification Tests on 7 Telecom. Equip.*,1989-91.

[3.77] Fenech A., Hijagi A., Danto Y., "Determination of thermomechanical behaviour of microel. packaging based on mictostructural analysis", Proc. ESREF" 94, Glasgow, 1994, pp. 405 - 410.
[3.78] Frear D.R.(Ed.), *The Mechanics of Solder Alloy Interconnections*, 1994,Van Nostrand Reinhold, N.Y.
[3.79] Grossmann G., *Zuverlässigkeit von Weichlotstellen*, 1993, Report L29, ETH Zurich, Rel. Lab.; *Produktion und Prüfung von Testprints der SMT Fine Pitch*, 1996, Rep.K12, ETH Zurich, Rel.Lab.; "Metallurgical consid. for acc. testing of el. equip.", *IEEE Trans. Comp., Pack. & Manuf. Technol.*, 20(1997)2, pp. 213-218; "The deformation of Sn62Pb36Ag2 and its impl. on the design of acc. tests for solder joints", *IEEE Trans. Comp., Pack.& Manuf. Technol.*, 22(1999)1, pp. 71-79; - *et al.* "Properties of thin layers of Sn62Pb36Ag2", *Proc. 1995 IEMT Symp.*, pp. 502-507; "Metallurgical consid. for accel. testing of el. equip.", *Proc. 1996 IEMT Symp.*, pp. 298-304; "Lifetime ass. of soft sold. joints on the base of the behav. of Sn62Pb36Ag2", *Proc. 1997 IEMT Symp.*, pp.256-263.
[3.80] Heiduschke K., "The logarithmic strain space description", *Int. J. Solids Structures*, 32 (1995), pp. 1047-1062 and 33(1996) pp. 747-760; *Kontinuumsmech. und Finite Element Modellierung* (Software package URMEL), 1996, Report K11, ETH Zurich, Rel. Lab.; - and Grossmann G., "Modeling fatigue cracks with spatial shape", *Proc. EuPac '94*, Essen 1994, pp. 16-23.
[3.81] IEEE, Special issue on: Plastic Encapsulated Microcircuits, *IEEE Trans. Rel.*, 42(1993)4.
[3.82] Jacob P., Held M., Scacco P., "Reliability Testing and Analysis of IGBT Power Semiconductor Modules ", *Proc.* ISTFA'94, Los Angeles CA 1994, pp. 319 - 325.
[3.83] Lau J., Harkins G., Rice D., Kral J., "Thermal fatigue reliability of SMT packages and interconnections", *Proc. Int. Rel. Phys. Symp.*, 1987, pp. 250-259.
[3.84] Lin R., Blackshear E.., Serisky P., "Moisture induced cracking in plastic encapsulated SMD during solder reflow process", *Proc. Int. Rel. Phys. Symp.*, 1988, pp. 83-89.
[3.85] Pecht M., "A model for moisture induced corrosion failures in microelectronic packages", *IEEE Trans. Comp., Pack., and Manuf. Technol.*, 13(1990)2, pp. 383-389; (Ed.) *Handbook of Electronic Package Design*, 1991, Dekker, New York; – and Ramappan V., "Are components still the major problem: a review of electronic system and device field failure reurns", *IEEE Trans. Comp., Pack., and Manuf. Technol.*, 15(1992)6, pp. 1160-1164.
[3.86] Philofsky E., "Design limits when using Au-Al bonds", *Proc. Int. Rel. Phys. Symp.*,1986, pp.114-19.
[3.87] Solomon H.D. et al. (Ed.), *Low Cycle Fatigue, 1988,* ASTM, Philadelphia.
[3.88] Weber L., *Material- & Schädigungsmodell bei Pb-Zn-Ag-Lot*,1996, Rep.K10, ETH Zurich,Rel. Lab.; *Creep-fatigue behaviour of eutectic Sn62Pb36Ag2 solder*, 1997, Ph.D.Thesis 12251, ETH Zurich.

see also [8.1 to 8.14] and [8.21 to 8.34]

4 Maintainability Analysis

[4.1] Bentz R.W., "Pitfalls to avoid in maint. testing", *Proc. Ann. Rel. & Maint. Symp.*, 1982, pp. 278-82.
[4.2] Birolini A., "Spare parts reservation of components subjected to wear-out and/or fatigue according to a Weibull distribution", *Nuclear Eng. & Design*, 27(1974), pp. 293-298.
[4.3] Blanchard B.S. , Lawery E.E., *Maintainability, Principles and Practices*, 1969, McGraw-Hill, N.Y.
[4.4] Collett R.E. and Bochant P.W., "Integration of BIT effectiveness with FMECA", *Proc. Ann. Rel. & Maint. Symp.*, 1984, pp. 300-305.
[4.5] Dobbins J.,"Error-correcting-code memory rel. calculations", *IEEE Trans. Rel.*,35(1986), pp.380-84.
[4.6] Hofstädt H., Gerner M., "Qualitative testability analysis and hierarchical test pattern generation: A new approach to design for testability", *Proc. Int. Test. Conf.*, 1987, pp. 538-546.
[4.7] IEC 60706: *Guide on Maintainability of Equipment* (Parts 1-4), 1982-92.
[4.8] IEEE Special issues on: Maintainability, *Trans. Rel.*, 30(1981)3; Fault-tolerant computing, *Computer*, 17(1984)8, *Trans. Rel.*, 36(1987)2, *Trans. Comp.*, 39(1990)4 & 41(1992)5; Fault tolerance in VLSI, *Proc. of the IEEE*, 74(1986)5; Testing, *Trans. Ind. El.*, 36(1989)2; Software tools for hardware test., *Computer*, 22(1989)4; Fault-tolerant systems, *Computer*, 23(1990)7.
[4.9] IEEE STD 1149.1: *Test Access Part and Boundary-Scan Architecture*, 1990.
[4.10] Lee K.W., Tillman F.A., Higgins J.J., "A literature survey of the human reliability component in a man-machine system", *IEEE Trans. Rel.*, 37(1988), pp. 24-34.

[4.11] Lee P.A., Anderson T., *Fault Tolerance, Principles and Practice*, 2nd Ed. 1990, Springer, Berlin.
[4.12] McCluskey E.J., *Logic Design Principles*, 1986, Prentice-Hall, Englewood Cliffs NJ.
[4.13] Pradhan D.K. (Ed.), *Fault-Tolerant Computing*, Vol.1&2, 1986, Prentice-Hall, Englewood Cliffs NJ.
[4.14] Retterer B.L., Kowalski R.A., "Maintainability - a historical perspective", *IEEE Trans. Rel.*, 33(1984)1, pp. 56-61.
[4.15] Rigby L.V., "Results of eleven sep. maintainability dem.", *IEEE Trans. Rel.*, 16(1967)1, pp.43-48.
[4.16] Robach C., Malecha P., Michel G., "CATA - a computer-aided test analysis system", *IEEE Design & Test of Computers Mag.*, (1984)5, pp. 68-79.
[4.17] Savir J., McAnney W.H., "Random pattern testability of delay faults", *IEEE Trans. Comp.*, 37(1988)3, pp. 291-300.
[4.18] Simpson W.R., Sheppard J.W., *System Test- and Diagnosis*, 1995, Kluwer Acad. Press, Boston.
[4.19] VDI 4003 Bl.3: *Allg. Forderungen an ein Sicherungsprogramm: Instandhaltbarkeit*, 1983.
[4.20] Wagner K.D., Chin C.K., McCluskey E.J., "Pseudorandom testing", *IEEE Trans. Comp.*, 36(1987)3, pp. 332-343.
[4.21] Williams T.W., Parker K.P, "Design for testability - a survey", *Proc. IEEE*, 71(1983)1, pp. 98-112; Williams T.W., "VLSI testing", *Computer*, 1984, pp. 126-136; Williams T.W. (Ed.), *VLSI Testing* (Vol. 5 of Advances in CAD for VLSI), 1986, North Holland, Amsterdam.
[4.22] Willimann B., *Optimale Auslegung der Logistik kompl. Syst.*, 1993, Ph.D.Thesis10313, ETH Zurich.

5 Design Rules for Reliability, Maintainability, and Software Quality

Design Guidelines for Reliability

[5.1] Boxleitner W., *Electrostatic Discharge and Electron. Equip.*, 1989, IEEE Press, Piscataway NJ.
[5.2] Catrysse J., "PCB & system design under EMC constraints", *Proc. 11th Int. Zurich EMC Symposium*, 1995, pp. 47-58.
[5.3] Deutsch A., "Electrical characteristics of interconnections for high-performance systems", *Proc. IEEE*, 86(1998)2, pp. 315-355.
[5.4] Gardner J.R., "The appropriateness of plastic encapsulated microcircuits in a specific wooden-round application", *IEEE Trans. Rel.*, 45(1996)1, pp. 10-17.
[5.5] Goedbloed J.J., *Electromagnetic Compatibility*, 1992, Prentice Hall, New York.
[5.6] Haseloff E., *Was nicht im Datenblatt steht*, 1992, Appl.-Bericht EB 192, Texas Instruments, Freising; "Entwicklungsrichtlinien für schnelle Logikschaltungen und Systemen", *Proc. ETH/IEEE Conf. on Design Rules for Rel., EMC, Maint., Soft. Qual.*, 1993, ETH Zurich, Rel.Lab., pp 5.1-5.17.
[5.7] Hellström S., *ESD-The Scourge of Electronics*, 1998, Springer, Berlin.
[5.8] Hirschi W., "EMV gerechte Auslegung elektron. Geräte", *Bull. SEV/VSE*, 83(1992)11, pp. 25-29.
[5.9] IEEE Special issues on: Design for Reliability of Telecom. Systems, *IEEE Trans. Rel.*, 40(1991)3; Design for Reliability, *IEEE Trans. Rel.*, 44(1995)2.
[5.10] IEEE Std 1100-1992: *IEEE Recom. Practice for Powering and Grounding Sensitive El. Equip.*
[5.11] Mannone P., " Careful design methods to prevent CMOS latch-up", *EDN*, Jan. 26, 1984, 6 pp.
[5.12] IPC, ANSI/IPC-SM-782: *Surface Mount Land Patterns (Config. and Design Rules)*, 1987.
[5.13] Ott H.W., *Noise Reduction Techniques in Electronic Systems*, 1976, Wiley, New York.
[5.14] RAC, SOAR-6: *ESD Control in the Manuf. Envir*, 1986; TR-82-172: *Thermal Guide for Rel. Eng.*, 1982; VZAP: *ESD Susceptibility Data*, 1991.
[5.15] Sergent J., Krum Al, *Thermal Management Handbook*, 1998, McGraw-Hill, New York.
[5.16] Solberg V., *Design Guidelines for Surface Mount and Fine Pitch Technolgy*, 1996, McGraw-Hill, New York.
[5.17] Vinson J.E., Liou J., "Electrostatic disch. in semicond dev", *Proc. IEEE*, 86(1998)2, pp.399-418.
[5.18] White D.R.J., *EMI Control in the Design of Printed Circuit Boards and Backplanes*, 1982, Interf. Control Tech., Gainesville VI.

see also [1.17, 1.21, 2.10 to 2.15, 3.1 to 3.88, 8.1 to 8.14]

Design Guidelines for Maintainabilty

[5.21] Abramovici M., Breuer M.A., Friedman A.D., *Digital System Testing and Testable Design*, 1990, Comp. Scient. Press, New York.
[5.22] Bennetts R.G., *Design of Testable Logic Circuits*, 1984, Addison-Wesley, London.
[5.23] DoD, AMCP-706-132: *Engineering Design Handbook - Maintenance Engineering Tech., 1975*; -133: *Engineering Design Handbook - Maintainability Eng. Theory and Practice*, 1975.
[5.24] Lala P.K., *Fault Tolerant and Fault Testable Hardware Design*, 1985, Prentice-Hall, Engl. Cliffs NJ.
[5.25] Maunder C., *The Board Designer's Guide to Testable Logic Circuits*, 1992, Addison-Wesley, Reading MA; A universal framework for managed Built-In Test, *Proc. Int. Test Conf.*, Paris 1995, 8 pp.
[5.26] Richards D.W., Klion J., "Smart BIT - an approach to better system-level built-in test", *Proc. Ann. Rel. & Maint. Symp.*, 1987, pp. 31-34.
[5.27] Robinson G., Deshayes J., "Interconnect testing of boards with partial boundary-scan", *Proc. Int. Test Conf.*, 1990, paper 27.3.
[5.28] Smith D.J. and Babb A.H., *Maintainability Engineering*, 1973, Pitman, London.

see also [4.3]

Software Quality

[5.31] ACM Special issues on: Software Testing, *Commun. of the ACM*, 31(1988)6; Software Quality, *Commun. of the ACM*, 36(1993)11.
[5.32] Adrion R.W., Branstad M.A., Cherniavsky J.C., "Validation, verification, and testing of computer software", *ACM Computing Surveys*, 14(1982)2, pp. 159-192.
[5.33] Arlat J., Karama K., Laprie J.C., "Dependability modeling and evaluation of software fault-tolerant systems", *IEEE Trans. Comp.*, 39(1990)4, pp. 504-513.
[5.34] Bergland G.D., "A guided tour of program design methodologies", *Computer*, 14(1990)10, pp.13-37.
[5.35] Boehm B.W., "Verifying and validating software requirements and design specifications", *Software*, 1(1984)1, pp. 75-88; "Improving software productivity", *Computer*, 20(1987) 9, pp. 43-57; Spiral model of software development and enhancement", *Computer*, 21 (1988)5, pp. 61-72.
[5.36] Brocklehurst S., Chan P.Y., Littlewood B., Snell J., "Recalibrating software reliability models", *IEEE Trans. Soft. Eng.*, 16(1990)4, pp. 458-469.
[5.37] BWB, *Software-Entwicklungsstandard der BWB - Vorgehensmodell*, 1991.
[5.38] ESA PSS-05-04: *Guide to the Software Architect*, 1992; -05: *Detailed Design and Prod.*, 1992; -08: *Project Management*, 1994; -09: *Configuration Manag.*, 1992; -11: *Quality Assurance*, 1993.
[5.39] Fenton N., Littlewood B., *Software Reliability and Metries*, 1991, Elsevier, London.
[5.40] Gomaa H., "A software design meth. for real-time systems", *Comm. ACM*, 27(1984)9, pp. 938-949.
[5.41] Grady R., "Practical res. from measur. soft. quality", *Commun. of the ACM*, 36(1993)11, pp.62-68.
[5.42] Hansen M.D., "Survey of available software-safety analysis techniques", *Proc. Ann. Rel. & Maint. Symp.*, 1989, pp. 46-49.
[5.43] Herrmann D.S., Peercy D.E., "The bridge between hardware, software, and system safety and reliability", *Proc. Ann. Rel. & Maint. Symp.*, 1999, pp. 396-402.
[5.44] IEC, 60300-3-6: *Depend. Manag. - Software Aspects of Dependability*, 1997.
[5.45] IEEE Special issues on: Software quality assurance, *Computer*, 12(1979)8; Fault Tolerant computing, *Computer*, 17(1984)8; Rapid prototyping, *Computer*, 22(1989)5; Verification and validation, *Software*, May 1988; Software Reliability, *IEEE Trans. Rel.*, 28(1979)3; Software eng. project manag., *IEEE Trans. Soft. Eng.*, 10(1984)1; Experimental computer science, *IEEE Trans. Soft. Eng.*, 16(1990) 2 ; Software Quality, *IEEE Software*, Jan. 1996.
[5.46] *IEEE Software Eng. Std.: Vol. 1 to Vol. 4*, 1999, see [A2.7].
[5.47] Kline M.B., "Software and Hardware R&M - what are the differences?", *Proc. Ann. Rel. & Maint. Symp.*, 1980, pp. 179-184.
[5.48] Leveson N.G.,, "Software safety in computer-controlled systems", *Computer*, (1984)2, pp. 48-55; "Software safety: why, what, and how", *ACM Computing Surveys*, 18(1986)2, pp. 125-163.
[5.49] Littlewood B., Strigini L., "The risk of software". *Scient. Amer.*, 1992, pp. 38-43; "Validation of ultrahigh dependability for software-based syst", *Commun. of the ACM*, 36(1993)11, pp. 69-80.
[5.50] Littlewood B., "Evaluation of software reliability - achievements and limitations", *Proc. ETH/IEEE Int. Symp. on Rel. Eng. 2'000*, ETH Zurich, Rel. Lab., Oct. 17, 1996, 22 pp.

[5.51] Musa J.D., Iannino A., Okumoto K., *Software Reliability: Measurement, Prediction, Application*, 1987, McGraw-Hill, New York.

[5.52] Parnas D.L., van Schouwen A.J., Kwan S.P., "Evaluation of safety-critical software", *Commun. of the ACM*, 33(1990)6, pp. 636-648.

[5.53] Pflegger S.L., "Measuring software reliability", *IEEE Spectrum,* Aug. 1992, pp. 56-60

[5.54] Reifer D.J., Software Failure Modes and Effects Anal.", *IEEE Trans. Rel.*, 28(1979)3, pp.247-49.

[5.55] SAQ, 10300: *Software Qualitätssich. und CASE*, 1995; 10301: *HDBK Beschaffung von Soft,* 1996; 10302: *HDBK für Audits im Soft.-Bereich*,1996; 10303: *Grundl. zum Umgang mit Soft.-Probl.*,1997.

[5.56] Singpurwalla N.D., "The failure rate of software: does it exist?", *IEEE Trans. Rel.*, 44(1995)3, pp. 463-469.

[5.57] Stankovic J.A., "A serious problem for next-generation system", *Computer,* 21(1988)10, pp.10-19.

[5.58] Wallace D. et al, "An analysis of selected software safety Std.", *IEEE AES Mag.*, 1992, pp. 3-14.

6 Reliability and Availability of Repairable Equipment and Systems

[6.1] Ascher H., Feingold H., *Repairable Systems Reliability*, 1984, Dekker, New York.

[6.2] Barlow R.E., Proschan F., *Mathematical Theory of Reliability*, 1965, Wiley, New York; *Statistical Theory of Reliability and Life Testing*, 1975, Holt Rinehart & Winston, New York.

[6.3] Beichelt F., Franken P., *Zuverlässigkeit und Instandhaltung - Math. Methoden*, 1983, Technik, Berlin; Beichelt F., *Zuverlässigkeits- und Instandhaltbarkeitstheorie*, 1993, Teubner, Stuttgart.

[6.4] Birolini A., "Comments on renewal theoretic aspects of two-unit redundant systems", *IEEE Trans. Rel.,* 21(1972)2, pp 122-123; "Generalization of the expressions for the rel. and availability of rep. items", *Proc. 2. Int. Conf. on Struct. Mech. in Reactor Techn.*, Berlin: 1973, Vol. VI, pp. 1-16; "Some appl. of regen. stoch. processes to reliability theory - part two: rel. and availability of 2-item redundant systems", *IEEE Trans. Rel.*, 24(1975)5, pp. 336-340; *On the Use of Stochastic Proc. in Modeling Rel. Problems,* 1985, Springer, Berlin (Lecture Notes in Ec. and Math. Systems Nr. 252); *Quality and Reliability of Technical Systems,* 2nd Ed. 1997, Springer, Berlin; *Zuverlässigkeit von Geräten und Systemen, 4th Ed.* 1997, Springer, Berlin.

[6.5] Bobbio A., Roberti L., "Distribution of the minimal completition time of parallel tasks in multi-reward semi-Markov models", *Performance Eval.*, 14(1992), pp. 239-256.

[6.6] Brenner A., *Performability and Dependability of Fault-Tolerant Systems, 1996,* Ph. D. Thesis 11623, ETH Zurich.

[6.7] Choi, C.Y. et al, "Safety issues in the comparative analysis of dependable architectures", *IEEE Trans. Rel.*, 46(1997)3, pp. 316-322.

[6.8] Dhillon B.S., Rayapati S.N., "Common-cause failures in repairable systems", *Proc. Ann. Rel. & Maint. Symp.*, 1988, pp. 283-289.

[6.9] Dyer D., "Unification of rel./availab. mod. for Markov syst.", *IEEE Trans. Rel.*, 38(1989)2, pp.246-52.

[6.10] Gaede K.W., *Zuverlässigkeit Mathematische Modelle,* 1977, Hanser, Munich.

[6.11] Gnedenko B.V., Beljajev J.K., Soloviev A.D., *Mathematical Methods of Reliability Theory,* 1969, Academic, New York (1968, Akademie, Berlin).

[6.12] Kovalenko I., Birolini A., "Uniform exponential boundes for the availability of a repairable system", in *Exploring Stochastic laws, Homage to V.S. Korolyuk*, 1995, VSP, Utrecht, pp.233-242.

[6.13] Kullstam A., "Availability, MTBF and MTTR for repairable M-out-of-N Systems", *IEEE Trans. Rel.*, 30(1981)4, pp. 393-394.

[6.14] Kumar A., Agarwal M., "A review of standby red. syst", *IEEE Trans. Rel.*, 29(1980)4, pp. 290-294.

[6.15] Osaki S., Nakagawa T., "Bibliography for reliability and availability of stochastic systems", *IEEE Trans. Rel.* 25(1976)4, pp. 284-287.

[6.16] Rai S., Agrawal D.P. (Ed.), *Advances in Distributed Systems Reliability* and *Distributed Computing Network Reliability*, 1990, IEEE Press, Piscataway NJ.

[6.17] Ravichandran N., *Stochastic Methods in Reliability Theory*, 1990, Wiley Eastern, New Dehli.

[6.18] Schneeweiss W.G., "Mean time to first failure of repairable systems with one cold spare", *IEEE Trans. Rel.*, 44(1995)4, pp. 567-574.

[6.19] Ushakov I.A., Harrison R., *Handbook of Reliability Engineering*, 1994, Wiley, New York.

see also [2.2, 2.33, 2.37, 2.40 to 2.44, A7.2, A7.4, A7.10, A7.20, A7.26 to A7.28]

7 Statistical Quality Control and Reliability Tests

Statistical Quality Control

[7.1] ANSI Z1.1 and Z1.2-1958: *Guide for Quality Control and Control Chart Method of Analyzing Data*; Z1.3-1959: *Control Chart Method of Controlling Quality During Production.*

[7.2] IEC 60410: *Sampling Plans and Procedures for Inspection by Attributes,* 1973; see also MIL-STD-105, DIN 40080, DGQ-SAQ-OeVQ 16-01, ISO 2859.

[7.3] SAQ-DGQ-OeVQ, DGQ16-01: *Attributprüfung* 9th Ed. 1986; 16-26: *Meth. zur Ermittlung geeigneter AQL-Werte.* 4rd Ed. 1990; 16-31/-32/-33: *SPC 1/2/3 Statistische Prozesslenkung*, 1990.

[7.4] Sarkadi K.,Vincze I., *Mathematical Methods of Statistical Quality Control*, 1974,Academ. Press, N.Y.

[7.5] Uhlmann W., *Statistische Qualitätskontrolle*, 2nd Ed. 1982, Teubner, Stuttgart.

Statistical Reliability and Maintainability Tests

[7.11] Ascher H.E., Hansen C.K., "Spurious exponentiality observed when incorrectly fitting a distribution of nonstationary data", *IEEE Trans. Rel.,* 47(1998)4, pp.451-459.

[7.12] IEC 60605: *Equipment Rel. Testing* (Part 1 to 4, 6), 1978 - 97; 60706: *Guide on Maintainability of Eq.* (Parts 5-6), 1994; 61070: *Compliance Test Proc. for Steady-State Availability*, 1991; 61123: *Rel. Testing - Compl. Test Plans for Success Ratio,* 1991; 61124: *Rel. Testing - Compl. Test Plans for Const. Failure Rate,* 1997; 61649: *Goodness-of-fit Tests & Conf. Intervals for Weibull Distr. Data,* 1997; 61650: *Rel. Data Analysis Techniques,* 1997; see also 60319: *Presentation and Spec. of Rel. Data*, 1978, rev. draft 1998.

[7.13] Hu J.M., Dasgupta A., Arora A.K., "Rate of failure-mechanisms identification in accelerated testing", *Proc. Ann. Rel. & Maint. Symp.*, 1992, pp. 181-188.

[7.14] Klinger D.J., "On the notion of activation energy in reliability: Arrhenius, Eyring, and thermodynamics", *Proc. Ann. Rel. & Maint. Symp.*, 1991, pp. 295-300.

[7.15] MIL-STD-471: *Maintainability Verif., Demonstr. Eval.*, Ed. A 1973 (see also -HDBK-472); -STD- 781: *Rel. Testing for Eng. Dev., Qualif., and Product.,* Ed. D 1986 (see also -HDBK -781).

[7.16] Nelson W., *Accelerated Testing*, 1990, Wiley, New York.

[7.17] Peck D.S., Trapp O.D., *Accelerated Testing HDBK*, 1987, Techn. Ass., Portola Valley CA.

[7.18] Shaked M., Singpurwalla N.D., "Nonparametric estimation and goodness-of-fit-testing of hypotheses for distributions in accelerated life testing", *IEEE Trans.Rel.,* 31(1982)1, pp. 69-74.

[7.19] Thomas E.F., "Reliability testing pitfalls", *Proc. Ann. Rel. & Maint. Symp.*, 1974, pp. 78-83.

[7.20] Viertl R., *Statistical Methods in Accelerated Life Testing*, 1988, Vandenhoeck, Göttingen.

see also [1.23, A8.1 to A8.30]

8 Quality and Reliability Assurance During the Production Phase

Production Processes

[8.1] Desplas E.P., "Rel. in the manufacturing cycle", *Proc. Ann. Rel. & Maint. Symp.*,1986, pp. 139-144.

[8.2] DGQ 16-31/-32/-33: *SPC 1/2/3 Statistische Prozesslenkung*, 1990.

[8.3] Ellis B.N., *Cleaning and Contamination of Electronics Components and Assemblies*, 1986, Electrochemical Publ., Ayr (Scotland).

[8.4] Hnatek E. R., *Integrated Circuit Quality and* Reliability, 2nd Ed. 1999, Dekker, New York.

[8.5] Grossmann G., "Contamination of various flux-cleaning combinations on SMT assemblies, *Soldering & SMT*, 22 (1996) Feb., pp. 16-21.

[8.6] Lau J.H. et al, "Experimental & statistical analysis of surface-mount technology PLCC solder-joint reliability", *IEEE Trans. Rel.*, 37(1988)5, pp. 524-530.

[8.7] Lea C., *A Scientific Guide to SMT*, 1988, Electrochemical Publ., Ayr (Scotland).

[8.8] Lenz E., *Automatisiertes Löten elektronischer Baugruppen*, 1985, Siemens, Munich.
[8.9] Pawling J.F. (Ed.), *Surface Mounted Assemblies, 1987*, Electrochemical Publ., Ayr (Scotland).
[8.10] Prasad R.P., *Surface Mount Technology*, 1989, Van Nostrand Reinhold, New York.
[8.11] Shewhart W.A., "Quality control charts", *Bell Tech. Jurn.*, 5(1926) pp. 593-603.
[8.12] Scraft R., Bleicher M. (Eds.), *Oberflächenmontage elektron. Bauelemente*, 1987, Hütig, Heidelberg.
[8.13] Vardaman J. (Ed.), *Surface Mount Technology: Recent Jap. Dev*, 1993, IEEE Press, Piscataway NJ.
[8.14] Wassink R.J.K., *Soldering in Electronic*, 2nd Ed. 1989, Electrochemical Publ., Ayr (Scotland).

see also [3.71 to 3.88]

Test and Screening Stategies

[8.21] Bennetts R.G., *Introduction to Digital Board Testing*, 1981, Crane Russak, New York.
[8.22] Birolini A., "Möglichkeiten und Grenzen der Qualifikation, Prüfung und Vorbeh. von ICs", *QZ*, 27(1982)11, pp. 321-326; "Prüfung und Vorbeh. von Bauelementen und Leiterplatten", *VDI-Bericht* Nr. 519, pp. 49-61 1984; "VLSI testing and screening", *Jorunal of Env. Sciences (IES)*, May/June 1989, pp. 42-48; "Matériels électroniques: stratégies de test et de déverminage", *La Revue des Lab. d'Essais*, 1989 pp.18-21; – and Grossmann G., "Experimentelle Ergebnisse zur Qualität und Zuverlässigkeit der SMT mit Pitch 0.5 mm", *me* (1995) 5, pp. 28-33.
[8.23] Bullock M., "Designing SMT boards for in-circuit testab", *Proc. Int. Test Conf.*, 1987, pp.606-613.
[8.24] De Cristoforo R.J., "Environmental stress screening - lesson learned", *Proc. Ann. Rel. & Maint. Symp.*, 1984, pp. 129-133.
[8.25] Desplas E., "Reliability in the manuf. cycle", *Proc. Ann. Rel. & Maint. Symp.*, 1986, pp. 139-144.
[8.26] Geniaux B. et al, *Déverminage des matériels électroniques*, 1986, ASTE, Paris; "Climatique et déverminage", *La Revue des Lab. d'essais*, Sept. 1989, pp. 5-8.
[8.27] IEC 61163: *Rel. Stress Screening - Part 1: Repair. Items*, 1995; *- Part 2: Electronic Comp.*, 1998.
[8.28] IES, *Environmental Stress Screening Guideline for Assemblies*, 1988; *Guidelines for Parts*, 1985; *Environmental Test Tailoring*, 1986; *Environmental Stress Screening*, 1989.
[8.29] Kallis J.M., Hammond W.K., Curry B., "Stress screening of electronic modules: Investigation of effects of temp. rate-of-change", *Proc. Ann. Rel. & Maint. Symp.*, 1990, pp. 59-66.
[8.30] Kindig W., McGrath J., "Vibration, random required", *Proc.Ann.Rel.&Maint.Symp.*,1984, pp.143-47.
[8.31] MIL-HDBK-344: *Environmental Stress Screening of Electronic Equipment*, 1986; see also -HDBK-263, -STD-810, -STD-883, and -STD-2164.
[8.32] Parker P., Webb C., "A study of failures identified during board level environmental stress testing", *IEEE Tran. Comp. Pack., and Manuf. Technol.*, 15(1992)6, pp. 1086-1092.
[8.33] Pynn C., *Strategies for Electronics Test*, 1986, McGraw-Hill, New York.
[8.34] Wennberg S.R., Freyler C.R., "Cost-effective vibration testing for automotive electronic", *Proc. Ann. Rel. & Maint. Symp.*, 1990, pp. 157-159.

see also [3.71 to 3.88]

Reliability Growth

[8.41] Barlow R., Bazovsky I., Wechsler S., "Classical and Bayes approach to ESS- a comparison", *Proc. Ann. Rel. & Maint. Symp.*, 1990, pp. 81-84.
[8.42] Benton A.W., Crow L.A., "Integrated reliability-growth testing", *Proc. Ann. Rel. & Maint. Symp.*, 1990, pp. 160-166.
[8.43] Brinkmann R., *Modellierung des Zuverlässigkeitswachstums komplexer, reparierbarer Systeme*, 1997, Ph. D. Thesis 11903, ETH Zurich.
[8.44] Chay S.C., "Reliability growth during a product development program", *Proc. Ann. Rel. & Maint. Symp.*, 1983, pp. 43-48.
[8.45] Crow L.H., "On tracking reliability growth", *Proc. Ann. Rel. & Maint. Symp. (AR & MS)*, 1975, pp. 438-442; "Methods for assessing rel. growth potential", *AR & MS* , 1982, pp. 74-78; "On the initial system rel.", *AR & MS*, 1986, pp. 115-119; "Evaluating the rel. of repairable systems", *AR & MS*, 1990, pp. 275-279; - and Basu A.P., "Rel. growth estimation with missing data", *AR & MS*, 1988, pp. 248-253; "Confidence intervals on the rel. of rep. systems", *AR & MS*, 1993, pp. 126-134.

[8.46] Duane J.T., "Learning curve appr. to rel. monitoring", *IEEE Trans. Aerosp.*, (1964)2, pp. 563-566.
[8.47] Fries A., Sen A., "A survey of discrete rel.-growth mod." *IEEE Trans. Rel.*, 45(1996)4, pp. 582-604.
[8.48] IEC 61014: *Programs for Reliability Growth*, 1989; 61164: *Reliability Growth - Statistical Tests and Estimation Methods*, 1995.
[8.49] IES, *Reliability Growth Processes and Management*, 1989.
[8.50] Jääskeläinen P., "Rel. growth and Duane learning curves", *IEEE Trans. Rel.*, 31(1982)2, pp.151-54.
[8.51] Jayachandran T., Moore L.R., "A comparison of reliability growth models", *IEEE Trans. Rel.*, 25(1976)1, pp. 49-51.
[8.52] Kasouf G. and Weiss D., "An integrated missile reliability growth program", *Proc. Ann. Rel. & Maint. Symp.*, 1984, pp. 465-470.
[8.53] VDI 4009 Bl.8: *Zuverlässigkeitswachstum bei Systemen*, 1985.
[8.54] Wong K.L., "A new environmental stress screening theory for electronics", *Proc. Ann. Tech. Meeting IES*, 1989, pp. 218-224; "Demonstrating reliability and reliability growth with data from env. stress screening", *Proc. Ann. Rel. & Maint. Symp.*, 1990, pp. 47-52.
[8.55] Yamada S. et al, "Reliability growth models for hardware and software systems based on nonhomogeneous Poisson processes - a survey", *Microel. & Rel.*, 23(1983), pp. 91-112.

A1 Terms and Definitions

[A1.1] EOQC, *Glossary of Terms Used in the Management of Quality*, 6th Ed. 1989.
[A1.2] ESA, ECSS-P-001A, Rev.1: *Definitions*, 1997.
[A1.3] IEC 60050-191: *Int. El. Vocabulary, Chapter 191 - Dependability and Quality of Service*, 1990.
[A1.4] IEEE, ANSI/IEEE Std 100-1997: *IEEE Standard Dictionary of Electrical and Electronics Terms*; Std 610.12-1990: *IEEE Standard Glossary of Software Eng. Terminology.*
[A1.5] ISO 8402: *Quality - Vocabulary*, 1994; ISO/IEC 2382-14: *Information technology vocabulary-Part 14*, 1997.
[A1.6] MIL-STD-109: *Quality Assurance Terms and Definitions*, Ed.B, 1969; -280: *Def. of Item Levels and Related Terms*, Ed.A, 1969; -721: *Def. of Terms for Rel. and Maintainability*, Ed.C, 1981.

A2 Customer Requirements

[A2.1] ANSI Z-1.15-1979: *Generic Guidelines for Quality Systems.*
[A2.2] Benski C., "Dependability Stds: An int. perspective", *Proc. Ann. Rel.&Maint. Symp.*,1996,pp.13-16.
[A2.3] prEN 9000-1-1998: *Aerospace Ind. Quality System* (Part 1), Europ. Ass. of Aerosp. Ind., Bruxelles.
[A2.4] ESA ECSS-M-00: *Space Project Manag.*, Ed. A 1996, see also -10 to -70; -Q-00: *Space Product Assurance*, Ed. A 1996, see also -20(QA), -30(Dependability), -40 (Safety), -60 (El. Components), -70(Mat. & Proc.), -80 (Software); -E-00 *Space Eng.* Ed. A 1996, see also -10 (System Eng.).
[A2.5] IEC 60068: *Environmental Testing* (Parts 1-5), 1969-97; 60300: *Dependability Manag.*, -1 (Program), -2 (Tasks), -3 (Appl. Guides), 1991-97; 60319: *Presentation and Spec. of Rel. Data*, 1978, rev. draft 1998; 60410 (Sampling plans), 1973; 60605: *Eq. Rel. Test*, -1 (Gen. Req.),1982, -2 (Test Cycles), 1994, -3 (Test Cond.)1986-96, -4 (Param. Estim.), 1986. -6 (Const. λ), 1997; 60706: *Maintainability* (Parts 1-6), 1982-94; 60721: Env. Conditions (Parts 1-3), 1982-97; 60749 (Semicond. Tests), 1996; 60812 (FMEA), 1985; 60863 (Rel. Pred.), 1986; 61000 (EMC, Part 1-6), 1990-98; 61014 & 61164 (Rel. Growth), 1989 & 1995; 61025 (FTA), 1990; 61070 (Avail. Valid.), 1991; 61078 (Rel. Block Diagr.), 1991; 61123 (Rel. Valid.), 1991; 61124 (Const. Fail. Rate Valid.), 1997; 61160 (Design Reviews), 1992; 61163: Screening (Parts 1-2), 1995-98; 61165 (Markov. Tech.), 1995; 61649 (Weibull Data), 1997; 61650 (Comp. of two λ), 1997; 61709 (Failure Rates), 1996; see also ISO/IEC 12207 (Software Life Cycle Proc.), 1995.
[A2.6] IEEE Std 1332-1998: IEEE Standard Reliability Program for the Development and Production of Electronic Systems and Equipment.

[A2.7] *IEEE Software Eng. Std.: Vol. 1 to Vol. 4*, 1999 (in particular 610.12, 730, 828, 829, 830, 982.1/.2, 1008, 1012, 1028, 1042, 1045, 1058, 1059, 1061, 1062, 1063, 1074, 1219, 1465), IEEE Standards Assoc., Piscataway NJ; see also ISO/IEC 12207 (Software Life Cycle Processes), 1995.

[A2.8] ISO 9000: *Quality Manag. and Quality Assurance Std.*, Part 1 to 4, 1993-97 (Ed. 2000 announced).
9001 : *Model for QA in Design, Develop., Prod., Install. & Servicing*, 1994 (ed 2000 announced).
9002 : *Model for QA in Production, Install. & Servicing*, 1994 (will be integrated in 9001-2000).
9003 : *Model for Quality Ass. in Final Inspection & Test*, 1994 (will be integrated in 9001-2000).
9004 : *Quality Manag. & Quality System - Guidelines*, Part 1 to 4, 1991-94 (Ed. 2000 announced).
10005: *Quality Management - Guidelines for Quality Plans*, 1995.
10007: *Quality Management - Guidelines for Configuration Management*, 1995.
see also *ISO Standard Compendium: ISO 9000*, 7th Ed. 1998 (announced are ISO 9000-2000 (Fundamentals & Vocabulary), ISO 9001-2000 (Requirements), and ISO 9004-2000 (Guidance)).

[A2.9] MIL-STD-454: *Standard General Req. for Electronic Equip*, Ed. M 1991; see also -470 (Maintainability), - 471 (Maintainability Dem.), - 483, - 499, - 781 (Reliability Testing), - 785 (Reliability Program), - 810(Environmental Test Methods), - 881, - 882 (Safety Progr.), - 1309, - 1388, -1472 (Human Eng.), -1521 (Technical Reviews), -1629 (FMECA), -1686 (ESD), -2164 (ESS), -2165 (Testability Progr.), -45662, -I-45208, -Q-9858 (Quality Progr. Req.), -HDBK-338 (El. Rel. Design), -HDBK-470 & -471 (Maint.), -HDBK-781 (Rel. Test); see also NATO AQAP-1 to -15.

[A2.10] NASA NHB 5300.4 (1A): *Rel. Progr. Prov. for Areo. & Space System Contractors*, 1970; (1B): *Quality Progr. Prov. for Aero. & Space Syst. Contr.*, 1969; (2B): *Quality Assurance Provisions for Government Agen.* 1971; (1D-1): *Safety, Rel., Maint., & Quality Prov. for Shuttle Progr*, 1974.

[A2.11] VDI: 4002, 4003, 4004, 4005, 4007 Bl, 4008, 5-9, 4009, 4010, 1981-86.

Guides for Evaluation

[A2.21] ASQC, *Compendium of Audit Standards* (Ed. by W. Willborn), 1983.

[A2.22] CECC 00804: *Interpretation of EN ISO 9000 - Rel. Aspects for Electronic Comp.*, 1996.

[A2.23] DGQ 12-63: *Systemaudit - Die Beurteilung des QS-Systems*, 1987.

[A2.24] ISO 10011-1: *Guidelines for Auditing Quality Systems-Part 1*, 1990, see also Part 2 to 3 (1991).
10012-1: *Quality Assurance Req. for Meas. Equip.-Part 1*, 1992, see also Part 2 (1997).
10013 : *Guidelines for Developing Quality Manuals*, 1995.

[A2.25] NATO: *AQAP-2, AQAP-5, AQAP-7, AQAP-14*, 1994.

[A2.26] SAQ, 10400: *Leitfaden zur Normreihe EN ISO 9001, 9002, 9003*, 1996; 10405: *Leitfaden ISO 9000 - die grosse Revision*, 1999; 10404: *Leitfaden Integrierte Produktsicherheit*, 1998.

A3 Quality and Reliability Assurance Program

[A3.1] Carrubba E.R., Commercial vs. DoD rel. progr.", *Proc. Ann. Rel. & Maint. Symp.*,1981, pp.289-92.

[A3.2] IEC 60300-1: *Dep. Manag. - Dependability Progr. Manag.*, 1993; -2: *Dependability Progr., Elements and Tasks*, 1995; -3-4: *Guide to the Spec. of Dependability Requirements*, 1996.

[A3.3] IEEE, "Special issue on Management of Reliability Progr", *IEEE Trans. Rel.*, 32(1984)3.

[A3.4] MIL-STD-785: *Rel. Progr. for Systems and Equip. Development and Production*, Ed. B 1980.

[A3.5] VDI 4003 Bl.1: *Anwendung zuverl.-bezogener Progr*, 1985; Bl.2: *Funktionszuverlässigkeit*, 1986.

A4 Design Reviews

[A4.1] ASQC, *Configuration Management*, 1969.

[A4.2] IEC 61160: *Formal Design Review*, 1992.

[A4.3] MIL-STD-1521: *Technical Review and Audits for Syst., Eq. and Comp. Prog*, Ed.B 1985.

[A4.4] Samaras T.T., *Fundamentals of Configuration Management*, 1971, Wiley, New York.

[A4.5] VDI Bericht 192: *Konzeptionen und Verfahrensweisen für Entwurfsüberprüfungen*, 1973.

see also [1.3, 1.5, 1.9, 1.15]

A5 Quality Data Reporting System

[A5.1] ASQC, *A Reliability Guide to Failure Reporting, Analysis, and Corrective Action Systems*, 1977.

[A5.2] Collins J.A., Hagan B.T., Bratt H.M., "Helicopter failure modes and corrective actions", *Proc. Ann. Rel. & Maint. Symp.*, 1975, pp. 504-510.

[A5.3] IEC 60300-3-2: *Dep. Manag. - Guide for the Collection of Dependability Data from Field*, 1993 (see also 60706-3).

[A5.4] MIL-STD-2155: *Failure Reporting, Analysis & Corrective Action System*, 1985.

[A5.5] NASA TND-5009: *An Introduction to Computer-Aided Reliability Data Analysis*, 1969.

[A5.6] Thomas E., "Appl. of unified data base technol", *Proc. Ann. Rel.&Maint. Symp.*,1984, pp.192 - 96.

A6 Probability Theory

[A6.1] Aitchison J., Brown J.A.C., *The Lognormal Distribution*, 1969, Univ. Press, Cambridge.

[A6.2] Breiman L., *Probability*, 1968, Addison-Wesley, Reading MA.

[A6.3] Bühlmann H. et al, *Einführung in die Theorie und Praxis der Entscheidung*, 1969, Springer,Berlin.

[A6.4] Crow E.L, Kunio S., *Lognormal Distributions - Theory and Applications*, 1988, Dekker, NY.

[A6.5] Evans D.M., *Probability and its Applications for Engineers*, 1992, Dekker, Milwaukee.

[A6.6] Feller W., *An Introduction to Probability Theory and its Applications*, Vol I 3th Ed. 1968, Vol II 2nd Ed. 1966, Wiley, New York.

[A6.7] Gnedenko B.W., *Theory of Probability*, 1967, Cheslea, New York; *Einführung in die Wahrscheinlichkeitstheorie*, 1991, Akademie, Berlin.

[A6.8] Gumbel E.J., *Statistical Theory of Extreme Values and Some Practical Applications*, 1954, National Bureau of Standards, Washington.

[A6.9] Johnson N.L, Kotz S., *Distributions in Statistics*, Vol. 1 to 4, 1969 to 1972, Wiley, New York.

[A6.10] Kolmogoroff A.N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933, Springer, Berlin.

[A6.11] Kuhn P.N., *Computational Probability*, 1980, Academic Press, New York.

[A6.12] Laha R.G. and Rohatgi V.K., *Probability Theory*. New York: Wiley, .

[A6.13] Ochi M.K., *Applied Probability and Stochastic Processes*, 1990, Wiley, New York.

[A6.14] Rao M.M., *Probability Theory with Appplications*, 1984, Academic Press, Orlando.

[A6.15] Roberts R.A., *An Introduction to Applied Probability*, 1992, Addison-Wesley, Reading MA.

[A6.16] Rényi A., *Wahrscheinlichkeitsrechnung*, 2nd Ed. 1966, VEB Deut. Verlag der Wiss., Berlin.

[A6.17] Shiryayev A.N., *Probability*, 1984, Springer, New York.

[A6.18] Stark H., Woods J.W., *Probability, Random Processes and Estimation Theory for Engineers*, 1986, Prentice Hall, Englewood Cliffs NJ.

[A6.19] Trivedi K.S., *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, 1982, Prentice Hall, Englewood Cliffs NJ.

[A6.20] Weibull W.,"A statistical distrib. function of wide applicability", *J.Appl. Mech.*,18(1951),pp. 293-97.

A7 Stochastic Processes

[A7.1] Asmussen S., *Applied Probability and Queues*, 1986, Wiley, Chicherster.

[A7.2] Birolini A., "Some appl. of regen. stoch. proc. to rel. theory - Part one and two", *IEEE Trans. Rel.*, 23 (1974) 3, pp. 186-194 and 24(1975)5, pp. 336-340; *Semi-Markoff und verwandte Prozesse: Erzeugung und Anwendungen auf Probleme der Zuverlässigkeits- und Übertragungsth*, 1974, Ph.D. Thesis 5375, ETH Zurich, also in *AGEN-Mitt.*, 18(1975), pp. 3-52; "Hardware simulation of semi-Markov and related proc., *Math. & Comp. in Simul.*, 19(1977), pp. 75-97 and 183-191; *On the Use of Stoch. Proc. in Modeling Rel. Problems*, 1985, Springer, Berlin (Lect. Notes Ec. & Math. Systems Nr. 252); *Quality and Reliability of Technical Systems*, 2nd Ed. 1997, Springer, Berlin; *Zuverlässigkeit von Geräten und Systemen, 4th Ed.* 1997, Springer, Berlin.

[A7.3] Cinlar E., *Introduction to Stochastic Processes*, 1975, Prentice Hall, Englewood Cliffs NJ.

[A7.4] Cox D.R., "The analysis of non-markovian stoch. proc. by the inclusion of sup. variables", *Proc. Cambridge Phil. Soc.*, 51(1955), pp. 433-441; *Renewal Theory*, 1962, Methuen, London.

[A7.5] Cox D.R., Lewis P.A., *The Stat. Analysis of Series of Events*, 2nd Ed. 1968, Methuen, London.

[A7.6] Cramér H., "Model building with the aid of stoch. proc.", *Technometrics*, 6 (1964), pp. 133-159.

[A7.7] Cramér H., Leadbetter M.R., *Stationary and Related Stochastic Processes*, 1967, Wiley,New York.

[A7.8] Doob I.L., *Stochastic Processes*, 7th Ed. 1967, Wiley, New York.

[A7.9] Feller W., "On the integral equation of renewal theory", *Ann. Math. Statistics*,12(1941), pp.243-67; "On semi-Markov-processes", *Proc. Nat. Acad. Scient.* (USA), 51(1964), pp. 653-659; *An Introduction to Prob. Theory and its Applic.*, Vol I 3th Ed. 1968, Vol II 2nd Ed. 1966, Wiley, N.Y.

[A7.10] Franken P., Streller A., "Reliability analysis of complex repairable systems by means of marked point processes", *J. Appl. Prob.*, 17(1980), pp. 154-167; - and Kirsten B.M., "Reliability analysis of complex systems with repair", *EIK*, 20(1984), pp. 407-422.

[A7.11] Franken P. et al, *Queues and Point Processes*, 1981, Akademie, Berlin.

[A7.12] Gnedenko B.W., König D. (Ed.), *Handbuch der Bedienungstheorie*, Vol.I & II 1983, Akad., Berlin.

[A7.13] Gnedenko B.W., Kovalenko I.N., *Introduction to Quening Theory*, 1989, Birkhäuser, Basel.

[A7.14] Grigelionis B.I., "Limit theorems for sums of repair processes", *Cybernetics in the Serv. of Comm.*, 2(1964), pp. 316-341.

[A7.15] Hunter J.J., *Mathematical Techniques of Applied Probability, Vol. 2, Discrete Time Models*, 1983, Academic Press, New York.

[A7.16] Karlin S., McGregor J.L., "The differential equations of birth and death processes, and the Stieltjes moment problem", *Trans. Amer. Math. Soc.*, 86(1957), pp. 489-546; "The classification of birth and death processes", *Trans. Amer. Math. Soc.*, 85(1957), pp. 366-400; "Coincidence properties of birth and death processes", *Pacific J. Math.*, 9(1959), pp. 1109-1140.

[A7.17] Khintchine A.Y., *Mathematical Methods in the Theory of Queuing*, 1960, Griffin, London.

[A7.18] Lévy P., "Processus semi-markoviens", *Proc. Int. Congr. Math. Amsterdam*, 3(1954), pp. 416-26.

[A7.19] Neuts M.F., *Matrix-geometric Solutions in Stochastic Models: an Algorithmic Approach*, 1981, Hopkins Univ. Press, Baltimore MA.

[A7.20] Osaki S. and Hatoyama Y. (Eds.), *Stochastic Models in Reliability Theory*, 1984, Springer, Berlin (Lect. Notes in Ec. and Math. Syst. Nr. 235).

[A7.21] Parzen E., *Stochastic Processes*, 3rd Printing 1967, Holden-Day, San Francisco.

[A7.22] Pavlov I.V., "The asymptotic distribution of the time until a semi-Markov process gets out of a kernel", *Eng. Cybernetics*, (1978)5, pp. 68-72.

[A7.23] Pyke R., "Markov renewal processes: definitions and preliminary properties", *Annals Math. Statistics*, 32(1961), pp. 1231-1242; "Markov renewal proc. with finitely many states", *Annals Math. Stat.*, 32(1961), pp.1243-1259; - and Schaufele R., "Limit theorems for Markov renewal proc.", *Annals Math. Stat.*, 35(1964), pp. 1746-1764; "The existence and uniqueness of stationary measures for Markov renewal proc.", *Annals Math. Stat.*, 37(1966), pp. 1439-1462.

[A7.24] Smith W.L., "Asymptotic renewal theorems", *Proc. Roy. Soc. Edinbourgh*, 64(1954), pp. 9-48; "Regenerative stochastic processes, *Proc. Int. Congress Math. Amsterdam*, 3(1954), pp. 304-305; "Regenerative stochastic processes", *Proc. Roy. Soc. London*, Ser. A, 232(1955), pp. 6-31; "Remarks on the paper: Regenerative stochastic processes", *Proc. Roy. Soc. London*, Ser. A, 256(1960), pp. 496-501; "Renewal theory and its ramifications", *J. Roy. Stat. Soc.*, Ser.B, 20(1958), pp. 243-302.

[A7.25] Snyder D.L., Miller M.I., *Random Point Processes in Time and Space*, 2nd Ed.1991,Springer,Berlin.

[A7.26] Solovyev A.D., "The problem of optimal servicing", *Eng. Cybernetics*, 5(1970), pp. 859-868; "Asymptotic distribution of the moment of first crossing of a high level by a birth and death process", *Proc. sixth Berkeley Symp. Math. Stat. Prob.*, 3(1970), pp. 71-86; "Asymptotic behavior of the time of first occurr. of a rare event in a reg. proc.", *Eng. Cybernetics*, 6(1971)9, pp.1038-048.

[A7.27] Srinivasan S.K., Mehata K.M., *Stochastic processes*, 2nd Ed. 1988, Tata McGraw-Hill, New Delhi.

[A7.28] Störmer H., *Semi-Markoff-Prozesse mit endlich vielen Zuständen*, 1970, Springer, Berlin (Lect. Notes in Op. Res. and Math. Syst. Nr. 34).

[A7.29] Takács L., "On certain sojurn time problems in the theory of stoch. proc.", *Acta Math. (Hungar)*, 8(1957), pp.169-191; Stochastic processes, problems and solutions, 4th Ed.1968, Methuen,London.

[A7.30] Thompson W.A., Jr., *Point Processes Models with Applications to Safety and Reliability*, 1988, Chapman & Hall, New York.

see also [6.1 to 6.19]

A8 Mathematical Statistics

[A8.1] Bain L., Engelhardt M., *Statistical Analysis of Rel. and Life-Testing Models*, 1991, Dekker NY.

[A8.2] Barlow R.E., Bartholomew D.J., Bremner J.M., Brunk H.D., *Statistical Inference Under Order Restrictions*, 1972, Wiley, New York; Barlow R.E., Campo R.A., *Total Time on Test Processes and Applications to Failure Data Analysis*, 1975, Tech. Rep. 75-0195, Aerospace Res. Lab., Ohio.

[A8.3] Belyayev J.K., "Unbiased estimation of the parameter of an exponential distribution", *Eng. Cybernetics*, (1983)3, pp.78-81.

[A8.4] Birnbaum Z.W., "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size", *Annals Stat. Ass.*, 47(1952), pp. 425-441.

[A8.5] Cantelli F.P., "Considerazioni sulla legge uniforme dei grandi numeri e sulla generalizzazione di un fondamentale teorema del Sig. Paul Lévy", *Giornale Attuari*, 1933, pp. 327-338; "Sulla determinazione empirica delle leggi di probabilità", *Giorn. Attuari*, 1933, pp. 421-424.

[A8.6] Chernoff H., Lehmann E.L., "The use of maximum Likelihood estimates in χ^2 goodness-of-fit", *Ann. Math. Stat.*, 25(1954), pp. 579-586.

[A8.7] Clopper C.J., Pearson E.S., "The use of confidence or fiducial limits illustrated in the case of the binomial", *Biometrika*, 26(1934), pp. 404-413.

[A8.8] Cochran W.G., "The $\chi2$ tests of goodness of fit", *Ann. Math. Stat.*, 23(1952), pp. 315-345.

[A8.9] Cramér H., *Mathematical Methods of Statistics*, 1958, Univ. Press, Princeton.

[A8.10] d'Agostino R.B., Stephens M. A., *Goodness-of-fit-Techniques*, 1986, Dekker, New York.

[A8.11] Darling D.A., "The Kolmogorov-Smirnov, Cramer-von Mises tests", *Ann. Math. Stat.*, 28(1957), pp. 823-838.

[A8.12] Epstein B, et al, "Life testing", *J. Amer. Stat. Ass.*, 48(1953), pp. 486-502; "Sequential life tests in the exponential case", *Ann. Math. Stat.*, 26(1955), pp. 82-93; "The exact analysis of sequential life tests with particular application to AGREE plans", *Rel. & Maint. Conf.*, 1963, pp. 284-310.

[A8.13] Feller W., "On the Kolmogorov-Smirnov limit theorems for empirical distributions", *Ann. Math. Stat.*, 19(1948), pp. 177-189.

[A8.14] de Finetti B., "Sull'approssimazione empirica di una legge di probabilità", *Giorn. Attuari*, 1933, pp. 415-420.

[A8.15] Fisher R.A.,"On the mathematical foundations of theoretical statistics", *Phil. Trans.*, A 222(1921), pp. 309-368; "The conditions under which χ^2 measures the discrepancy between observation and hypothesis", *J. Roy Stat. Soc.*, 87(1924), pp. 442-450; "Theory of statistical estimation", *Proc. Cambridge Phil. Soc.*, 22(1925), pp. 700-725.

[A8.16] Gliwenko V.,"Sulla determinazione emp. delle leggi di probabilità", *Giorn. Attuari*, 1933, pp.92-99.

[A8.17] Gumbel E.J., *Statistics of Extremes*, 1958, Columbia Univ. Press, New York.

[A8.18] Heinhold J., Gaede K.W., *Ingenieur-Statistik*, 4th Ed. 1979, Oldenbourg, Munich.

[A8.19] Kalbfleisch J.D., Prentice R.L., *Statistical Analysis of Failure Time Data*, 1988, Wiley, New York.

[A8.20] Kolmogoroff A.N., "Sulla determinazione empirica di una legge di distribuzione", *Giorn. Attuari*, 1933, pp. 83-91.

[A8.21] Lawless J. F., *Statistical Models and Methods for Lifetime Data*, 1982, Wiley, New York.

[A8.22] Lehmann E.L., *Testing Statistical Hypotheses*, 1959, Wiley, New York.

[A8.23] Mann N.R., Schafer R.E., Singpurwalla N.D., *Methods for Statistical Analysis of Reliability and Life Data*, 1974, Wiley, New York.

[A8.24] Martz H.F. and Waller R.A., *Bayesian Reliability Analysis*, 1982, Wiley, New York.

[A8.25] Miller L.H., "Table of % points of Kolmogorov statistics", *J. Amer. Stat. Ass.*, 51(1956), pp.111-121.

[A8.26] Pearson K., "On deviations from the probable in a correlated system of variables", *Phil. Magazine*, 50(1900), pp. 157-175.

[A8.27] Rohatgi V.K., *An Introduction to Probability Theory and Mathematical Statistics*, 1976, Wiley, New York.

[A8.28] Serfling R.J., *Approximation Theorems of Mathematical Statistics*, 1980, Wiley, New York.

[A8.29] Smirnov N., "On the estimation of the discrepancy between empirical curves of distribution for two independent samples", *Bull. Math. Moscow Univ.*, 2(1939), fasc. 2

[A8.30] Wald A., *Sequential Analysis* 1947, Wiley, New York; *Statistical Decision Functions*, 1950, Wiley, New York.