3

REDUNDANCY, SPARES, AND REPAIRS

3.1 INTRODUCTION

This chapter deals with a variety of techniques for improving system reliability and availability. Underlying all these techniques is the basic concept of redundancy, providing alternate paths to allow the system to continue operation even when some components fail. Alternate paths can be provided by parallel components (or systems). The parallel elements can all be continuously operated, in which case all elements are powered up and the term *parallel redundancy* or *hot standby* is often used. It is also possible to provide one element that is powered up (on-line) along with additional elements that are powered down (standby), which are powered up and switched into use, either automatically or manually, when the on-line element fails. This technique is called *standby redundancy* or *cold redundancy*. These techniques have all been known for many years; however, with the advent of modern computer-controlled digital systems, a rich variety of ways to implement these approaches is available. Sometimes, system engineers use the general term *redundancy management* to refer to this body of techniques. In a way, the ultimate cold redundancy technique is the use of spares or repairs to renew the system. At this level of thinking, a spare and a repair are the same thing—except the repair takes longer to be effected. In either case for a system with a single element, we must be able to tolerate some system downtime to effect the replacement or repair. The situation is somewhat different if we have a system with two hot or cold standby elements combined with spares or repairs. In such a case, once one of the redundant elements fails *and we detect the failure*, we can replace or repair the failed element while the system continues to operate; as long as the

Reliability of Computer Systems and Networks

MARTIN L. SHOOMAN

replacement or repair takes place before the operating element fails, the system never goes down. The only way the system goes down is for the remaining element(s) to fail before the replacement or repair is completed.

This chapter deals with conventional techniques of improving system or component reliability, such as the following:

1. Improving the manufacturing or design process to significantly lower the system or component failure rate. Sometimes innovative engineering does not increase cost, but in general, improved reliability requires higher cost or increases in weight or volume. In most cases, however, the gains in reliability and decreases in life-cycle costs justify the expenditures.

2. Parallel redundancy, where one or more extra components are operating and waiting to take over in case of a failure of the primary system. In the case of two computers and, say, two disk memories, synchronization of the primary and the extra systems may be a bit complex.

3. A standby system is like parallel redundancy; however, power is off in the extra system so that it cannot fail while in standby. Sometimes the sensing of primary system failure and switching over to the standby system is complex.

4. Often the use of replacement components or repairs in conjunction with parallel or standby systems increases reliability by another substantial factor. Essentially, once the primary system fails, it is a race to fix or replace it before the extra system(s) fails. Since the repair rate is generally much higher than the failure rate, the repair almost always wins the race, and reliability is greatly increased.

Because fault-tolerant systems generally have very low failure rates, it is hard and expensive to obtain failure data from tests. Thus second-order factors, such as common mode and dependent failures, may become more important than they usually are.

The reader will need to use the concepts of probability in Appendix A, Sections A1–A6.3 and those of reliability in Appendix B3 for this chapter. Markov modeling will appear later in the chapter; thus the principles of the Markov model given in Appendices A8 and B6 will be used. The reader who is unfamiliar with this material or needs review should consult these sections.

If we are dealing with large complex systems, as is often the case, it is expedient to divide the overall problem into a number of smaller subproblems (the "divide and conquer" strategy). An approximate and very useful approach to such a strategy is the method of apportionment discussed in the next section.

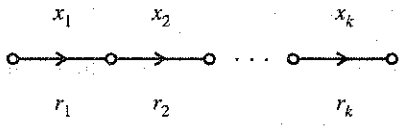


Figure 3.1 A system model composed of k major subsystems, all of which are necessary for system success.

3.2 APPORTIONMENT

One might conceive system design as an optimization problem in which one has a budget of resources (dollars, pounds, cubic feet, watts, etc.), and the goal is to achieve the highest reliability within the constraints of the available budget. Such an approach is discussed in Chapter 7; however, we need to use some of the simple approaches to optimization as a structure for comparison of the various methods discussed in this chapter. Also, in a truly large system, there are too many possible combinations of approach; a top–down design philosophy is therefore useful to decompose the problem into simpler subproblems. The technique of apportionment serves well as a "divide and conquer" strategy to break down a large problem.

Apportionment techniques generally assume that the highest level—the overall system—can be divided into 5–10 major subsystems, all of which must work for the system to work. Thus we have a series structure as shown in Fig. 3.1.

We denote x_1 as the event success of element (subsystem) 1, x_1' is the event failure of element 1, $P(x_1) = 1 - P(x_1')$ is the probability of success (the reliability, r_1). The system reliability is given by

$$R_s = P(x_1 \cap x_2 \cdots \cap x_k) \tag{3.1a}$$

and if we use the more common engineering notation, this equation becomes

$$R_s = P(x_1 x_2 \cdots x_k) \tag{3.1b}$$

If we assume that all the elements are independent, Eq. (3.1a) becomes

$$R_s = \prod_{i=1}^{k} r_i \tag{3.2}$$

To illustrate the approach, let us assume that the goal is to achieve a system reliability equal to or greater than the system goal, R_0, within the cost budget, c_0. We let the single constraint be cost, and the total cost, c, is given by the sum of the individual component costs, c_i.

$$c = \sum_{i=1}^{k} c_i \tag{3.3}$$

We assume that the system reliability given by Eq. (3.2) is below the system specification or goal, and that the designer must improve the reliability of the system. We further assume that the maximum allowable system cost, c_0, is generally sufficiently greater than c so that the system reliability can be improved to meet its reliability goal, $R_s \geq R_0$; otherwise, the goal cannot be reached, and the best solution is the one with the highest reliability within the allowable cost constraint.

Assume that we have a method for obtaining optimal solutions and, in the case where more than one solution exceeds the reliability goal within the cost constraint, that it is useful to display a number of "good" solutions. The designer may choose to just meet the reliability goal with one of the suboptimal solutions and save some money. Alternatively, there may be secondary factors that favor a good suboptimal solution. Lastly, a single optimum value does not give much insight into how the solution changes if some of the cost or reliability values assumed as parameters are somewhat in error. A family of solutions and some sensitivity studies may reveal a good suboptimal solution that is less sensitive to parameter changes than the true optimum.

A simple approach to solving this problem is to assume an equal apportionment of all the elements $r_i = r_1$ to achieve R_0 will be a good starting place. Thus Eq. (3.2) becomes

$$R_0 = \prod_{i=1}^{k} r_i = (r_1)^k \tag{3.4}$$

and solving for r_1 yields

$$r_1 = (R_0)^{1/k} \tag{3.5}$$

Thus we have a simple approximate solution for the problem of how to apportion the subsystem reliability goals based on the overall system goal. More details of such optimization techniques appear in Chapter 7.

3.3 SYSTEM VERSUS COMPONENT REDUNDANCY

There are many ways to implement redundancy. In Shooman [1990, Section 6.6.1], three different designs for a redundant auto-braking system are compared: a split system, which presently is used on American autos either front/rear or LR–RF/RR–LF diagonals; two complete systems; or redundant components (e.g., parallel lines). Other applications suggest different possibilities. Two redundancy techniques that are easily classified and studied are component and system redundancy. In fact, one can prove that component redundancy is superior to system redundancy in a wide variety of situations.

Consider the three systems shown in Fig. 3.2. The reliability expression for system (a) is

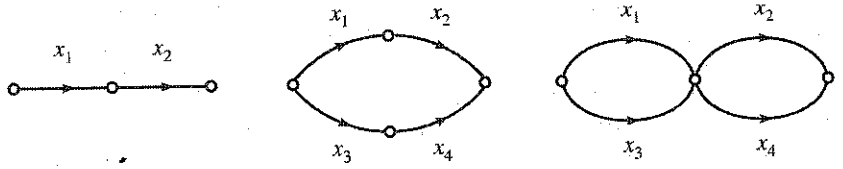


Figure 3.2 Comparison of three different systems: (a) single system, (b) unit redundancy, and (c) component redundancy.

$$R_a(p) = P(x_1)P(x_2) = p^2 \tag{3.6}$$

where both x_1 and x_2 are independent and identical and $P(x_1) = P(x_2) = p$. The reliability expression for system (b) is given simply by

$$R_b(p) = P(x_1x_2 + x_3x_4) \tag{3.7a}$$

For independent identical units (IIU) with reliability of p,

$$R_b(p) = 2R_a - R_a^2 = p^2(2 - p^2) \tag{3.7b}$$

In the case of system (c), one can combine each component pair in parallel to obtain

$$R_b(p) = P(x_1 + x_3)P(x_2 + x_4) \tag{3.8a}$$

Assuming IIU, we obtain

$$R_c(p) = p^2(2 - p)^2 \tag{3.8b}$$

To compare Eqs. (3.8b) and (3.7b), we use the ratio

$$\frac{R_c(p)}{R_b(p)} = \frac{p^2(2-p)^2}{p^2(2-p^2)} = \frac{(2-p)^2}{(2-p^2)} \tag{3.9}$$

Algebraic manipulation yields

$$\frac{R_c(p)}{R_b(p)} = \frac{(2-p)^2}{(2-p^2)} = \frac{4-4p+p^2}{2-p^2} = \frac{(2-p^2)+2(1-p)^2}{2-p^2} = 1 + \frac{2(1-p)^2}{2-p^2} \tag{3.10}$$

Because $0 < p < 1$, the term $2 - p^2 > 0$, and $R_c(p)/R_b(p) \geq 1$; thus component redundancy is superior to system redundancy for this structure. (Of course, they are equal at the extremes when $p = 0$ or $p = 1$.)

We can extend these chain structures into an n-element series structure, two parallel n-element system-redundant structures, and a series of n structures of two parallel elements. In this case, Eq. (3.9) becomes

$$\frac{R_c(p)}{R_b(p)} = \frac{(2-p)^n}{(2-p^n)} \tag{3.11}$$

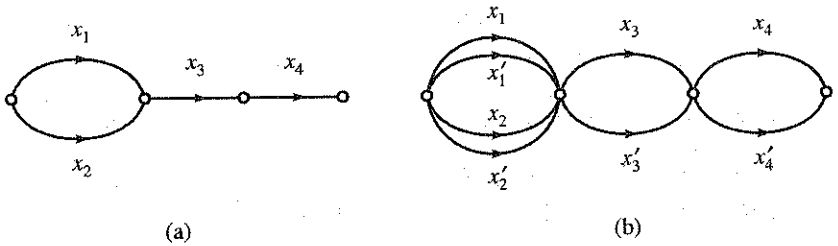
Roberts [1964, p. 260] proves by induction that this ratio is always greater than 1 and that component redundancy is superior regardless of the number of elements n.

The superiority of component redundancy over system redundancy also holds true for nonidentical elements; an algebraic proof is given in Shooman [1990, p. 282].

A simpler proof of the foregoing principle can be formulated by considering the system tie-sets. Clearly, in Fig. 3.2(b), the tie-sets are x_1x_2 and x_3x_4, whereas in Fig. 3.2(c), the tie-sets are x_1x_2, x_3x_4, x_1x_4, and x_3x_2. Since the system reliability is the probability of the union of the tie-sets, and since system (c) has the same two tie-sets as system (b) as well as two additional ones, the component redundancy configuration has a larger reliability than the unit redundancy configuration. It is easy to see that this tie-set proof can be extended to the general case.

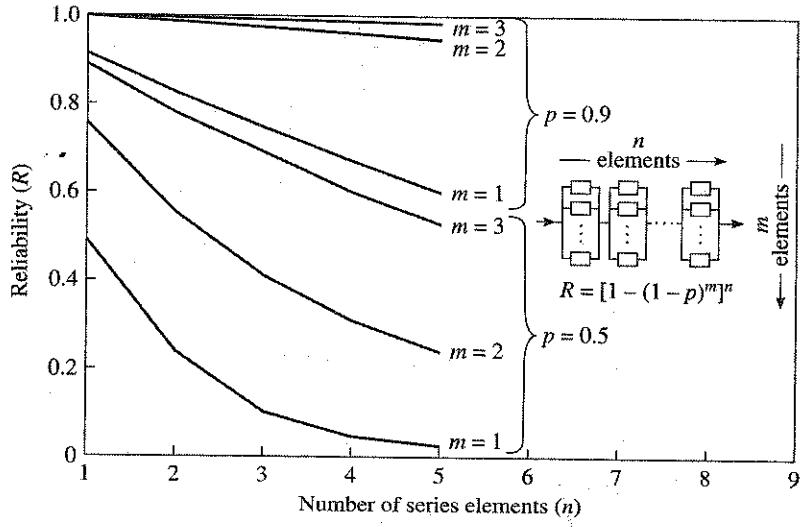
The specific result can be broadened to include a large number of structures. As an example, consider the system of Fig. 3.3(a) that can be viewed as a simple series structure if the parallel combination of x_1 and x_2 is replaced by an equivalent branch that we will call x_5. Then x_5, x_3, and x_4 form a simple chain structure, and component redundancy, as shown in Fig. 3.3(b), is clearly superior. Many complex configurations can be examined in a similar manner. Unit and component redundancy are compared graphically in Fig. 3.4.

Another interesting case in which one can compare component and unit

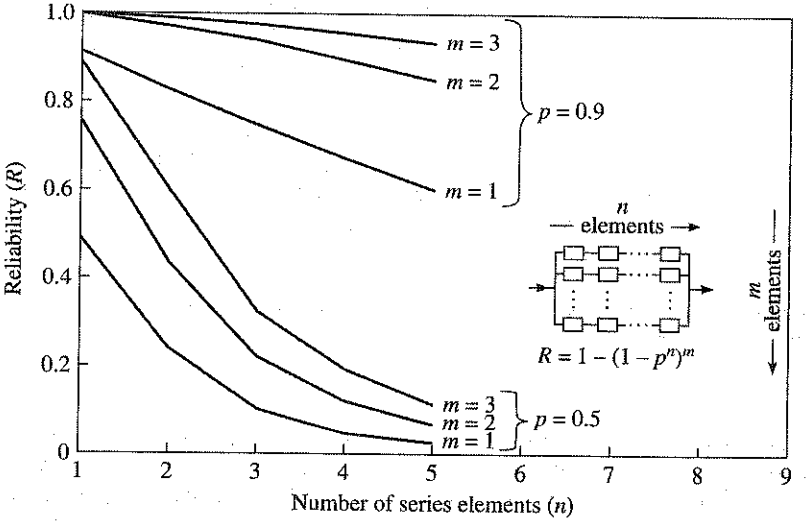


(a) (b)

Figure 3.3 Component redundancy: (a) original system and (b) redundant system.

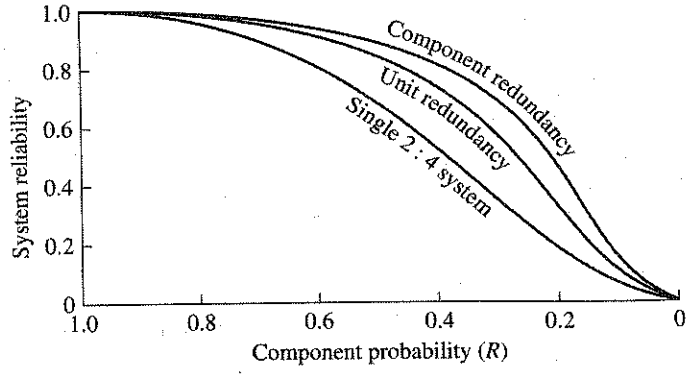


(a)

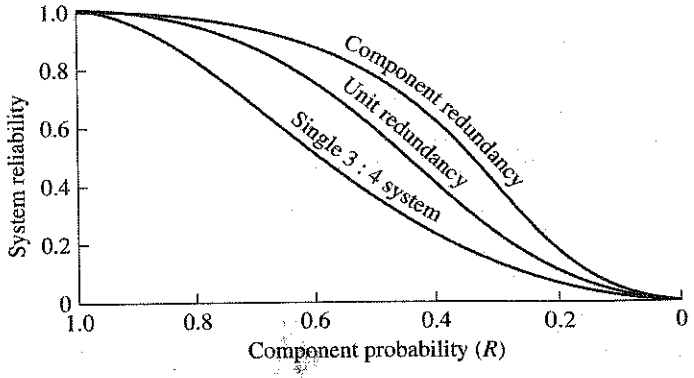


(b)

Figure 3.4 Redundancy comparison: (a) component redundancy and (b) unit redundancy. [Adapted from Figs. 7.10 and 7.11, *Reliability Engineering*, ARINC Research Corporation, used with permission, Prentice-Hall, Englewood Cliffs, NJ, 1964.]



(a)

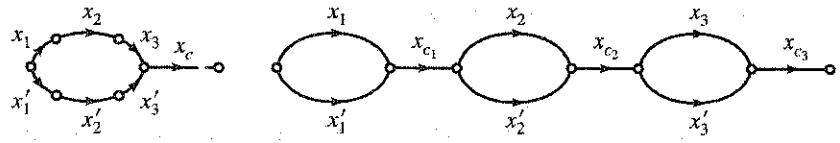


(b)

Figure 3.5 Comparison of component and unit redundancy for r-out-of-n systems: (a) a 2-out-of-4 system and (b) a 3-out-of-4 system.

redundancy is in an r-out-of-n system (the system succeeds if r-out-of-n components succeed). Immediately, one can see that for $r = n$, the structure is a series system, and the previous result applies. If $r = 1$, the structure reduces to n parallel elements, and component and unit redundancy are identical. The interesting cases are then $2 \le r < n$. The results for 2-out-of-4 and 3-out-of-4 systems are plotted in Fig. 3.5. Again, component redundancy is superior. The superiority of component over unit redundancy in an r-out-of-n system is easily proven by considering the system tie-sets.

All the above analysis applies to two-state systems. Different results are obtained for multistate models; see Shooman [1990, p. 286].



(a) System redundancy (one coupler)

(b) Component redundancy (three couplers)

Figure 3.6 Comparison of system and component redundancy, including coupling.

In a practical case, implementing redundancy is a bit more complex than indicated in the reliability graphs used in the preceding analyses. A simple example illustrates the issues involved. We all know that public address systems consisting of microphones, connectors and cables, amplifiers, and speakers are notoriously unreliable. Using our principle that component redundancy is better, we should have two microphones that are connected to a switching box, and we should have two connecting cables from the switching box to dual inputs to amplifier 1 or 2 that can be selected from a front panel switch, and we select one of two speakers, each with dual wires from each of the amplifiers. We now have added the reliability of the switches in *series* with the parallel components, which lowers the reliability a bit; however, the net result should be a gain. Suppose we carry component redundancy to the extreme by trying to parallel the resistors, capacitors, and transistors in the amplifier. In most cases, it is far from simple to merely parallel the components. Thus how low a level of redundancy is feasible is a decision that must be left to the system designer.

We can study the required circuitry needed to allow redundancy; we will call such circuitry or components *couplers*. Assume, for example, that we have a system composed of three components and wish to include the effects of coupling in studying system versus component reliability by using the model shown in Fig. 3.6. (Note that the prime notation is used to represent a "companion" element, not a logical complement.) For the model in Fig. 3.6(a), the reliability expression becomes

$$R_a = P(x_1x_2x_3 + x_1'x_2'x_3')P(x_c) \tag{3.12}$$

and if we have IIU and $P(x_c) = Kp(x_c) = Kp$,

$$R_a = (2p^3 - p^6)Kp \tag{3.13}$$

Similarly, for Fig. 3.6(b) we have

$$R_b = P(x_1 + x_1')P(x_2 + x_2')P(x_3 + x_3')P(x_{c1})P(x_{c2})P(x_{c3}) \tag{3.14}$$

and if we have IIU and $P(x_{c1}) = P(x_{c2}) = P(x_{c3}) = Kp$,

$$R_b = (2p - p^2)^3 k^3 p^3 \tag{3.15}$$

We now wish to explore for what value of K Eqs. (3.13) and (3.15) are equal:

$$(2p^3 - p^6)Kp = (2p - p^2)^3 K^3 p^3 \tag{3.16a}$$

Solving for K yields

$$K^2 = \frac{(2p^3 - p^6)}{(2p - p^2)^3 p^2} \tag{3.16b}$$

If $p = 0.9$, substitution in Eq. (3.16) yields $K = 1.085778501$, and the coupling reliability Kp becomes 0.9772006509. The easiest way to interpret this result is to say that if the component failure probability $1 - p$ is 0.1, then component and system reliability are equal if the coupler failure probability is 0.0228. In other words, if the coupler failure probability is less than 22.8% of the component failure probability, component redundancy is superior. Clearly, coupler reliability will probably be significant in practical situations.

Most reliability models deal with two element states—good and bad; however, in some cases, there are more distinct states. The classical case is a diode, which has three states: good, failed-open, and failed-shorted. There are also analogous elements, such as leaking and blocked hydraulic lines. (One could contemplate even more than three states; for example, in the case of a diode, the two "hard"-failure states could be augmented by an "intermittent" short-failure state.) For a treatment of redundancy for such three-state elements, see Shooman [1990, p. 286].

3.4 APPROXIMATE RELIABILITY FUNCTIONS

Most system reliability expressions simplify to sums and differences of various exponential functions once the expressions for the hazard functions are substituted. Such functions may be hard to interpret; often a simple computer program and a graph are needed for interpretation. Notwithstanding the case of computer computations, it is still often advantageous to have techniques that yield approximate analytical expressions.

3.4.1 Exponential Expansions

A general and very useful approximation technique commonly used in many branches of engineering is the *truncated series expansion*. In reliability work, terms of the form e^{-z} occur time and again; the expressions can be simplified by

series expansion of the exponential function. The Maclaurin series expansion of e^{-z} about $Z = 0$ can be written as follows:

$$e^{-Z} = 1 - Z + \frac{Z^2}{2!} - \frac{Z^3}{3!} + \cdots + \frac{(-Z)^n}{n!} + \cdots \tag{3.17}$$

We can also write the series in n terms and a remainder term [Thomas, 1965, p. 791], which accounts for all the terms after $(-Z)^n/n!$

$$e^{-Z} = 1 - Z + \frac{Z^2}{2!} - \frac{Z^3}{3!} + \cdots + \frac{(-Z)^n}{n!} + R_n(Z) \tag{3.18}$$

where

$$R_n(Z) = (-1)^{n+1} \int_0^Z \frac{(Z - \xi)^n}{n!} e^{-\xi} \, d\xi \tag{3.19}$$

We can therefore approximate e^{-Z} by n terms of the series and use $R_n(Z)$ to approximate the remainder. In general, we use only two or three terms of the series, since in the high-reliability region $e^{-Z} \sim 1$, Z is small, and the high-order terms Z^n in the series expansion becomes insignificant. For example, the reliability of two parallel elements is given by

$$\begin{aligned}(2e^{-Z}) + (-e^{-2Z}) &= \left(2 - 2Z + \frac{2Z^2}{2!} - \frac{2Z^3}{3!} + \cdots + \frac{2(-Z)^n}{n!} + \cdots\right) \\ &\quad + \left(-1 + 2Z - \frac{(2Z)^2}{2!} + \frac{(2Z)^3}{3!} - \cdots - \frac{(2Z)^n}{n!} + \cdots\right) \\ &= 1 - Z^2 + Z^3 - \frac{7}{12} Z^4 + \frac{1}{4} Z^5 - \cdots + \end{aligned} \tag{3.20}$$

Two- and three-term approximations to Eqs. (3.17) and (3.20) are compared with the complete expressions in Fig. 3.7(a) and (b). Note that the two-term approximation is a "pessimistic" one, whereas the three-term expression is slightly "optimistic"; inclusion of additional terms will give a sequence of alternate upper and lower bounds. In Shooman [1990, p. 217], it is shown that the magnitude of the nth term is an upper bound on the error term, $R_n(Z)$, in an n-term approximation.

If the system being modeled involves repair, generally a Markov model is used, and oftentimes Laplace transforms are used to solve the Markov equations. In Section B8.3, a simplified technique for finding the series expansion of a reliability function—cf. Eq. (3.20)—directly from a Laplace transform is discussed.

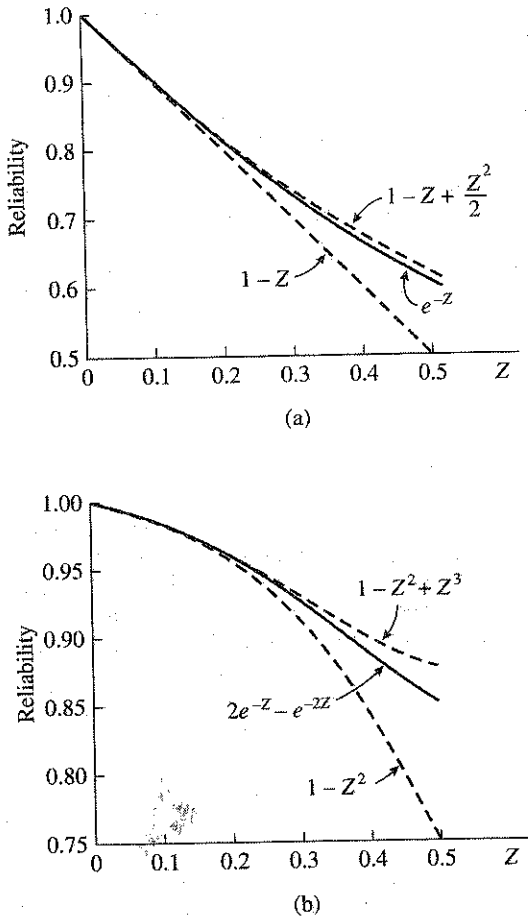


Figure 3.7 Comparison of exact and approximate reliability functions: (a) single unit and (b) two parallel units.

3.4.2 System Hazard Function

Sometimes it is useful to compute and study the system hazard function (failure rate). For example, suppose that a system consists of two series elements, x_2x_3, in parallel with a third, x_1. Thus, the system has two "success paths": it succeeds if x_1 works or if x_2 and x_3 both work. If all elements have identical constant hazards, λ, the reliability function is given by

$$R(t) = P(x_1 + x_2x_3) = e^{-\lambda t} + e^{-2\lambda t} - e^{-3\lambda t} \tag{3.21}$$

From Appendix B, we see that $z(t)$ is given by the density function divided by the reliability function, which can be written as the negative of the time derivative of the reliability function divided by the reliability function.

$$z(t) = \frac{f(t)}{R(t)} = -\frac{\dot{R}(t)}{R(t)} = \frac{\lambda(1 + 2e^{-\lambda t} - 3e^{-2\lambda t})}{1 + e^{-\lambda t} - e^{-2\lambda t}} \tag{3.22}$$

Expanding $z(t)$ in a Taylor series,

$$z(t) = 1 + \lambda t - 3\lambda^2 t^2/2 + \cdots \tag{3.23}$$

We can use such approximations to compare the equivalent hazard of various systems.

3.4.3 Mean Time to Failure

In the last section, it was shown that reliabiilty calculations become very complicated in a large system when there are many components and a diverse reliability structure. Not only was the reliability expression difficult to write down in such a case, but computation was lengthy, and interpretation of the individual component contributions was not easy. One method of simplifying the situation is to ask for less detailed information about the system. A useful figure of merit for a system is the mean time to failure (MTTF).

As was derived in Eq. (B51) of Appendix B, the MTTF is the expected value of the time to failure. The standard formula for the expected value involves the integral of $tf(t)$; however, this can be expressed in terms of the reliability function.

$$\text{MTTF} = \int_0^\infty R(t)\, dt \tag{3.24}$$

We can use this expression to compute the MTTF for various configurations. For a series reliability configuration of n elements in which each of the elements has a failure rate $z_i(t)$ and $Z(t) = \int z(t)\, dt$, one can write the reliability expression as

$$R(t) = \exp\left[-\sum_{i=1}^{n} Z_i(t)\right] \tag{3.25a}$$

and the MTTF is given by

$$\text{MTTF} = \int_0^\infty \left\{\exp\left[-\sum_{i=1}^{n} Z_i(t)\right]\right\} dt \tag{3.25b}$$

If the series system has components with more than one type of hazard model, the integral in Eq. (3.25b) is difficult to evaluate in closed form but can always be done using a series approximation for the exponential integrand; see Shooman [1990, p. 20].

Different equations hold for a parallel system. For two parallel elements, the reliability expression is written as $R(t) = e^{-Z_1(t)} + e^{-Z_2(t)} - e^{[-Z_1(t)+Z_2(t)]}$. If both system components have a constant-hazard rate, and we apply Eq. (3.24) to each term in the reliability expression,

$$\text{MTTF} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_1 + \lambda_2} \tag{3.26}$$

In the general case of n parallel elements with constant-hazard rate, the expression becomes

$$\begin{aligned}\text{MTTF} = &\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \cdots + \frac{1}{\lambda_n}\right) - \left(\frac{1}{\lambda_1+\lambda_2} + \frac{1}{\lambda_1+\lambda_3} + \cdots + \frac{1}{\lambda_i+\lambda_j}\right) \\ &+ \left(\frac{1}{\lambda_1+\lambda_2+\lambda_3} + \frac{1}{\lambda_1+\lambda_2+\lambda_4} + \cdots + \frac{1}{\lambda_i+\lambda_j+\lambda_k}\right) \\ &- \cdots + (-1)^{n+1}\frac{1}{\sum_{i=1}^{n}\lambda_i}\end{aligned} \tag{3.27}$$

If the n units are identical—that is, $\lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda$—then Eq. (3.27) becomes

$$\text{MTTF} = \left[\frac{\binom{n}{1}}{1} - \frac{\binom{n}{2}}{2} + \frac{\binom{n}{3}}{3} - \cdots + (-1)^{n+1}\frac{\binom{n}{n}}{n}\right] = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i} \tag{3.28a}$$

The preceding series is called the harmonic series; the summation form is given in Jolley [1961, p. 26, Eq. (200)] or Courant [1951, pp. 380]. This series occurs in number theory, and a series expansion is attributed to the famous mathematician Euler; the constant in the expansion (0.577) is called Euler's constant [Jolley, 1961, p. 14, Eq. (70)].

$$\frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i} = \frac{1}{\lambda}\left[0.577 + \ln n + \frac{1}{2n} - \frac{1}{12n(n+1)}\cdots\right] \tag{3.28b}$$

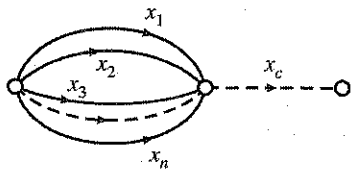


Figure 3.8 Parallel reliability configuration of n elements and a coupling device x_c.

3.5 PARALLEL REDUNDANCY

3.5.1 Independent Failures

One classical approach to improving reliability is to provide a number of elements in the system, any one of which can perform the necessary function. If a system of n elements can function properly when only one of the elements is good, a parallel configuration is indicated. (A parallel configuration of n items is shown in Fig. 3.8.) The reliability expression for a parallel system may be expressed in terms of the probability of success of each component or, more conveniently, in terms of the probability of failure (coupling devices ignored).

$$R(t) = P(x_1 + x_2 + \cdots + x_n) = 1 - P(\bar{x}_1\bar{x}_2\cdots\bar{x}_n) \tag{3.29}$$

In the case of constant-hazard components, $P_f = P(\bar{x}_i) = 1 - e^{-\lambda_i t}$, and Eq. (3.29) becomes

$$R(t) = 1 - \left[\prod_{i-1}^{n}(1 - e^{-\lambda_i t})\right] \tag{3.30}$$

In the case of linearly increasing hazard, the expression becomes

$$R(t) = 1 - \left[\prod_{i-1}^{n}(1 - e^{-K_i t^2/2})\right] \tag{3.31}$$

We recall that in the example of Fig. 3.6(a), we introduced the notion that a coupling device is needed. Thus, in the general case, the system reliability function is

$$R(t) = \left\{1 - \left[\prod_{i-1}^{n}(1 - e^{-Z_i(t)})\right]\right\}P(x_c) \tag{3.32}$$

If we have IIU with constant-failure rates, then Eq. (3.32) becomes

$$R(t) = [1 - (1 - e^{-\lambda t})^n]e^{-\lambda_c t} \tag{3.33a}$$

where λ is the element failure rate and λ_c is the coupler failure rate. Assuming $\lambda_c t < \lambda t \ll 1$, we can simplify Eq. (3.33) by approximating $e^{-\lambda_c t}$ and $e^{-\lambda t}$ by the first two terms in the expansion—cf. Eq. (3.17)—yielding $(1 - e^{-\lambda t}) \approx \lambda t$, $e^{-\lambda_c t} \approx 1 - \lambda_c t$. Substituting these approximations into Eq. (3.33a),

$$R(t) \approx [1 - (\lambda t)^n](1 - \lambda_c t) \tag{3.33b}$$

Neglecting the last term in Eq. (3.33b), we have

$$R(t) \approx 1 - \lambda_c t - (\lambda t)^n \tag{3.34}$$

Clearly, the coupling term in Eq. (3.34) must be small or it becomes the dominant portion of the probability of failure. We can obtain an "upper limit" for λ_c if we equate the second and third terms in Eq. (3.34) (the probabilities of coupler failure and parallel system failure) yielding

$$\frac{\lambda_c}{\lambda} < (\lambda t)^{n-1} \tag{3.35}$$

For the case of $n = 3$ and a comparison at $\lambda t = 0.1$, we see that $\lambda_c/\lambda < 0.01$. Thus the failure rate of the coupling device must be less than 1/100 that of the element. In this example, if $\lambda_c = 0.01\lambda$, then the coupling system probability of failure is equal to the parallel system probability of failure. This is a limiting factor in the application of parallel reliability and is, unfortunately, sometimes neglected in design and analysis. In many practical cases, the reliability of the several elements in parallel is so close to unity that the reliability of the coupling element dominates.

If we examine Eq. (3.34) and assume that $\lambda_c \approx 0$, we see that the number of parallel elements n affects the curvature of $R(t)$ versus t. In general, the more parallelism in a reliability block diagram, the less the initial slope of the reliability curve. The converse is true with more series elements. As an example, compare the reliability functions for the three reliability graphs in Fig. 3.9 that are plotted in Fig. 3.10.

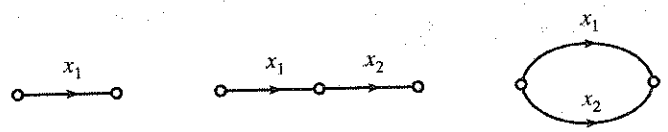
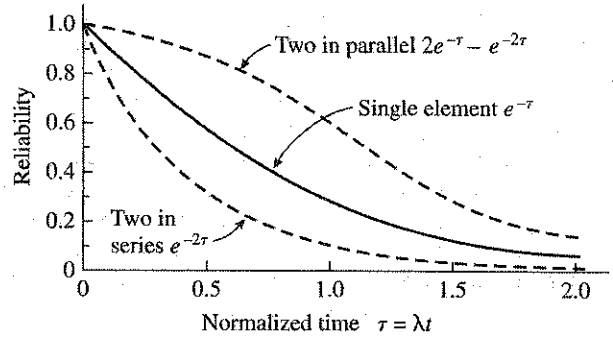
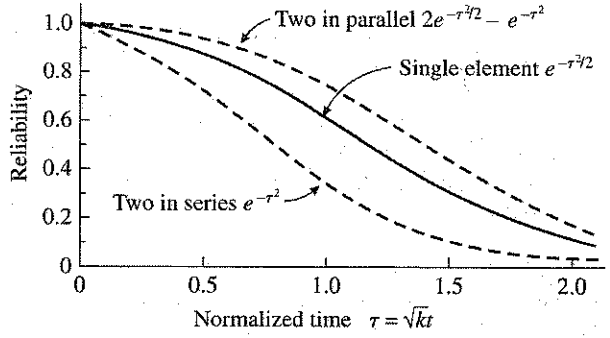


Figure 3.9 Three reliability structures: (a) single element, (b) two series elements, and (c) two parallel elements.



(a)



(b)

Figure 3.10 Comparison of reliability functions: (a) constant-hazard elements and (b) linearly increasing hazard elements.

3.5.2 Dependent and Common Mode Effects

There are two additional effects that must be discussed in analyzing a parallel system: that of *common mode* (common cause) failures and that of *dependent failures*. A common mode failure is one that affects all the elements in a redundant system. The term was popularized when the first reliability and risk analyses of nuclear reactors were performed in the 1970s [McCormick, 1981, Chapter 12]. To protect against core melt, reactors have two emergency core-cooling systems. One important failure scenario—that of an earthquake—is likely to rupture the piping on both cooling systems.

Another example of common mode activity occurred early in the space program. During the reentry of a *Gemini* spacecraft, one of the two guidance computers failed, and a few minutes later the second computer failed. Fortunately,

the astronauts had an additional backup procedure. Based on rehearsed procedures and precomputations, the Ground Control advised the astronauts to maneuver the spacecraft, to align the horizon with one of a set of horizontal scribe marks on the windows, and to rotate the spacecraft so that the Sun was aligned with one set of vertical scribe marks. The Ground Control then gave the astronauts a countdown to retro-rocket ignition and a second countdown to rocket cutoff. The spacecraft splashed into the ocean—closer to the recovery ship than in any previous computer-controlled reentry. Subsequent analysis showed that the temperature inside the two computers was much higher than expected and that the diodes in the separate power supply of each computer had burned out. From this example, we learn several lessons:

1. The designers provided two computers for redundancy.
2. Correctly, two separate power supplies were provided, one for each computer, to avoid a common power-supply failure mode.
3. An unexpectedly high ambient temperature caused identical failues in the diodes, resulting in a common mode failure.
4. Fortunately, there was a third redundant mode that depended on a completely different mechanism, the scribe marks, and visual alignment. When parallel elements are purposely chosen to involve devices with different failure mechanisms to avoid common mode failures, the term *diversity* is used.

In terms of analysis, common mode failures behave much like failures of a coupling mechanism that was studied previously. In fact, we can use Eq. (3.33) to analyze the effect if we use λ_c to represent the sum of coupling and common mode failure rates. (A fortuitous choice of subscript!)

Another effect to consider in parallel systems is the effect of dependent failures. Suppose we wish to use two parallel satellite channels for reliable communication, and the probability of each channel failure is 0.01. For a single channel, the reliability would be 0.99; for two parallel channels, c_1 and c_2, we would have

$$R = P(c_1 + c_2) = 1 - P(\bar{c}_1\bar{c}_2) \tag{3.36}$$

Expanding the last term in Eq. (3.36) yields

$$R = 1 - P(\bar{c}_1\bar{c}_2) = 1 - P(\bar{c}_1)P(\bar{c}_2|\bar{c}_1) \tag{3.37}$$

If the failures of both channels, c_1 and c_2, are independent, Eq. (3.37) yields $R = 1 - 0.01 \times 0.01 = 0.9999$. However, suppose that one-quarter of satellite transmission failures are due to atmospheric interference that would affect both channels. In this case, $P(\bar{c}_2|\bar{c}_1)$ is 0.25, and Eq. (3.37) yields $R = 1 - 0.01 \times 0.25 = 0.9975$. Thus for a single channel, the probability of failure is 0.01; with two independent parallel channels, it is 0.0001, but for dependent channels, it is 0.0025. This means that dependency has reduced the expected 100-fold reduction in failure probabilities to a reduction by only a factor of 4. In general, a modeling of dependent failures requires some knowledge of the failure mechanisms that result in dependent modes.

The above analysis has explored many factors that must be considered in analyzing parallel systems: coupling failures, common mode failures, and dependent failures. Clearly, only simple models were used in each case. More complex models may be formulated by using Markov process models—to be discussed in Section 3.7, where we analyze standby redundancy.

3.6 AN r-OUT-OF-n STRUCTURE

Another simple structure that serves as a useful model for many reliability problems is an r-out-of-n structure. Such a model represents a system of n components in which r of the n items must be good for the system to succeed. (Of course, r is less than n.) An example of an r-out-of-n structure is a fiber-optic cable, which has a capacity of n circuits. If the application requires r channels of the transmission, this is an r-out-of-n system $(r:n)$. If the capacity of the cable n exceeds r by a significant amount, this represents a form of parallel redundancy. We are of course assuming that if a circuit fails it can be switched to one of the n–r "extra circuits."

We may formulate a structural model for an r-out-of-n system, but it is simpler to use the binomial distribution if applicable. The binomial distribution can be used only when the n components are independent and identical. If the components differ or are dependent, the structural-model approach must be used. Success of exactly r-out-of-n identical and independent items is given by

$$B(r:n) = \binom{n}{r} p^r (1-p)^{n-r} \tag{3.38}$$

where $r:n$ stands for r out of n, and the success of at least r-out-of-n items is given by

$$P_s = \sum_{k=r}^{n} B(k:n) \tag{3.39}$$

For constant-hazard components, Eq. (3.38) becomes

$$R(t) = \sum_{k=r}^{n} \binom{n}{k} e^{-k\lambda t}(1 - e^{-\lambda t})^{n-k} \tag{3.40}$$

Similarly, for linearly increasing or Weibull components, the reliability functions are

$$R(t) = \sum_{k=r}^{n} \binom{n}{k} e^{-kKt^2/2}(1 - e^{-Kt^2/2})^{n-k} \tag{3.41a}$$

and

$$R(t) = \sum_{k=r}^{n} \binom{n}{k} e^{-kKt^{m+1}/(m+1)}(1 - e^{-Kt^{m+1}/(m+1)})^{n-k} \tag{3.41b}$$

Clearly, Eqs. (3.39)–(3.41) can be studied and evaluated by a parametric computer study. In many cases, it is useful to approximate the result, although numerical evaluation via a computer program is not difficult. For an r-out-of-n structure of identical components, the exact reliability expression is given by Eq. (3.38). As is well known, we can approximate the binomial distribution by the Poisson or normal distributions, depending on the values of n and p (see Shooman, 1990, Sections 2.5.6 and 2.6.8). Interestingly, we can also develop similar approximations for the case in which the n parameters are not identical.

The Poisson approximation to the binomial holds for $p \le 0.05$ and $n \ge 20$, which represents the low-reliability region. If we are interested in the high-reliability region, we switch to failure probabilities, requiring $q = 1 - p \le 0.05$ and $n \ge 20$. Since we are assuming different components, we define average probabilities of success and failure $\overline{p}$ and $\overline{q}$ as

$$\overline{p} = \frac{1}{n}\sum_{i=1}^{n} p_i \qquad 1 - \overline{q} = 1 - \frac{1}{n}\sum_{i=1}^{n}(1 - p_i) \tag{3.42}$$

Thus, for the high-reliability region, we compute the probability of n–r or fewer failures as

$$R(t) = \sum_{k=0}^{n-r} \frac{(n\overline{q})^k e^{-n\overline{q}}}{k!} \tag{3.43}$$

and for the low-reliability region, we compute the probability of r or more successes as

$$R(t) = \sum_{k=r}^{n} \frac{(n\overline{p})^k e^{-n\overline{p}}}{k!} \tag{3.44}$$

Equations (3.43) and (3.44) avoid a great deal of algebra in dealing with nonidentical r-out-of-n components. The question of accuracy is somewhat difficult to answer since it depends on the system structure and the range of values of p that make up $\overline{p}$. For example, if the values of q vary only over a 2 : 1 range, and if $\overline{q} \le 0.05$ and $n \ge 20$, intuition tells us that we should obtain reasonably accurate results. Clearly, modern computer power makes explicit enumeration of Eqs. (3.39)–(3.41) a simple procedure, and Eqs. (3.43) and (3.44) are useful mainly as simplified analytical expressions that provide a check on computations. [Note that Eqs. (3.43) and (3.44) also hold true for IIU with $\overline{p} = p$ and $\overline{q} = q$.]

We can appreciate the power of an $r : n$ design by considering the following example. Suppose we have a fiber-optic cable with 20 channels (strands) and a system that requires all 20 channels for success. (For simplicity of the discussion, assume that the associated electronics will not fail.) Suppose the probability of failure of each channel within the cable is $q = 0.0005$ and $p = 0.9995$. Since all 20 channels are needed for success, the reliability of a 20-channel cable will be $R_{20} = (0.9995)^{20} = 0.990047$. Another option is to use two parallel 20-channel cables and associated electronics switch from cable A to cable B whenever there is any failure in cable A. The reliability of such an ordinary parallel system of two 20-channel cables is given by $R_{2/20} = 2(0.990047) - (0.990047)^2 = 0.9999009$. Another design option is to include extra channels in the single cable beyond the 20 that are needed—in such a case, we have an $r : n$ system. Suppose we approach the design in a trial-and-error fashion. We begin by trying $n = 21$ channels, in which case we have

$$\begin{aligned} R_{21} &= B(21:21) + B(20:21) = p^{21}q^0 + 21p^{20}q \\ &= (0.9995)^{21} + 21(0.9995)^{20}(0.0005) = 0.98755223 + 0.010395497 \\ &= 0.999947831 \end{aligned} \tag{3.45}$$

Thus R_{21} exceeds the design with two 20-channel cables. Clearly, all the designs require some electronic steering (couplers) for the choice of channels, and the coupler reliability should be included in a detailed comparison. Of course, one should worry about common mode failures, which could completely change the foregoing results. Construction damage—that is, line-severing by a contractor's excavating maching (backhoe)—is a significant failure mode for in-soil fiber-optic telephone lines.

As a check on Eq. (3.45), we compute the approximation Eq. (3.43) for $n = 21$, $r = 20$.

$$\begin{aligned} R(t) &= \sum_{k=0}^{1} \frac{(n\overline{q})^k e^{-n\overline{q}}}{k!} = (1 + nq)e^{-nq} = [1 + 21(0.0005)]e^{-22\times 0.0005} \\ &= 0.999831687 \end{aligned} \tag{3.46}$$

These values are summarized in Table 3.1.

TABLE 3.1 Comparison of Design for Fiber-Optic Cable Example

System	Reliability, R	Unreliability, $(1-R)$
Single 20-channel cable	0.990047	0.00995
Two 20-channel cables in parallel	0.9999009	0.000099
A 21-channel cable (exact)	0.999948	0.000052
A 21-channel cable (approx.)	0.99983	0.00017

Essentially, the efficiency of the $r:n$ system is because the redundancy is applied at a lower level. In practice, a 24- or 25-channel cable would probably be used, since a large portion of the cable cost would arise from the land used and the laying of the cable. Therefore, the increased cost of including four or five extra channels would be "money well spent," since several channels could fail and be locked out before the cable failed. If we were discussing the number of channels in a satellite communications system, the major cost would be the launch; the economics of including a few extra channels would be similar.

3.7 STANDBY SYSTEMS

3.7.1 Introduction

Suppose we consider two components, x_1 and x_1', in parallel. For discussion purposes, we can think of x_1 as the primary system and x_1' as the backup; however, the systems are identical and could be interchanged. In an ordinary parallel system, both x_1 and x_1' begin operation at time $t=0$, and both can fail. If t_1 is the time to failure of x_1, and t_2 is the time to failure of x_2, then the time to system failure is the maximum value of (t_1, t_2). An improvement would be to energize the primary system x_1 and have backup system x_1' unenergized so that it cannot fail. Assume that we can immediately detect the failure of x_1 and can energize x_1' so that it becomes the active element. Such a configuration is called a standby system, x_1 is called the *on-line* system, and x_1' the *standby* system. Sometimes an ordinary parallel system is called a "hot" standby, and a standby system is called a "cold" standby. The time to system failure for a standby system is given by $t = t_1 + t_2$. Clearly, $t_1 + t_2 > \max(t_1, t_2)$, and a standby system is superior to a parallel system. The "coupler" element in a standby system is more complex than in a parallel system, requiring a more detailed analysis.

One can take a number of different approaches to deriving the equations for a standby system. One is to determine the probability distribution of $t = t_1 + t_2$, given the distributions of t_1 and t_2 [Papoulis, 1965, pp. 193–194]. Another approach is to develop a more general system of probability equations known

TABLE 3.2 States for a Parallel System

$s_0 = x_1 x_2$ = Both components good.
$s_1 = x_1 \bar{x}_2 = x_1$, good; x_2, failed.
$s_2 = \bar{x}_1 x_2 = x_1$, failed; x_2, good.
$s_3 = \bar{x}_1 \bar{x}_2$ = Both components failed.

as Markov models. This approach is developed in Appendix B and will be used later in this chapter to describe repairable systems.

In the next section, we take a slightly simpler approach: we develop two difference equations, solve them, and by means of a limiting process develop the needed probabilities. In reality, we are developing a simplified Markov model without going through some of the formalism.

3.7.2 Success Probabilities for a Standby System

One can characterize an ordinary parallel system with components x_1 and x_2 by the four states given in Table 3.2. If we assume that the standby component in a standby system won't fail until energized, then the three states given in Table 3.3 describe the system. The probability that element x fails in time interval Δt is given by the product of the failure rate λ (failures per hour) and Δt. Similarly, the probability of no failure in this interval is $(1 - \lambda \Delta t)$. We can summarize this information by the probabilistic state model (probabilistic graph, Markov model) shown in Fig. 3.11.

The probability that the system makes a transition from state s_0 to state s_1 in time Δt is given by $\lambda_1 \Delta t$, and the transition probability for staying in state s_0 is $(1 - \lambda_1 \Delta t)$. Similar expressions are shown in the figure for staying in state s_1 or making a transition to state s_2. The probabilities of being in the various system states at time $t = t + \Delta t$ are governed by the following difference equations:

$$P_{s_0}(t + \Delta t) = (1 - \lambda_1 \Delta t) P_{s_0}(t), \tag{3.47a}$$

$$P_{s_1}(t + \Delta t) = \lambda_1 \Delta t P_{s_0}(t) + (1 - \lambda_2 \Delta t) P_{s_1}(t) \tag{3.47b}$$

$$P_{s_2}(t + \Delta t) = \lambda_2 \Delta t P_{s_1}(t) + (1) P_{s_2}(t) \tag{3.47c}$$

We can rewrite Eq. (3.47) as

TABLE 3.3 States for a Standby System

$s_0 = x_1 x_2$ = On-line and standby components good.
$s_1 = \bar{x}_1 x_2$ = On-line failed and standby component good.
$s_2 = \bar{x}_1 \bar{x}_2$ = On-line and standby components failed.

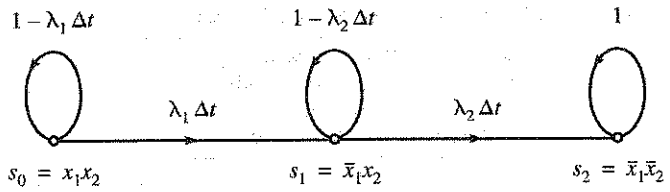


Figure 3.11 A probabilistic state model for a standby system.

$$P_{s_0}(t+\Delta t) - P_{s_0}(t) = -\lambda_1 \Delta t P_{s_0}(t) \tag{3.48a}$$

$$\frac{P_{s_0}(t+\Delta t) - P_{s_0}(t)}{\Delta t} = -\lambda_1 P_{s_0}(t) \tag{3.48b}$$

Taking the limit of the left-hand side of Eq. (3.48b) as $\Delta t \to 0$ yields the time derivative, and the equation becomes

$$\frac{dP_{s_0}(t)}{dt} + \lambda_1 P_{s_0} = 0 \tag{3.49}$$

This is a linear, first-order, homogeneous differential equation and is known to have the solution $P_{s_0} = Ae^{-\lambda_1 t}$. To verify that this is a solution, we substitute into Eq. (3.49) and obtain

$$-\lambda_1 A e^{-\lambda_1 t} + \lambda_1 A e^{-\lambda_1 t} = 0$$

The value of A is determined from the initial condition. If we start with a good system, $P_{s_0}(t=0) = 1$; thus $A = 1$ and

$$P_{s_0} = e^{-\lambda_1 t} \tag{3.50}$$

In a similar manner, we can rewrite Eq. (3.47b) and take the limit obtaining

$$\frac{dP_{s_1}(t)}{dt} + \lambda P_{s_1}(t) = \lambda_1 P_{s_0} \tag{3.51}$$

This equation has the solution

$$P_{s_1}(t) = B_1 e^{-\lambda_1 t} + B_2 e^{-\lambda_2 t} \tag{3.52}$$

Substitution of Eq. (3.52) into Eq. (3.51) yields a group of exponential terms that reduces to

$$[\lambda_2 B_1 - \lambda_1 B_1 - \lambda_1] e^{-\lambda_1 t} = 0 \tag{3.53}$$

and solving for B_1 yields

$$B_1 = \frac{\lambda_1}{\lambda_2 - \lambda_1} \tag{3.54}$$

We can obtain the other constant by substituting the initial condition $P_{s_1}(t=0) = 0$, and solving for B_2 yields

$$B_2 = -B_1 = \frac{\lambda_1}{\lambda_1 - \lambda_2} \tag{3.55}$$

The complete solution is

$$P_{s_1}(t) = \frac{\lambda_1}{\lambda_2 - \lambda_1}[e^{-\lambda_1 t} - e^{-\lambda_2 t}] \tag{3.56}$$

Note that the system is successful if we are in state 0 or state 1 (state 2 is a failure). Thus the reliability is given by

$$R(t) = P_{s_0}(t) + P_{s_1}(t) \tag{3.57}$$

Equation (3.57) yields the reliability expression for a standby system where the on-line and the standby components have two different failure rates. In the more general case, both the on-line and standby components have the same failure rate, and we have a small difficulty since Eq. (3.56) becomes 0/0. The standard approach in such cases is to use l'Hospital's rule from calculus. The procedure is to take the derivative of the numerator and the denominator separately with respect to λ_2; then to take the limit as $\lambda_2 \to \lambda_1$. This results in the expression for the reliability of a standby system with two identical on-line and standby components:

$$R(t) = e^{-\lambda t} + \lambda t e^{-\lambda t} \tag{3.58}$$

A few general comments are appropriate at this point.

1. The solution given in Eq. (3.58) can be recognized as the first two terms in the Poisson distribution, the probability of zero occurrences in time t plus the probability of one occurrence in time t hours, where λ is the occurrence rate per hour. Since the "exposure time" for the standby component does not start until the on-line element has failed, the occurrences are a sequence in time that follows the Poisson distribution.
2. The model in Fig. 3.11 could have been extended to the right to incorporate a very large number of components and states. The general solution of such a model would have yielded the Poisson distribution.

3. A model could have been constructed composed of four states: (x_1x_2, $x_1\bar{x}_2$, $\bar{x}_1x_2$, $\bar{x}_1\bar{x}_2$). Solution of this model would yield the probability expressions for a parallel system. However, solution of a parallel system via a Markov model is seldom done except for tutorial purposes because the direct methods of Section 3.5 are simpler.
4. Generalization of a probabilistic graph, the resulting differential equations, the solution process, and the summing of appropriate probabilities leads to a generalized Markov model. This is further illustrated in the next section on repair.
5. In Section 3.8.2 and Chapter 4, we study the formulation of Markov models using a more general algorithm to derive the equations, and we use Laplace transforms to solve the equations.

3.7.3 Comparison of Parallel and Standby Systems

It is assumed that the reader has studied the material in Sections A8 and B6 that cover Markov models. We now compare the reliability of parallel and standby systems in this section. Standby systems are inherently superior to parallel systems; however, much of this superiority depends on the reliability of the standby switch. Also, the reliability of the coupler in a parallel system must also be considered in the comparison. The reliability of the standby system with an imperfect switch will require a more complex Markov model than that developed in the previous section, and such a model is discussed below.

The switch in a standby system must perform three functions:

1. It must have some sort of decision element or algorithm that is capable of sensing improper operation.
2. The switch must then remove the signal input from the on-line unit and apply it to the standby unit, and it must also switch the output as well.
3. If the element is an active one, the power must be transferred from the on-line to the standby element (see Fig. 3.12). In some cases, the input and output signals can be permanently connected to the two elements; only the power needs to be switched.

Often the decision unit and the input (and output) switch can be incorporated into one unit: either an analog circuit or a digital logic circuit or processor algorithm. Generally, the power switch would be some sort of relay or electronic switch, or it could be a mechanical device in the case of a mechanical, hydraulic, or pneumatic system. The specific implementation will vary with the application and the ingenuity of the designer.

The reliability expression for a two-element standby system with constant hazards and a perfect switch was given in Eqs. (3.50), (3.56), and (3.57) and for identical elements in Eq. (3.58). We now introduce the possibility that the switch is imperfect.

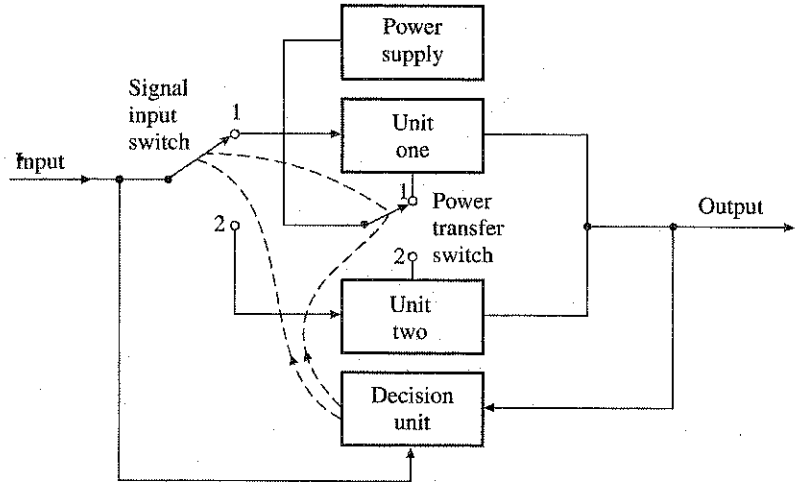


Figure 3.12 A standby system in which input and power switching are shown.

We begin with a simple model for the switch where we assume that any failure of the switch is a failure of the system, even in the case where both the on-line and the standby components are good. This is a conservative model that is easy to formulate. If we assume that the switch failures are independent of the on-line and standby component failures and that the switch has a constant failure rate λ_s, then Eq. (3.58) holds. Thus we obtain

$$R_1(t) = e^{-\lambda_s t}(e^{-\lambda t} + \lambda t e^{-\lambda t}) \tag{3.59}$$

Clearly, the switch reliability multiplies the reliability of the standby system and degrades the system reliability. We can evaluate how significant the switch reliability problem is by comparing it with an ordinary parallel system. A comparison of Eqs. (3.59) and (3.30) (for $n = 2$ and identical failure rates) is given in Fig. 3.13. Note that when the switch failure rate is only 10% of the component failure rates ($\lambda_s = 0.1\lambda$), the degradation is only minor, especially in the high-reliability region of most interest: ($1 \geq R(t) \geq 0.9$). The standby system degrades to about the same reliability as the parallel system when the switch failure rate is about half the component failure rate.

A simple way to improve the switch reliability model is to assume that the switch failure mode is such that it only fails to switch from on-line to standby when the on-line element fails (it never switches erroneously when the on-line element is good). In such a case, the probability of no failures is a good state and the probability of one failure and no switch failure is also a good state, that is, the switch reliability only multiplies the second term in Eq. (3.58). In such a case, the reliability expression becomes

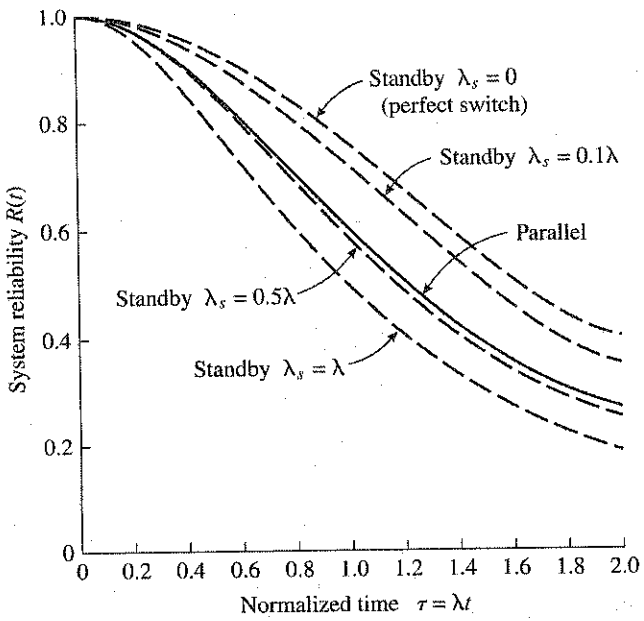


Figure 3.13 A comparison of a two-element ordinary parallel system with a two-element standby system with imperfect switch reliability.

$$R_2(t) = e^{-\lambda t} + \lambda t e^{-\lambda t} e^{-\lambda_s t} \tag{3.60}$$

Clearly, this is less conservative and a more realistic switch model than the previous one.

One can construct even more complex failure models for the switch in a standby system [Shooman, 1990, Section 6.9].

1. Switch failure modes where the switching occurs even when the on-line element is good or where the switch jitters between elements can be included.
2. The failure rate of n nonidentical standby elements was first derived by Bazovsky [1961, p. 117]; this can be shown as related to the gamma distribution and to approach the normal distribution for large n [Shooman, 1990].
3. For n identical standby elements, the system succeeds if there are n–1 or fewer failures, and the probabilities are given by the Poisson distribution that leads to the expression

$$R(t) = e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} \tag{3.61}$$

3.8 REPAIRABLE SYSTEMS

3.8.1 Introduction

Repair or replacement can be viewed as the same process, that is, replacement of a failed component with a spare is just a fast repair. A complete description of the repair process takes into account several steps: (a) detection that a failure has occurred; (b) diagnosis or localization of the cause of the failure; (c) the delay for replacement or repair, which includes the logistic delay in waiting for a replacement component or part to arrive; and (d) test and/or recalibration of the system. In this section, we concentrate on modeling the basics of repair and will not decompose the repair process into a finer model that details all of these substates.

The decomposition of a repair process into substates results in a nonconstant-repair rate (see Shooman [1990, pp. 348–350]). In fact, there is evidence that some repair processes lead to lognormal repair distributions or other nonconstant-repair distributions. One can show that a number of distributions (e.g., lognormal, Weibull, gamma, Erlang) can be used to model a repair process [Muth, 1967, Chapter 3]. Some software for modeling system availability permits nonconstant-failure and -repair rates. Only in special cases is such detailed data available, and constant-repair rates are commonly used. In fact, it is not clear how much difference there is in compiling the steady-state availability for constant- and nonconstant-repair rates [Shooman, 1990, Eq. (6.106) ff.]. For a general discussion of repair modeling, see Ascher [1984].

In general, repair improves two different measures of system performance: the reliability and the availability. We begin our discussion by considering a single computer and the following two different types of computer systems: an air traffic control system and a file server that provides electronic mail and network access to a group of users. Since there is only a single system, a failure of the computer represents a system failure, and repair will not affect the system *reliability* function. The *availability* of the system is a measure of how much of the operating time the system is up. In the case of the air traffic control system, the fact that the system may occasionally be down for short time periods while repair or replacement goes on may not be tolerable, whereas in the case of the file server, a small amount of downtime may be acceptable. Thus a computation of both the reliability and the availability of the system is required; however, for some critical applications, the most important measure is the reliability. If we say the basic system is composed of two computers in parallel or standby, then the problem changes. In either case, the system can tolerate one computer failure and stay up. It then becomes a race to see if the