CMU-CS-81-128

Transient Error Reliability Models Based on Data Analysis

Stephen R. McConnel
Daniel P. Siewiorek

June 1981

Departments of Electrical Engineering and Computer Science
Carnegie-Mellon University
Pittsburgh, Pa. 15213

This research was sponsored by the Office of Naval Research under Contract NR-048-645. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the U.S. Government.

Table of Contents

1. Introduction **3**

1.1 Background 3
1.2 Significance of Transient Errors 4
1.3 Overview of Paper 4

2. Data Collection and Analysis **5**

2.1 Data Collection 5
2.1.1 PDP-10 Systems 5
2.1.2 Experimental LSI-11 Systems 7
2.2 Data Analysis 9
2.2.1 Review of the Weibull Distribution 9
2.2.2 Maximum Likelihood Estimation and Goodness of Fit Tests 10
2.2.3 Confidence Intervals and Consonance Sets 13
2.2.4 Conclusions of the Data Analysis 16

3. Modeling Techniques **17**

3.1 Markov Models 17
3.1.1 Time Invariant Markov Models 17
3.1.2 Time Varying Markov Models 18
3.2 Monte Carlo Simulation 19

4. System Models **21**

4.1 Simplex Models 22
4.2 Duplex Models 23
4.3 TMR Models 24
4.4 Hybrid Models 26

5. Modeling Results and Comparisons **29**

5.1 Parameter Values 29
5.2 System Reliability 29
5.3 Reliability Difference Relative to Constant Hazard Rate Systems 30
5.4 Mission Time Improvement Relative to Constant Hazard Rate Systems 35
5.5 Reliability Difference Relative to Simplex System 44
5.6 Mission Time Improvement Relative to Simplex System 48

6. Conclusions and Comments **53**

References **57**

List of Figures

Figure 2-1: 90% Consonance Sets for Weibull Parameters 15
Figure 4-1: Simplex Model Without Error Process Renewal 22
Figure 4-2: Simplex Model With Error Process Renewal 22
Figure 4-3: State Diagram of Duplex Model Without Error Process Renewal 24
Figure 4-4: State Diagram of Duplex Model With Error Process Renewal 25
Figure 4-5: State Diagram of TMR Model Without Error Process Renewal 25
Figure 4-6: State Diagram of TMR Model With Error Process Renewal 26
Figure 4-7: State Diagram of Hybrid Model Without Error Process Renewal 27
Figure 4-8: State Diagram of Hybrid Model With Error Process Renewal 28
Figure 5-1: Simplex System Reliability 31
Figure 5-2: Duplex System Reliability 32
Figure 5-3: TMR System Reliability 33
Figure 5-4: Hybrid System Reliability 34
Figure 5-5: Reliability Difference of Simplex System Relative to $\alpha_e = 1.0$ 36
Figure 5-6: Reliability Difference of Duplex System Relative to $\alpha_e = 1.0$ 37
Figure 5-7: Reliability Difference of TMR System Relative to $\alpha_e = 1.0$ 38
Figure 5-8: Reliability Difference of Hybrid System Relative to $\alpha_e = 1.0$ 39
Figure 5-9: Mission Time Improvement of Simplex System Relative to $\alpha_e = 1.0$ 40
Figure 5-10: Mission Time Improvement of Duplex System Relative to $\alpha_e = 1.0$ 41
Figure 5-11: Mission Time Improvement of TMR System Relative to $\alpha_e = 1.0$ 42
Figure 5-12: Mission Time Improvement of Hybrid System Relative to $\alpha_e = 1.0$ 43
Figure 5-13: Reliability Difference Comparing Duplex to Simplex System 45
Figure 5-14: Reliability Difference Comparing TMR to Simplex System 46
Figure 5-15: Reliability Difference Comparing Hybrid to Simplex System 47
Figure 5-16: Mission Time Improvement Comparing Duplex to Simplex System 49
Figure 5-17: Mission Time Improvement Comparing TMR to Simplex System 50
Figure 5-18: Mission Time Improvement Comparing Hybrid to Simplex System 51

List of Tables

Table 1-1: Ratios of Transient Errors to Permanent Failures
Table 2-1: Summary of Collected PDP-10 Data
Table 2-2: Summary of Collected Cm* and C.mmp Data
Table 2-3: Test of Exponential Distribution
Table 2-4: Test of Weibull Distribution
Table 2-5: 90% Confidence Intervals for Weibull Parameters
Table 5-1: Parameter Values for Experiment
Table 5-2: Test Results for Validating Simulations
Table 6-1: Mission Reliabilities for Design Example

Abstract

Experimental data on transient errors from several digital computer systems is presented and analyzed. This is the first large scale public study on the statistical distribution of transient errors. The systems for which data has been collected are the DEC PDP-10 series computers, the Cm* multiprocessor, and the C.vmp fault tolerant microprocessor. Statistical tests indicate that transient errors follow a decreasing hazard rate distribution. This is at variance with the standard assumption of constant hazard rates (exponential distribution) used in reliability modeling, and requires models of greater complexity for accurate results. Models of common fault tolerant redundant structures are developed using the Weibull distribution, which has a time-varying hazard rate. Both analytical and simulation models are used to analyze the differences between the reliabilities predicted by Weibull based transient error models and those predicted by exponential based models. The analysis indicates a significant difference between the models based on the exponential distribution and those based on the decreasing hazard rate Weibull distribution. Reliability differences ranging from -0.10 to +0.20 and factors greater than 2.0 in Mission Time Improvement for Weibull parameters equivalent to measured system behavior are seen in the model results. System designers should be aware of these differences.

Chapter 1
Introduction

1.1 Background

The problems posed to digital computers by transient faults have largely been neglected in the research literature, except for occasional items of limited scope (e.g., [Wakerly 75]). A general model for transient fault occurrence and recovery based on a Markov model with the assumed exponential distribution for interarrival times was presented in [Avizienis 77]. This approach is typical of efforts to model all types of faults -- permanent [DoD 74] and intermittent [Spillman 77, Su et al. 78] as well as transient.

The original focus of this research was the collection of transient error data for analysis. Many studies of actual failure data have supported the use of the constant failure rate (exponential) process as an accurate mathematical model for hard failures. No such study has been published for transient errors. Without data, there has been no basis for either supporting or disputing the extension of earlier probabilistic models for hard failures to cover transient errors also.

Data collected from Cm*, a multi-microprocessor built at Carnegie-Mellon University, indicated a decreasing error rate [Tsao 78]. This spurred analysis of data from other systems, notably large timesharing computers in use at Carnegie-Mellon, and C.vmp, another experimental microprocessor system. When data from these systems was analyzed, similar decreasing error rates were also observed [McConnel et al. 79]. These consistent results, which will be developed in detail later in the paper, suggested a new goal of investigating reliability models which incorporate time-varying error rates. Such models present difficuities which do not arise when constant error rates can be assumed.

1.2 Significance of Transient Errors

The importance of transient faults comes from their relatively frequent occurrence. Several studies have shown that permanent faults cause only a small fraction of all detected errors. The available figures for several systems are given in Table 1-1 below [Fuller & Harbison 78, Morganti et al. 78, Morganti 78, Siewiorek et al. 78a]. The last row of figures are estimates comparing the hard and soft failure rates for a one megaword by 37 bit memory array composed of 4K MOS RAMs [Geilhufe 79, Ohm 79]. (Soft failures are transient failures caused by the radioactivity in the packaging material of the chips.)

		Detection	MTTF	MTTF	
System	Technology	Mechanism	(transient)	(permanent)	MTTE/MTTF
CMUA PDP-10	ECL	Parity	44 hrs.	800-1600 hrs.	0.03-0.06
Cm* LSI-11	NMOS	Diagnostics	128 hrs.	4200 hrs.	0.03
C.vmp	TMR LSI-11	Crash	97-328 hrs.	4900 hrs.	0.02-0.07
Telettra	TTL	Mismatch	80-170 hrs.	1300 hrs.	0.06-0.13
1M x 37 RAM	MOS	(Parity)	106 hrs.	1450 hrs.	0.07

Table 1-1: Ratios of Transient Errors to Permanent Failures

1.3 Overview of Paper

Chapter 2 presents the data analysis results, showing that a decreasing hazard rate distribution fits the collected data much better than the constant hazard rate (exponential) distribution. Techniques for modeling the reliability of systems whose modules exhibit nonconstant error rates are described in Chapter 3, and the models developed for this paper are given in Chapter 4. The results of the models are presented and compared in Chapter 6, and the major results are summarized in Chapter 7.

Chapter 2
Data Collection and Analysis

2.1 Data Collection

2.1.1 PDP-10 Systems

The PDP-10[1] is a general purpose 36-bit computer packaged either as a DECsystem-10 running a time sharing operating system called TOPS-10, or as a DECsystem-20 running TOPS-20 [Bell et al. 78]. The main system for the Computer Science Department at Carnegie-Mellon University is a DECsystem-10 which has a KL-10 (ECL) processor, one megaword of memory, eight disk drives totaling 1600 megabytes of online storage, and two magnetic tape drives. Data was collected on this system between January 15, 1980 and January 2, 1981.

Another system for which data has been collected in this study is a DECsystem-20 operated by the university's Computation Center to support the general educational needs of the university. This system has 512K words of memory, three disk drives totaling 528 megabytes of online storage, and two magnetic tape drives. Data was collected on this system from September 24, 1978 to March 17, 1979.

The core of the PDP-10 error reporting system is the online error log file maintained by the TOPS-10 and TOPS-20 operating systems. Entries are made in this file for a variety of reasons, most notably system reloads, memory parity errors, and I/O errors [DEC 78]. Each entry contains the date and time at which it was made, the processor serial number, and information about the type of error or other condition being reported.

To facilitate statistical analysis of transient errors on PDP-10's, a program to derive interarrival

[1] DEC, PDP-10, DECsystem-10, TOPS-10, KL-10, DECsystem-20, TOPS-20, PDP-11, and LSI-11 are all registered trademarks of Digital Equipment Corporation.

times and time-of-day distributions from the system error log files has been written. This program is named SEADS, which stands for Statistical Error Analysis Data Summarization. The outputs generated include graphs of the distributions of interarrival times for various types of entries and lists of those interarrival times.

Two types of entries in the PDP-10 error log files were chosen as being likely to reflect transient errors: system reloads and memory parity errors. System reloads were chosen because they are commonly caused by the ubiquitous "crash". In systems with stable hardware and matured software, the most frequent cause of crashes appears to be transient errors. Attempts to separate reloads caused by transient errors from those caused by permanent failures, software failures, and scheduled maintenance, have proven to be unsuccessful due to the paucity of information provided in the reload entries. Thus this data is only indicative of transient error behavior. On the other hand, there seems to be a general consensus that transients are responsible for the great majority of memory parity errors, so these should provide a better measure.

In scanning the data processed by SEADS, it became clear that the PDP-10 systems frequently recorded several errors for one fault. To mask out the effects of this, error entries within five minutes of a previous entry were counted as a part of the previous fault. It was felt that five minutes was a reasonable threshold for the study. The software allowed any choice for the threshold, facilitating examination of the the sensitivity of the data to threshold values[1]. This process of eliminating very short interarrival times is referred to as "filtering" in this paper.

Table 2-1 summarizes the data collected from the PDP-10 systems. Detailed analyses of these data sets are performed later in this section. Unfortunately for purposes of data analysis, too few parity errors occurred on the TOPS-20 system to be statistically significant. Thus only three data sets are analyzed: system reloads for both the TOPS-10 and TOPS-20 systems, and memory parity errors for the TOPS-10 system.

[1] Threshold values of one minute and ten minutes were also tried without significantly changing the results presented in this paper.

	TOPS-10	TOPS-20
Total Hours Covered	8,272	3,282
Approximate Coverage	0.979	0.790
Total Entries	84,918	19,690
Total Reloads	748	276
Reload MTTE (hours)	11.1	11.9
"Filtered" Reloads	637	222
"Filtered" MTTE (hours)	11.4	13.0
Total Parity Errors	120	31
Parity MTTE (hours)	68.9	105.9
"Filtered" Parity Errors	103	12
"Filtered" MTTE (hours)	70.7	239.9

Table 2-1: Summary of Collected PDP-10 Data

2.1.2 Experimental LSI-11 Systems

In addition to the PDP-10 system error log files, data was also collected from two experimental LSI-11 systems at Carnegie-Mellon University. The first of these is a multiprocessor system named Cm*, whose structure has been widely reported in the literature [Swan et al. 77, Swan 78]. Cm* uses a network of buses to give processors (slightly modified LSI-11's) access to a large shared memory. The system is built from processor-memory pairs called Computer Modules or Cm's. The memory local to a processor is also the shared memory of the system. Each Cm has a local switch, or Slocal, which interfaces its bus to the rest of the system. A mapping controller, or Kmap, is shared by several Cm's, which are connected to it by their Slocal's to form a "cluster". Each Kmap in the system is connected via multiple intercluster buses to other Kmaps, completing the interconnection scheme. The Kmaps perform all the functions necessary for meeting both intracluster and intercluster memory requests.

The configuration of Cm* for which data was collected contained ten Cm's connected to three Kmaps to form three clusters, two with four Cm's and one with two Cm's. Each Cm had a serial link to the Cm* host computer, a message switching PDP-11 [Seelza 77], to facilitate user access and program loading.

Automatic diagnostic software was developed to exercise idle modules in Cm*. One such idle module was used to run the master control program CMDIAG. This module operated in a special mode which enabled it to control other Cm's as though it was a user at a terminal. With this ability, CMDIAG acquired control of all unassigned modules and continuously cycled each of them through a series of four diagnostic tests. The program was able to dynamically acquire and release modules in response to the changing needs of the other users.

The automatic diagnostic system for Cm* was exercising at least some of the modules for almost 50% of the time between May 1977 and April 1978. A total of 104 separate transient errors were recorded during a total diagnostic session time of 4223 hours. Some of these transients caused multiple errors on single modules, and several produced simultaneous errors in more than one module [Tsao 78].

The second experimental system for which data was collected is C.vmp, a triplicated NMOS LSI-11 microprocessor with voting at the bus level [Siewiorek et al. 78a, Siewiorek et al. 78b]. There are three processor-memory pairs, each pair connected via a voter circuit. During the time that data was collected, C.vmp contained 12K words of dynamic MOS memory and a dual floppy disk drive per processor.

The transient error data from C.vmp was gleaned from the system maintenance notebook entries for August 1977 to April 1978. During this period, C.vmp was used for several months under actual load conditions with students doing projects for an introductory real time programming course. The students were supplied with a standard system software manual and a small pamphlet on information specific to C.vmp (i.e., location of the power switches, a reminder to load three diskettes, etc.). Because of the disparate nature of the user community and the complexity of the system, the causes of many crashes remain ambiguous. During the 4921 hours of recorded operation, at least 15 crashes were definitely due to transient hardware errors; however, the actual number of crashes caused by transient errors may have been as high as 51. The only transients which should cause C.vmp to crash are those occurring simultaneously in more than one module. According to the data from Cm*, such transients make up 17% of the total, occurring roughly every 250 hours [Siewiorek et al. 78b]. The mean time to crash for C.vmp should equal or exceed this figure. Indeed, the "best case" and "worst case" figures for MTTE fall around this value, as shown in Table 2-2 below. Note that crashes which *may* have been software or user caused are included only in the worst case data for C.vmp.

	Cm*	C.vmp (best)	C.vmp (worst)
Total Hours	4223	4921	4921
Total Errors	104	15	51
MTTE (hours)	40.6	328.1	96.5
Standard Deviation	59.8	470.8	167.8

Table 2-2: Summary of Collected Cm* and C.vmp Data

2.2 Data Analysis

Two questions must be answered by examining the data:

1. Which probability distribution best fits the data?

2. What are reasonable parameter values for the collected data and chosen distribution?

In theory, the second question is meaningless until the first question is answered. In practice, parameter values must be estimated in order to test the applicability of different distributions. The choice of which probability distributions to test is based on such evidence as the distribution histogram of interarrival times produced by SEADS. The general shape of these histograms indicated that perhaps some generalized form of the exponential distribution would be appropriate. The Weibull distribution is one such extension of the exponential, and is widely used in reliability modeling and analysis.

2.2.1 Review of the Weibull Distribution

The Weibull distribution has two parameters: α (the shape parameter) and λ (the scale parameter). The probability density function, cumulative distribution function, reliability function, and hazard (error rate) function, of the Weibull distribution are shown in Equations (2.1) through (2.4) ($\alpha > 0$, $\lambda > 0$):

$$\text{pdf} = f(t) = \alpha\lambda(\lambda t)^{\alpha-1}\, e^{-(\lambda t)^{\alpha}} \tag{2.1}$$

$$\text{CDF} = F(t) = 1 - e^{-(\lambda t)^{\alpha}} \tag{2.2}$$

$$\text{Reliability} = R(t) = e^{-(\lambda t)^{\alpha}} \tag{2.3}$$

$$\text{Hazard Function} = z(t) = \alpha\lambda(\lambda t)^{\alpha-1} \tag{2.4}$$

Note that the values of all these functions depends on time only through the product of the scale actor and time -- λt.

The shape parameter directly influences the error rate:

- if $\alpha < 1$, the error rate is decreasing with time;
- if $\alpha = 1$, the error rate is constant with time, resulting in an exponential distribution; and
- if $\alpha > 1$, the error rate is increasing with time.

If t takes only the discrete times 0,1,2,..., then the discrete Weibull distribution is obtained from the Weibull distribution by substituting q for $e^{-\lambda^\alpha}$, and n for t [Nakagawa & Osaki 75]. The probability mass function, cumulative distribution function, reliability function, and hazard function of the discrete Weibull distribution are shown in Equations (2.5) through (2.8).(0 < q < 1):

$$\text{pmf} = f(n) = q^{n^\alpha} - q^{(n+1)^\alpha} = q^{n^\alpha}\left(1 - q^{(n+1)^\alpha - n^\alpha}\right) \tag{2.5}$$

$$\text{CDF} = F(n) = 1 - q^{n^\alpha} \tag{2.6}$$

$$\text{Reliability} = R(n) = q^{n^\alpha} \tag{2.7}$$

$$\text{Hazard function} = z(n) = 1 - q^{(n+1)^\alpha - n^\alpha} \tag{2.8}$$

The mean μ of the discrete Weibull function is given by

$$\mu = E(k) = \sum_{k=0}^{\infty} q^{k^\alpha} \tag{2.9}$$

In this paper, the only use of the discrete Weibull distribution is to approximate the Weibull distribution.

2.2.2 Maximum Likelihood Estimation and Goodness of Fit Tests

A common technique for estimating parameter values is maximum likelihood estimation. The maximum likelihood estimator (M.L.E.) of a parameter θ is the value $\hat{\theta}$ which maximizes the probability ("likelihood") of the observed data sample. For the exponential distribution, the M.L.E. of λ is given by

$$\hat{\lambda} = N \Big/ \sum_{j=1}^{N} t_j \tag{2.10}$$

The maximum likelihood estimators (MLE) $\hat{\alpha}$ and $\hat{\lambda}$ for the Weibull distribution satisfy the following equations [Thoman et al. 69]:

$$(N / \hat{\alpha}) + \sum_{j=1}^{N} \log_e(t_j) = N \times \left(\sum_{j=1}^{N} t_j^{\hat{\alpha}} \times \log_e(t_j)\right) \Big/ \left(\sum_{j=1}^{N} t_j^{\hat{\alpha}}\right) \tag{2.11}$$

$$(\hat{\lambda})^{\hat{\alpha}} = N \Big/ \sum_{j=1}^{N} t_j^{\hat{\alpha}} \tag{2.12}$$

Once the value of the shape parameter $\hat{\alpha}$, is known, Equation (2.12) can be used to calculate the scale parameter $\hat{\lambda}$. Equation (2.11) can be used to derive a difference equation in the form

$$\hat{\alpha}_{i+1} = \text{Function}\,(\hat{\alpha}_i, \vec{T}_N)$$

where $\vec{T}_N$ is the vector of observed data. A quickly converging solution can be found using the Newton-Raphson method, as was presented in [Thoman et al. 69].

The problem of fitting a suitable probability distribution to a random sample of N observations, $\{t_1, t_2, ..., t_N\}$, is addressed by the χ^2 goodness of fit test. For data taken from a continuous distribution (such as the exponential and Weibull distributions), the real number line is divided into K intervals, and the deviations between the observed and expected number of data points within each interval is recorded by the following statistic:

$$Q = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

where

O_i = observed number of samples in i^th interval;

E_i = expected number of samples in i^th interval.

The statistic Q approximately follows a χ^2 distribution (hence the name of the test). For meaningful results, the expected number of samples in any interval should be at least five.

The level of significance of a goodness of fit test is the *prespecified* probability of erroneously rejecting the hypothesis that the data is from the given distribution. (Typical values chosen for the level of significance range between 0.01 and 0.10.) However, a different quantity called the ***p-value*** is frequently reported for goodness of fit tests. The p-value is the empirical probability that rejection of the hypothesis under test would be erroneous. In other words, a reported p-value of 0.25 means that the hypothesis being tested would pass the goodness of fit test for any a priori level of significance less than or equal to that figure. A large p-value is generally considered to be good evidence that the data fits the given distribution.

Typically, the value of E_i in the expression above is calculated for each interval using the M.L.E. parameter values. If this is the case, then only lower and upper bounds for the test p-value can be found. For a distribution under test with n estimated parameters, the lower bound is given by the χ^2 distribution with K-n-1 degrees of freedom. The upper bound is given by the χ^2 distribution with K-1 degrees of freedom.

For data drawn from an unknown continuous probability distribution, an alternative means for testing the goodness of fit of a proposed distribution is provided by the Kolmogorov-Smirnov test. Let the empirical cumulative distribution function be defined as

$$F_N(t) = \frac{i}{N} \quad \text{for } t_i \le t < t_{i+1}$$

and the cumulative distribution function under test

$$F^*(t) = F(t\,;\hat{\theta})$$

where $\hat{\theta}$ is the M.L.E. parameter derived from the data sample $\{t_1, t_2, ..., t_N\}$. Define the following statistic:

$$D_N = \sup_{-\infty<t<\infty} \left| F_N(t) - F^*(t) \right|$$

where "sup" denotes the *supremum*, or least upper bound, of its argument. This statistic D_N is called the two-sided Kolmogorov-Smirnov statistic, and follows a probability distribution of the same name. D_N is the maximum difference between the empirical $F_N(t)$ and the hypothesized $F^*(t)$. Values of D_N tend to be smaller (for given sample size N) for hypothesized distributions which fit the data better. The p-value for this test can be calculated by evaluating the cumulative distribution function of the Kolmogorov-Smirnov distribution.

Table 2-3 presents the M.L.E. parameter values and goodness of fit test results for applying the exponential distribution to the transient error data described earlier. Table 2-4 contains the same information for testing the Weibull distribution. In all cases, the Weibull distribution shows a much better fit to the data than does the exponential distribution.

	TOPS-10 Reload	TOPS-20 Reload	TOPS-10 Parity	Cm*	C.vmp[1]
M.L.E. λ	0.0875	0.0771	0.0141	0.0246	0.0104
χ^2 test p-value / degrees of freedom					
lower bound for worst fit	0.000 / 1	0.000 / 4	0.006 / 4	0.013 / 3	0.001 / 1
upper bound for worst fit	0.000 / 2	0.000 / 5	0.012 / 5	0.029 / 4	0.005 / 2
lower bound for best fit	0.000 / 1	0.007 / 38	0.090 / 12	0.299 / 13	0.010 / 2
upper bound for best fit	0.000 / 2	0.009 / 39	0.126 / 13	0.369 / 14	0.027 / 3
Kolmogorov-Smirnov test p-value	0.000	0.001	0.054	0.056	0.002

Table 2-3: Test of Exponential Distribution

The presentation of four numbers for the p-value of the χ^2 test is due to a combination of two factors. First is the problem mentioned earlier that only upper and lower bounds are obtainable when

[1] The "worst case" data for C.vmp is used here because there are too few crashes in the "best case" data. Even that limited data supports the major conclusions of this section.

	TOPS-10 Reload	TOPS-20 Reload	TOPS-10 Parity	Cm*	C.vmp
M.L.E. α	0.778	0.793	0.751	0.779	0.654
M.L.E. λ	0.101	0.0882	0.0166	0.0288	0.0146
χ^2 test p-value / degrees of freedom					
lower bound for worst fit	0.000 / 40	0.043 / 2	0.169 / 1	0.239 / 1	0.121 / 3
upper bound for worst fit	0.000 / 42	0.179 / 4	0.595 / 3	0.709 / 3	0.324 / 5
lower bound for best fit	0.134 / 8	0.691 / 6	0.822 / 10	0.934 / 12	0.813 / 7
upper bound for best fit	0.258 / 10	0.867 / 8	0.920 / 12	0.975 / 14	0.930 / 9
Kolmogorov-Smirnov test p-value	0.330	0.687	0.756	0.935	0.690

Table 2-4: Test of Weibull Distribution

the M.L.E. parameter values are used in this test. The second problem is the choice of how many intervals to use in dividing up the data. The decision made was to treat this as another upper bound/lower bound problem, and to present the values indicating the best fit to the data and the worst fit. It is especially significant that the worst fit to the Weibull distribution is almost always much better than the best fit to the exponential distribution. This emphasizes that the decreasing hazard rate distribution describes the data more accurately than the constant hazard rate distribution.

2.2.3 Confidence Intervals and Consonance Sets

Point estimates, such as those obtained by maximum likelihood estimation, are only approximations which very rarely match exactly the values they are intended to estimate. Because of this, interval estimates are often desirable. These are intervals for which it can be asserted with some certainty that they contain the actual value of the parameter under consideration. The most common use of this idea is expressed in "confidence intervals". For $0 < p < 1$, a p-level confidence interval is a range within which the actual value of the estimated parameter would fall with probability p, if the experiment were repeated many times. To restate this, saying that a certain range of values is a 0.90 confidence interval for a parameter means that in repeated sampling, 90% of the confidence intervals so constructed would contain the actual parameter values [Miller & Freund 65].

The concept of "consonance sets" has been developed in an effort to simultaneously answer both of the basic questions of statistical data analysis: whether a particular probability distribution $F(t\ ;\ \theta)$ describes the data sample, and (if so) which values for the parameter θ are reasonable [Easterling 76]. A p-level consonance set has much the same properties as a p-level confidence interval, in that, for repeated sampling from a given distribution, the fraction p of the sets (intervals) calculated would contain the true value(s) of the parameter(s). For a fixed sample size and significance level, data which is more consonant with a proposed distribution produces a larger consonance set. This is due to the goodness of fit considerations involved with constructing consonance sets.

The general method for constructing consonance sets proceeds as follows [Salvia 79].

1. Choose an appropriate goodness of fit statistic, such as the Kolmogorov-Smirnov statistic D_N. (Evidence shows that the χ^2 statistic is *not* a good choice [Salvia 80].)
2. Set K equal to the value at which the cumulative distribution function of D_N (in this case, the Kolmogorov-Smirnov distribution) reaches the probability level p. (This is sometimes called the p^{th} fractile of the distribution of D_N.)
3. Consider the statistic D_N as a function of the parameter θ of the distribution under consideration. The *consonance set* C is defined as those values of θ for which $D_N \leq K$. (Recall that the calculation of D_N depends on the value of θ.)

A consonance set obtained in this way plots as a line segment for one-parameter distributions, and as a bounded area for distributions with two parameters. Specific techniques for constructing consonance sets for the Weibull and two-parameter exponential distributions are given in [Salvia 79].

Figure 2-1 displays 90% consonance sets for the Weibull parameters constructed from each of the five sets of data. As explained above, these sets contain all values of the parameters for which the Kolmogorov-Smirnov goodness of fit test passes the data at a significance level less than or equal to 0.10 (i.e., 1-p). The TOPS-10 and TOPS-20 reload data yield no parameter values consistent with the exponential distribution ($\alpha = 1$). The other three data sets, however, do produce some parameter values for which the exponential distribution is plausible. Because of this, 90% confidence intervals were constructed for these particular data sets using techniques developed in [Thoman et al. 69]. These confidence intervals are given in Table 2-5.

	TOPS-10 Parity	Cm*	C.vmp
Confidence Interval for α	[0.655 , 0.844]	[0.680 , 0.875]	[0.531 , 0.767]
Confidence Interval for λ	[0.0132 , 0.0209]	[0.0231 , 0.0359]	[0.0099 , 0.0214]

Table 2-5: 90% Confidence Intervals for Weibull Parameters

As seen in the table, none of the confidence intervals for the shape parameter α includes the value one, i.e., the constant hazard rate (exponential) function. Also, it should be noted that none of the corresponding consonance sets contains the M.L.E. scale parameter λ for the exponential distribution (indicated in lower case on the graphs). Thus, even though the exponential distribution cannot be ruled out absolutely, it does seem most unlikely. The decreasing hazard rate Weibull distribution, on the other hand, provides a good fit to these three data sets as well as to the PDP-10 reload data.

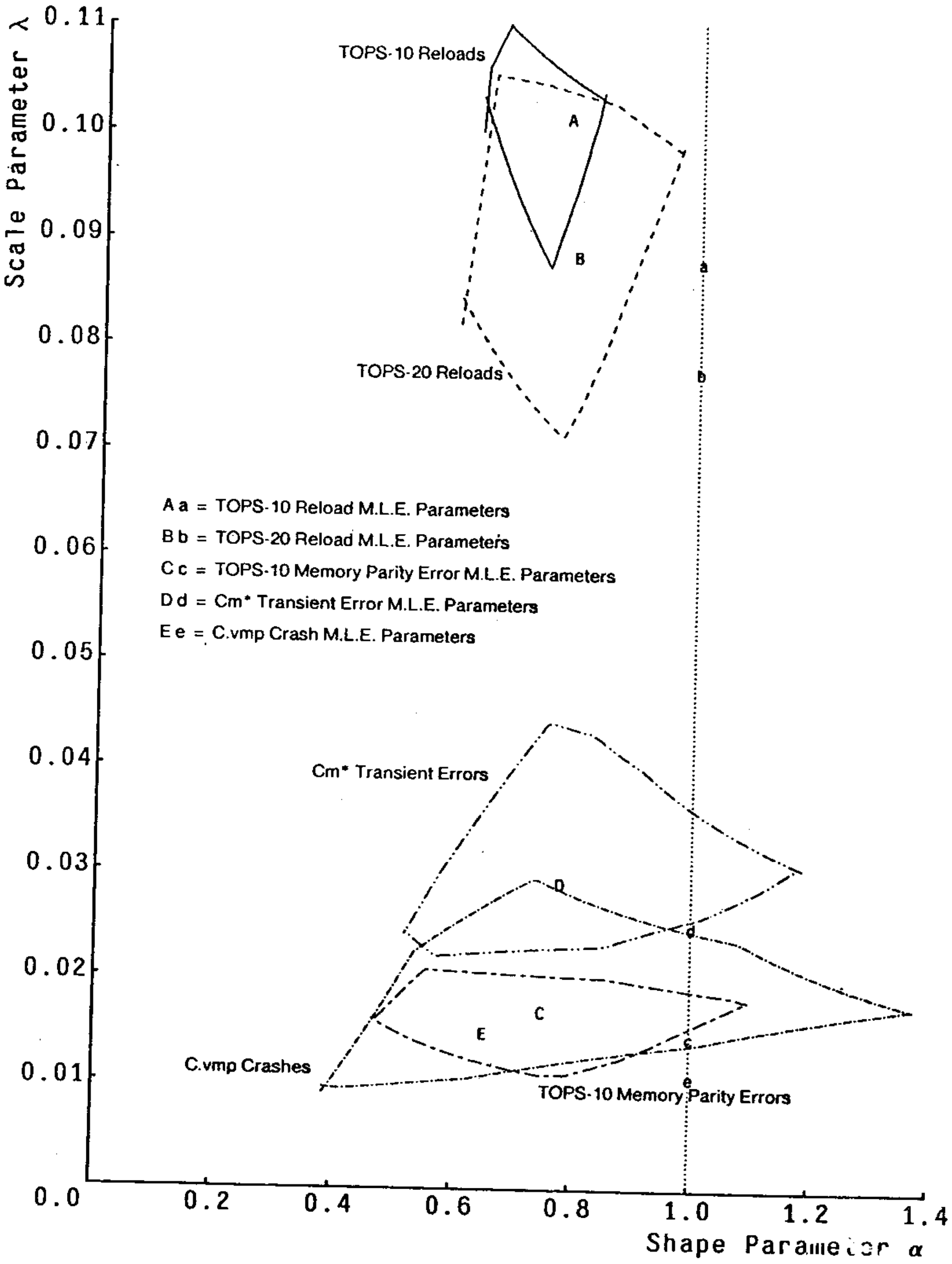


Figure 2-1: 90% Consonance Sets for Weibull Parameters

2.2.4 Conclusions of the Data Analysis

Data collected from several different systems has been presented and analyzed. These systems range in size from an NMOS microprocessor with 12K words of memory to ECL mainframes with one megaword of memory, and range in redundancy from nonredundant to parity to triplication. In every case, the data shows an acceptable fit to a decreasing hazard rate Weibull distribution. In fact, most of the data sets show excellent fits to the Weibull family of distributions. For only two sets of data (TOPS-10 memory parity errors and Cm* transient errors) does the exponential distribution appear to be even plausible. In each case, the fit to the exponential distribution is poor, while the fit to the decreasing hazard rate Weibull distribution is extremely good.

This result completely upsets the common assumption of constant hazard rates for transient errors used in reliability modeling. The impact of time-varying hazard rates on system reliability models is the topic of the remainder of this paper.

Chapter 3
Modeling Techniques

3.1 Markov Models

A powerful tool for analyzing complex probabilistic systems is the Markov process model. The two central concepts of such models are "state" and "state transition". The *state* of a system represents all that must be known to describe the system at any instant of time. For reliability models, each state represents a distinct combination of working and failed modules. If each module is in one of two conditions -- working or failed -- then the complete model for a system of n modules has 2^n states. As time passes, the system goes from state to state as modules fail and recover. These changes of state are called *state transitions.* Discrete time models require all state transitions to occur at fixed time intervals, and assign probabilities to each possible transition. Continuous time models allow state transitions to take place at varying, random intervals, with differential transition rates assigned to possible transitions. For reliability models, the transition rates are based on the module error hazard rates and recovery hazard rates.

3.1.1 Time Invariant Markov Models

The basic assumption underlying Markov models is that the probability of a given state transition depends only on the current state. For continuous time Markov processes, it is also assumed that the length of time already spent in a state does not influence either the probability distribution of the next state or the probability distribution of remaining time in the same state before the next transition. These are very strong assumptions, and imply that the waiting time spent in any one state is geometrically distributed in the discrete time case, or exponentially distributed in the continuous time case [Howard 71]. Thus, the Markov model naturally fits in with the standard assumption that error rates are constant, leading to exponentially distributed interarrival times for errors.

3.1.2 Time Varying Markov Models

A useful generalization of the Markov model for reliability modeling is to allow state transition probabilities to change over time. This causes difficulties in obtaining solutions, since it generally makes the use of transform analysis impossible [Howard 71]. Nevertheless, if error rates are functions of time, the techniques discussed in this section can be used.

Define $q_{ij}(m,n)$ as the probability that the system is in state j at-time n given that it was in state i at time m ($m \leq n$). For consistency, $Q(m,m) = I$ (the identity matrix). With this notation, in matrix form the "Chapman-Kolmogorov" equation is

$$Q(m,n) = Q(m,k)\, Q(k,n) \quad \text{for } m \leq k \leq n$$

Letting $k = n - 1$

$$Q(m,n) = Q(m,n-1)\, Q(n-1,n)$$

Defining $P(n) = Q(n,n+1)$

$$Q(m,n) = Q(m,n-1)\, P(n-1) \tag{3.1}$$

This equation can be expanded recursively

$$Q(m,n) = Q(m,n-2)\, P(n-2)\, P(n-1)$$

$$Q(m,n) = Q(m,n-3)\, P(n-3)\, P(n-2)\, P(n-1)$$

The solution finally becomes

$$Q(m,n) = \prod_{i=m}^{n-1} P(i) \tag{3.2}$$

For $m = 0$ and all $P(i) = P$, this becomes P^n, the classical solution for discrete time Markov models.

Corresponding equations can be derived for the continuous time case. This is usually done for time invariant models, as the the model is easily set up in terms of differential rates. However, the solution of time-varying models requires the use of numerical integration techniques due to their complexity [Stiffler et al. 79]. An alternative method is to approximate the continuous time processes with discrete time equivalents. Since numerical integration involves some degree of approximation anyway, this is frequently a good choice. The major difficulty is that many transition rates which are effectively zero in the continuous time differential transition rate matrix assume small but nonzero probabilities in the discrete time transition probability matrix. For the systems modeled in this paper, the number of states (and therefore the number of state transitions) is small enough to encourage the use of discrete time Markov models.

For converting from continuous time hazard functions (error and recovery rate functions) to discrete time hazard functions, a discrete time probability distribution corresponding to the continuous time distribution defined by that hazard function must be found. The corresponding parameters can then be calculated for the desired time step. Recall that the Weibull distribution hazard function

$$z(t) = \alpha\lambda(\lambda t)^{\alpha-1}$$

has a corresponding discrete distribution hazard function

$$z(k) = 1 - q^{(k+1)^{\alpha}-k^{\alpha}}$$

Assuming a unit time step, the correspondence between these two functions is given by

$$q = e^{-\lambda^{\alpha}}$$

The continuous time hazard function represents a differential rate of change from one state to another. The discrete time hazard function is a probability of changing state at the next time step.

3.2 Monte Carlo Simulation

The techniques considered thus far are insufficient to obtain results when even quite reasonable changes are made in the modeling assumptions. Consider the issue of error process renewal. It seems rather intuitive that a recovered module should be "as good as new", but that is *not* the assumption behind the Markov models discussed above. In those models, the error processes are not reset to time $t=0$ $(n=0)$ when a module recovers. This can make a dramatic difference in the error rates. In the Weibull hazard function, for α less than one, the error rate asymptotically approaches zero; for α greater than one, it grows without limit. The error rate immediately following recovery can thus vary tremendously depending on which assumption regarding renewals is made. (Of course, for constant error rates, there is no difference in effect between the two assumptions.) Consider the discrete Weibull hazard function z(n)

$$z(n) = 1 - q^{(n+1)^{\alpha}-n^{\alpha}}$$

If this error process is "renewed" (reset to time zero) whenever a recovery occurs, then the conditional hazard function of the process given the renewal time N_R is

$$z(n) = 1 - q^{(n-N_R+1)^{\alpha}-(n-N_R)^{\alpha}}$$

In general, the hazard function of the error process with renewal is given by

$$z(n) = 1 - \sum_{k=0}^{n} \left(q^{(n-k+1)^{\alpha}-(n-k)^{\alpha}}\right) Pr\{N_R = k|n\}$$

The second term in the summation is the conditional probability that the renewal time has any particular value given the current time. Calculation of this value depends on the entire past history of the system, which makes it rather intractable to compute in practice. Therefore a new technique to attack the problem of reliability modeling is needed.

A standard method of studying the reliability of systems which are too complex to model analytically is to simulate their performance and examine the results [Almassy 79, Yakowitz 77]. The basis of such "Monte Carlo" simulation schemes is a pseudo-random number generator, which produces a sequence of numbers between zero and one (0,1) that approximately follow the uniform distribution. For good results, simulations should be run using two or more independent pseudo-random number generators, and the generators used should be thoroughly tested [Knuth 69]. This was done for the simulation-based models of this chapter. Details of the generators chosen, and their test performance, are given in [McConnel 81]. Three separate generators were chosen after extensive testing.

Chapter 4
System Models

Four system organizations are modeled in this paper:

1. A nonredundant system *(simplex)*

2. A symmetric reconfigurable dual redundant system *(duplex)*

.3. A triple modular redundant system with majority voting *(TMR)*

4. A hybrid redundant system which has a triple modular redundant core plus a standby spare module *(hybrid)*.

Two types of models are developed for each of these systems. The first is a time-varying Markov model with uniform discrete Weibull error rate functions. In this model, all time-varying error rates and recovery rates follow a single global monotonic time scale. The second type of model allows independent renewals of the different error processes and recovery processes. This means that each individual error process is reset to time $t=0$ whenever a recovery occurs. (Recovery processes are reset whenever errors occur.) With this assumption, Monte Carlo simulation is used to solve for the system reliability. These simulations require two ancillary functions: a pseudo-random number generator RANDOM() and mission time function MT(r ; α,λ). For the simple Weibull reliability function

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

the corresponding mission time function is

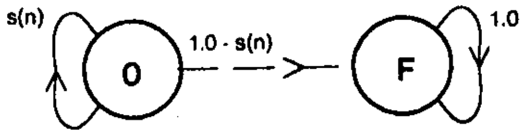
$$MT(r) = \frac{\sqrt[\alpha]{-\log_e(r)}}{\lambda} \tag{4.1}$$

Essentially, the mission time function is used to change the uniformly distributed pseudo-random numbers to follow the desired Weibull function.

4.1 Simplex Models

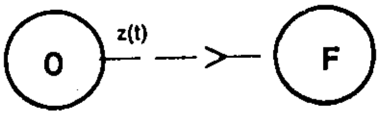
These models are trivial, as shown in Figures 4-1 and 4-2. Virtually the same model applies regardless of the assumption concerning process renewals, because the first module error causes the system to fail.

For the first model, that without error process renewals, the discrete time-varying Markov solution method using matrix multiplications is utilized. For the second model, although the results should be identical, a Monte Carlo simulation is performed. This allows a check on whether or not the simulation results (including those for other models) are reasonable. For this model (and the others described following), a total of 3000 simulations was performed, using three different pseudo-random number generators for 1000 simulations each. Each simulation of the simplex system consists only of generating r as the next pseudo-random number in sequence (r = RANDOM()), and transforming it via the mission time function MT(r,α_e,λ_e) to a time which follows the Weibull distribution.



$$\Lambda(n) = 1.0 - q_e^{(n+1)^{\alpha_e} - n^{\alpha_e}}$$

Figure 4-1: Simplex Model Without Error Process Renewal



$$z(t) = \alpha_e\lambda_e(\lambda_e t)^{\alpha_e - 1}$$

Figure 4-2: Simplex Model With Error Process Renewal

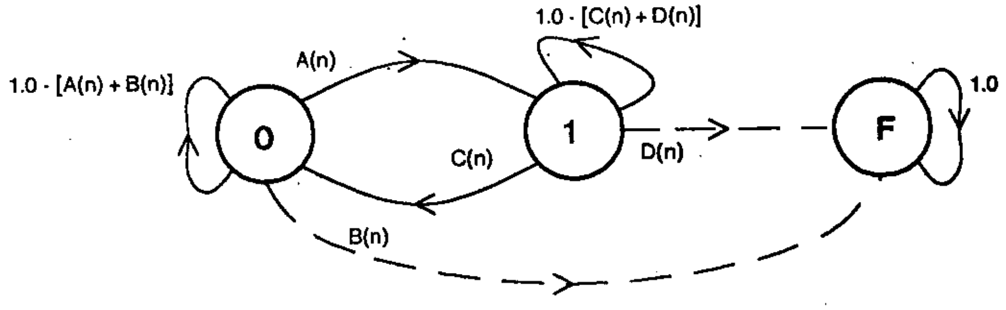
4.2 Duplex Models

The duplex system modeled in this thesis is organized as a symmetric dual system which reconfigures to a simplex system when either of the two modules fails to work properly. The first model (without error process renewal) has identical error and recovery processes for both modules at all times. Therefore the two states with one module working and one module crashed have been merged into one state. In this model, the states are labeled with the number of erring modules. State 2 thus represents the system failed state. The coverage factor c reflects the probability that a single error will cause the system to fail. (c is always assumed to equal 0.99 in this study.) The function s(n) is the time-dependent probability that an error will *not* occur at time step n, and the function r(n) is the probability that a module which has crashed will *not* recover at time step n.

The reliability of redundant systems is considerably more complex to simulate than for the simplex system. Consider the state transition diagram of a duplex system shown in Figure 4-4. Because of the need to distinguish between transient errors (and recovery from same) of individual modules, a four-state model is required. Otherwise, this models the same system as Figure 4-3, albeit with continuous rather than discrete time. Simulation of the duplex system reliability follows the algorithm given below. The algorithm assumes that the two modules are labeled 1 and 2. STATE records which module has an active error (0 if neither). The two times T[1] and T[2] record the (randomly generated) times of the next event, either error or recovery, for the corresponding module. NEXT records which module has the earlier occurring event to cause a state transition.

```
1. Global initialization
        STATE ← 0
        T[1] ← MT( RANDOM(), $\alpha_e$, $\lambda_e$ )
        T[2] ← MT( RANDOM(), $\alpha_e$, $\lambda_e$ )

2. Loop: choose next state transition
        if (T[1] < T[2]) then NEXT ← 1 else NEXT ← 2

3. Choose next state
        case STATE of
                0: if (RANDOM() $\le$ C) then STATE ← NEXT else goto step 6;
                1: if (NEXT = 1) then STATE ← 0 else goto step 6;
                2: if (NEXT = 2) then STATE ← 0 else goto step 6;
                end case

4. Calculate new transition time
        if (STATE = 0)
                then T[NEXT] ← T[NEXT] + MT( RANDOM(), $\alpha_e$, $\lambda_e$ )
                else T[NEXT] ← T[NEXT] + MT( RANDOM(), $\alpha_r$, $\lambda_r$ )
```

```
5. Repeat loop -- goto step 2

6. Return simulated mission time
        TIME ← T[NEXT]
```



$$A(n) = 2c\, s(n)[1\text{-}s(n)]$$
$$B(n) = 2[1\text{-}c]s(n)[1\text{-}s(n)] + [1\text{-}s(n)]^2$$
$$C(n) = s(n)[1\text{-}r(n)]$$
$$D(n) = [1\text{-}s(n)]r(n)$$
$$s(n) = q_e^{(n+1)^{\alpha_e} - n^{\alpha_e}}$$
$$r(n) = q_r^{(n+1)^{\alpha_r} - n^{\alpha_r}}$$

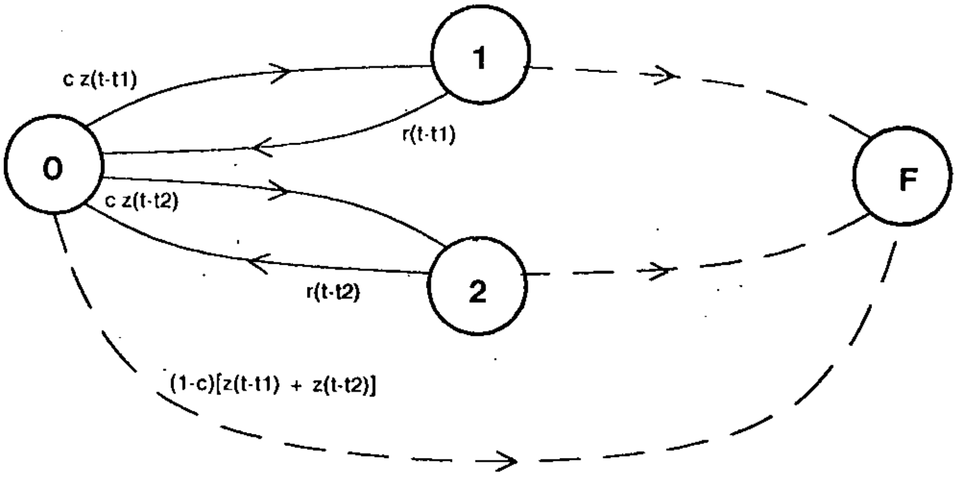
s(n) is the probability of *not* having an error occur within a module at time step n, and r(n) is the probability of not having a recovery occur at that time.

Figure 4-3: State Diagram of Duplex Model Without Error Process Renewal

4.3 TMR Models

The third structure modeled is a triple modular redundant (TMR) system. Th without error process renewal is shown in Figure 4-5. As with the duplex model, the multi vith one module not working have been merged into one state.

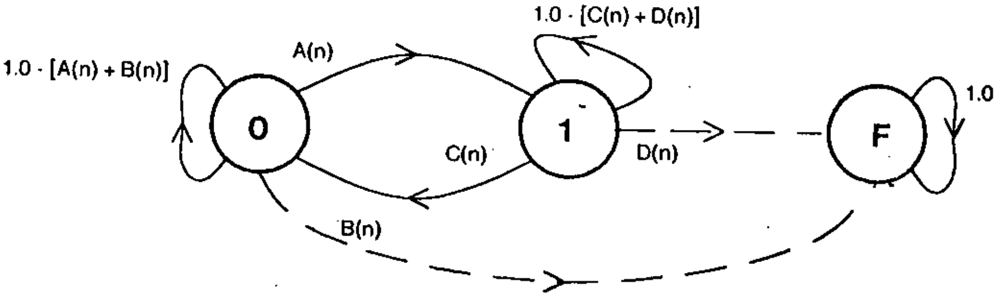
The state transition diagram of the TMR model with error process renewal is shown in Figure 4-6. The simulation algorithm is similar in outline to that presented for the duplex system model. Details can be found in [McConnel 81].



$$z(t) = \alpha_e \lambda_e (\lambda_e t)^{\alpha_e - 1}$$

$$r(t) = \alpha_r \lambda_r (\lambda_r t)^{\alpha_r - 1}$$

Figure 4-4: State Diagram of Duplex Model With Error Process Renewal

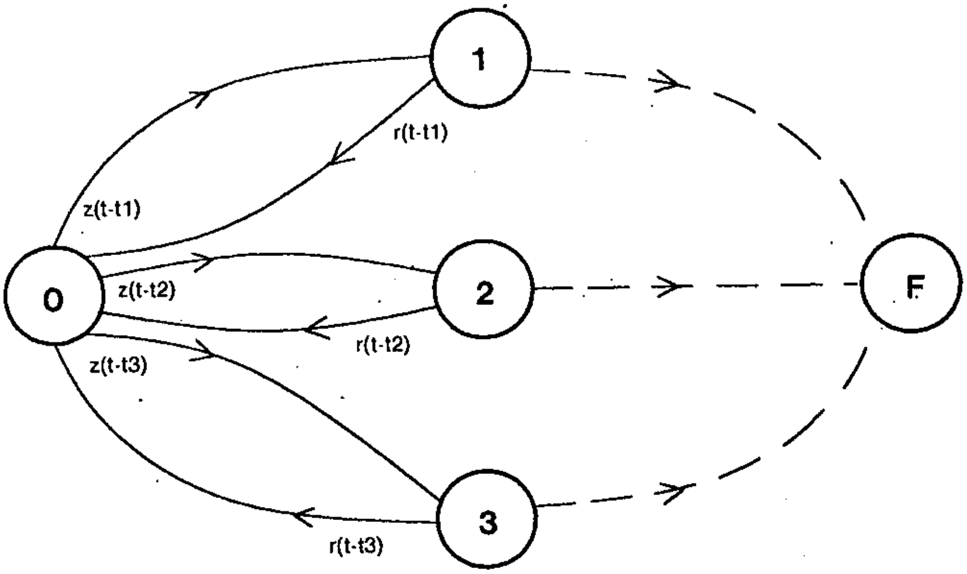


$$A(n) = [s(n)]^2[1-s(n)]$$
$$B(n) = 3s(n)[1-s(n)]^2 + [1-s(n)]^3$$
$$C(n) = [s(n)]^2[1-r(n)]$$
$$D(n) = 2s(n)[1-s(n)]r(n) + [1-s(n)]^2$$

$$s(n) = q_e^{(n+1)^{\alpha_e} - n^{\alpha_e}}$$

$$r(n) = q_r^{(n+1)^{\alpha_r} - n^{\alpha_r}}$$

Figure 4-5: State Diagram of TMR Model Without Error Process Renewal



$$z(t) = \alpha_e \lambda_e (\lambda_e t)^{\alpha_e - 1}$$

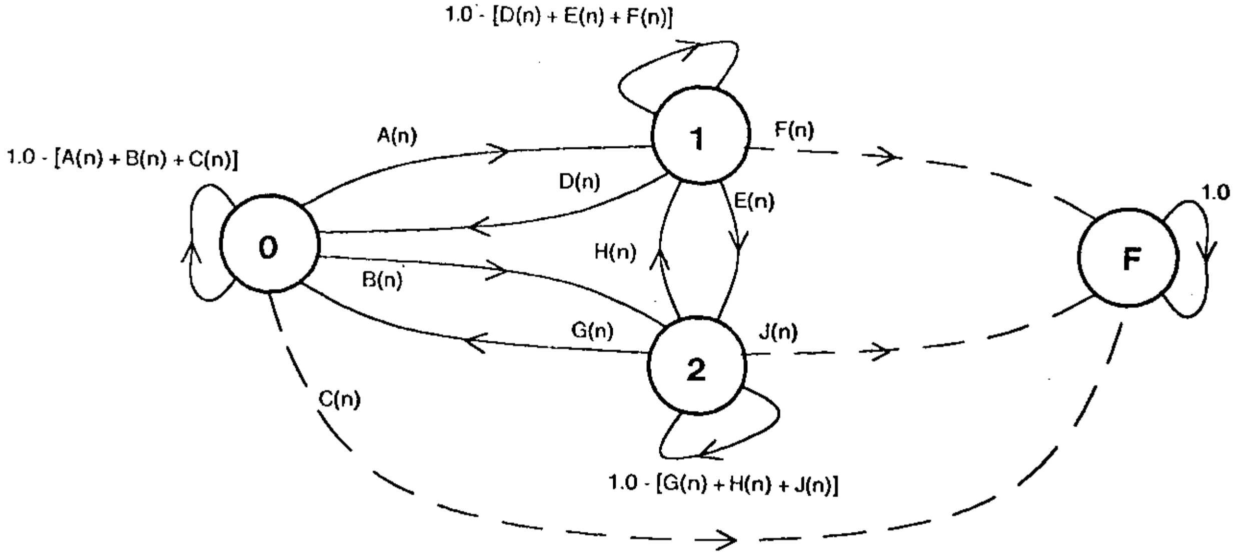
$$r(t) = \alpha_r \lambda_r (\lambda_r t)^{\alpha_r - 1}$$

Figure 4-6: State Diagram of TMR Model With Error Process Renewal

4.4 Hybrid Models

The last type of system modeled is a hybrid redundant structure, with a triple modular redundant core and a single standby replacement module. As in the other models without error process renewal, the multiple states with one module crashed, and with two modules crashed but the spare switched in, have been merged into single states. This results in the model shown in Figure 4-7. As with the other models, the states are labeled with the number of erring modules. State 3 is the system failed state. The coverage factor c is the probability that the spare module successfully replaces a module in which an error is detected. (As with the duplex model, c is always equal to 0.99 in this study.) The state transition probabilities are otherwise similarly defined to those for the other system models.

Figure 4-8 displays the state transition diagram of the hybrid system model with error process renewals. The simulation of this system proceeds similarly to that of the duplex and TMR systems.


1.0 - [D(n) + E(n) + F(n)]
1.0 - [A(n) + B(n) + C(n)]
A(n)
1
F(n)
D(n)
E(n)
1.0
0
H(n)
F
B(n)
G(n)
J(n)
2
C(n)
1.0 - [G(n) + H(n) + J(n)]

$$A(n) = (3c+1)[s(n)]^3[1-s(n)]$$
$$B(n) = 3(1-c)[s(n)]^3[1-s(n)] + 3(1+c)[s(n)]^2[1-s(n)]^2$$
$$C(n) = 3(1-c)[s(n)]^2[1-s(n)]^2 + 4s(n)[1-s(n)]^3 + [1-s(n)]^4$$
$$D(n) = [s(n)]^3[1-r(n)]$$
$$E(n) = 3[s(n)]^2[1-s(n)]r(n) + 3s(n)[1-s(n)]^2[1-r(n)]$$
$$F(n) = 3s(n)[1-s(n)]^2r(n) + [1-s(n)]^3$$
$$G(n) = [s(n)]^2[1-r(n)]^2$$
$$H(n) = 2s(n)[1-s(n)][1-r(n)]^2 + 2[s(n)]^2r(n)[1-r(n)]$$
$$J(n) = 2s(n)[1-s(n)][r(n)]^2 + [1-s(n)]^2[r(n)]^2 + 2[1-s(n)]^2r(n)[1-r(n)]$$

$$s(n) = q_e^{(n+1)^{\alpha_c} - n^{\alpha_e}}$$

$$r(n) = q_r^{(n+1)^{\alpha_r} - n^{\alpha_r}}$$

Figure 4-7: State Diagram of Hybrid Model Without Error Process Renewal

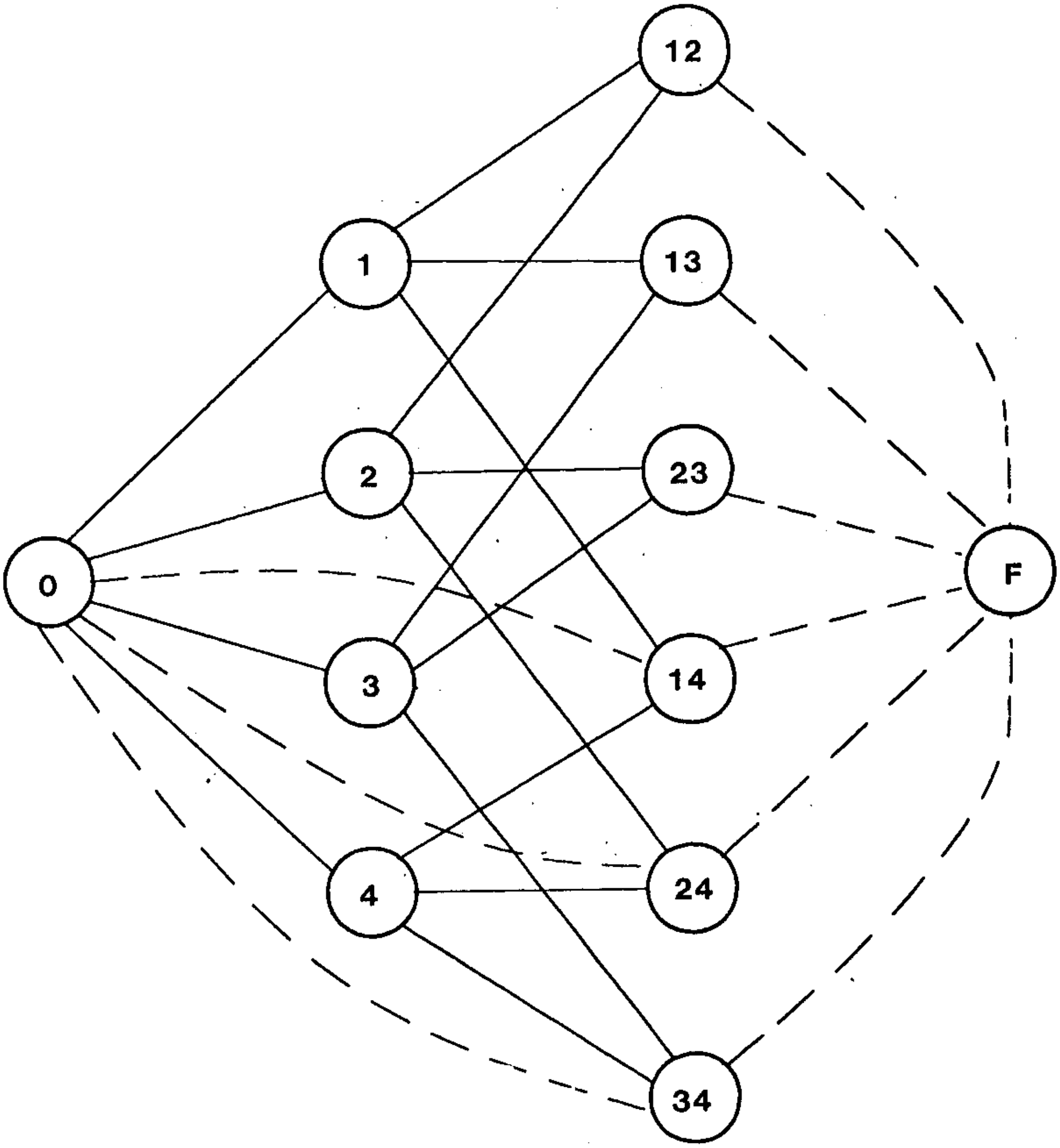


Figure 2-8: State Diagram of Hybrid Model With Error Process Renewal

Chapter 5
Modeling Results and Comparisons

5.1 Parameter Values

For purposes of comparison, error processes of equal means are used throughout. The values of λ are changed along with the values of α to maintain a constant value for the mean of each process. This is done because the mean value of a probabilistic system is one of the commonest statistics used to describe system reliability. For simplicity, recovery processes are assumed to be exponential.

The parameter values for this experiment are chosen to result in a module mean time to error of 100 time steps, and a mean time to recovery for a module of 10 time steps. The corresponding values of q_e for each value of α_e are calculated by finding values of λ_e in the Weibull function which result in the same mean time to error, and then defining $q_e = e^{-(\lambda_e)^{\alpha_e}}$. The values obtained are shown in Table 5-1 below. Their accuracy was checked by performing the summation of Equation (2.9) for several thousand terms. In each case, the sum came to within less than 1% of the desired value of 100.

α_e	λ_e	q_e	α_r	λ_r	q_r	Coverage, c
0.6	0.01505	0.922543	1.0	0.100	0.904837	0.99
0.8	0.01133	0.972624	1.0	0.100	0.904837	0.99
1.0	0.01000	0.990050	1.0	0.100	0.904837	0.99
1.2	0.00941	0.996308	1.0	0.100	0.904837	0.99

Table 5-1: Parameter Values for Experiment

5.2 System Reliability

Figures 5-1 through 5-4 display the reliability curves for each of the four system models under the two different modeling assumptions. The general effect noticed is that systems with a decreasing hazard rate ($\alpha_e < 1$) initially are *less* reliable than the one with a constant hazard rate ($\alpha_e = 1$), but eventually reach a crossover point and become *more* reliable[1]. An opposite effect is evident for the systems with an increasing hazard rate ($\alpha_e > 1$). Error process renewal delays the crossover points to a later time and lower reliability. This is most evident in Figure 5-2.

In order to judge the validity of the system reliability simulations, available analytical model results can be compared to the simulation results using the Kolmogorov-Smirnov goodness of fit test. The simplest model (for the simplex system) also provides a general test of how well the uniform number generator produces pseudo-random numbers following the Weibull distribution. The test results for the simplex system, and also the duplex and TMR systems with α_e equal to one, are given in Table 5-2. All of the tests produce acceptable results. This is encouraging for the studies in the remainder of this paper.

System	α_e	Fit Significance
Simplex	0.6	0.998
Simplex	0.8	0.362
Simplex	1.0	0.235
Simplex	1.2	0.434
Duplex	1.0	0.657
TMR	1.0	0.409

Table 5-2: Test Results for Validating Simulations

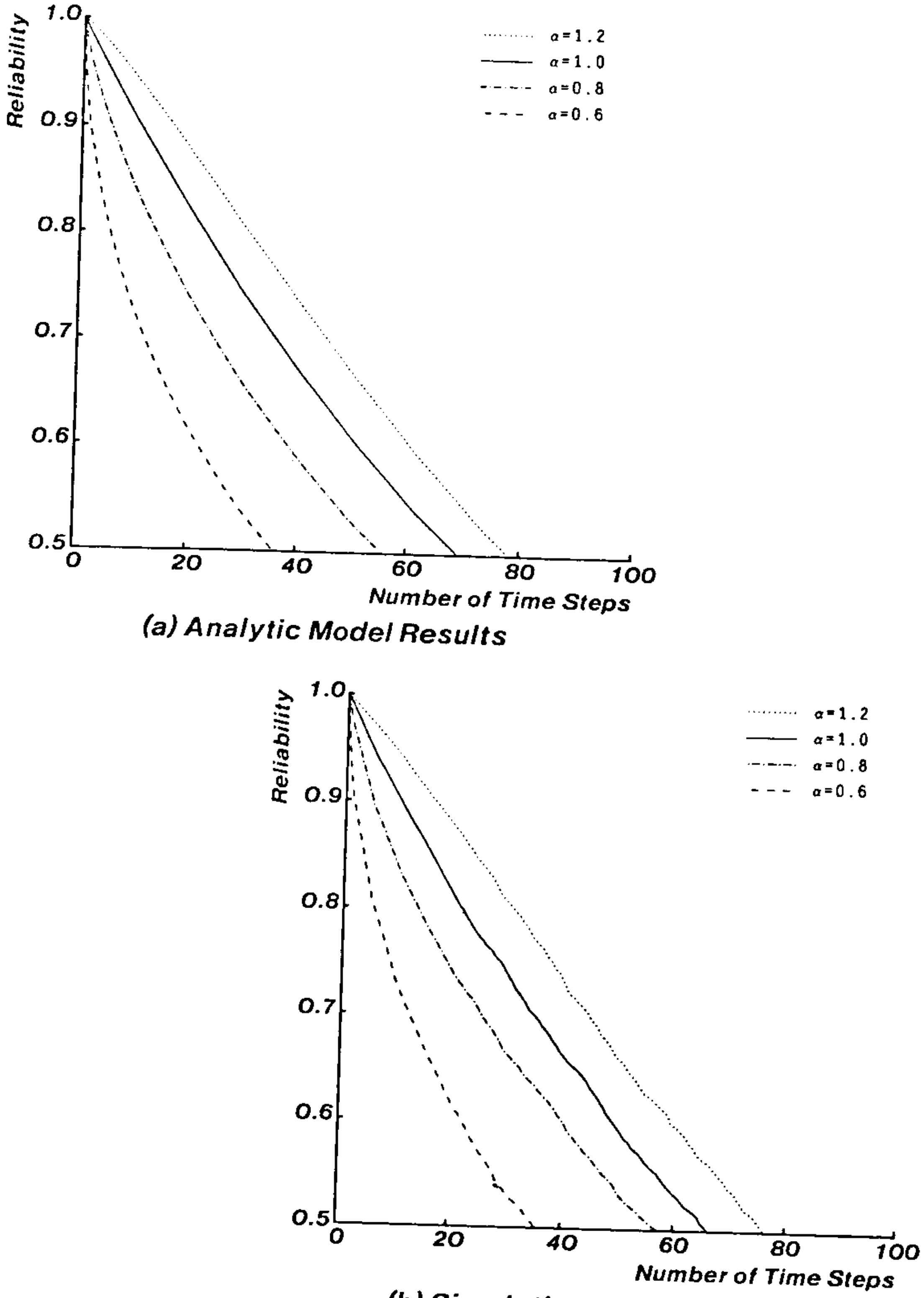
5.3 Reliability Difference Relative to Constant Hazard Rate Systems

Many methods exist for comparing the reliability results of different systems or different models. One of the simplest metrics is the reliability difference function RD(t), which is defined as

$$RD(t) = R(t) - R_b(t)$$

where $R_b(t)$ is a baseline system reliability. The impact of decreasing ($\alpha_e < 1$) or increasing ($\alpha_e > 1$) hazard rates is directly shown in Figures 5-5 through 5-8. Plotted in these figures are the reliability differences relative to the same system with a constant ($\alpha_e = 1$) hazard rate. Large differences are seen in these plots, especially for the redundant structures. For regions of high system reliability, these deviations can be as great as 30-40%. For the redundant structures in regions of lower reliability, the deviation can be several orders of magnitude if the decreasing hazard rate systems without error process renewal are compared to the constant hazard rate system. For the simplex system, and for the redundant systems with error process renewal, the largest effects of nonconstant

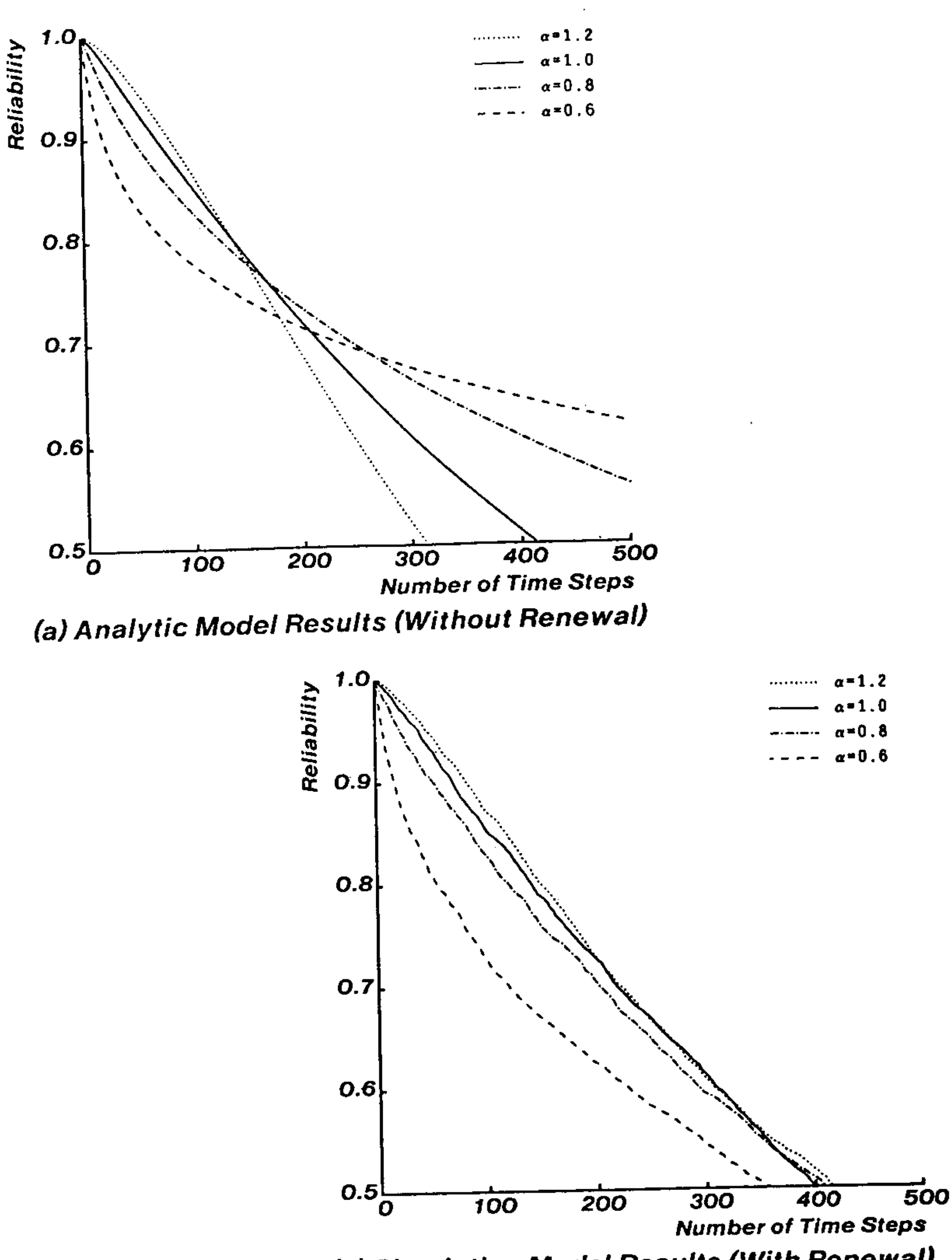
[1] The crossover points for some systems are not shown because they occur at reliability levels below 0.5.



(a) Analytic Model Results

(b) Simulation Model Results

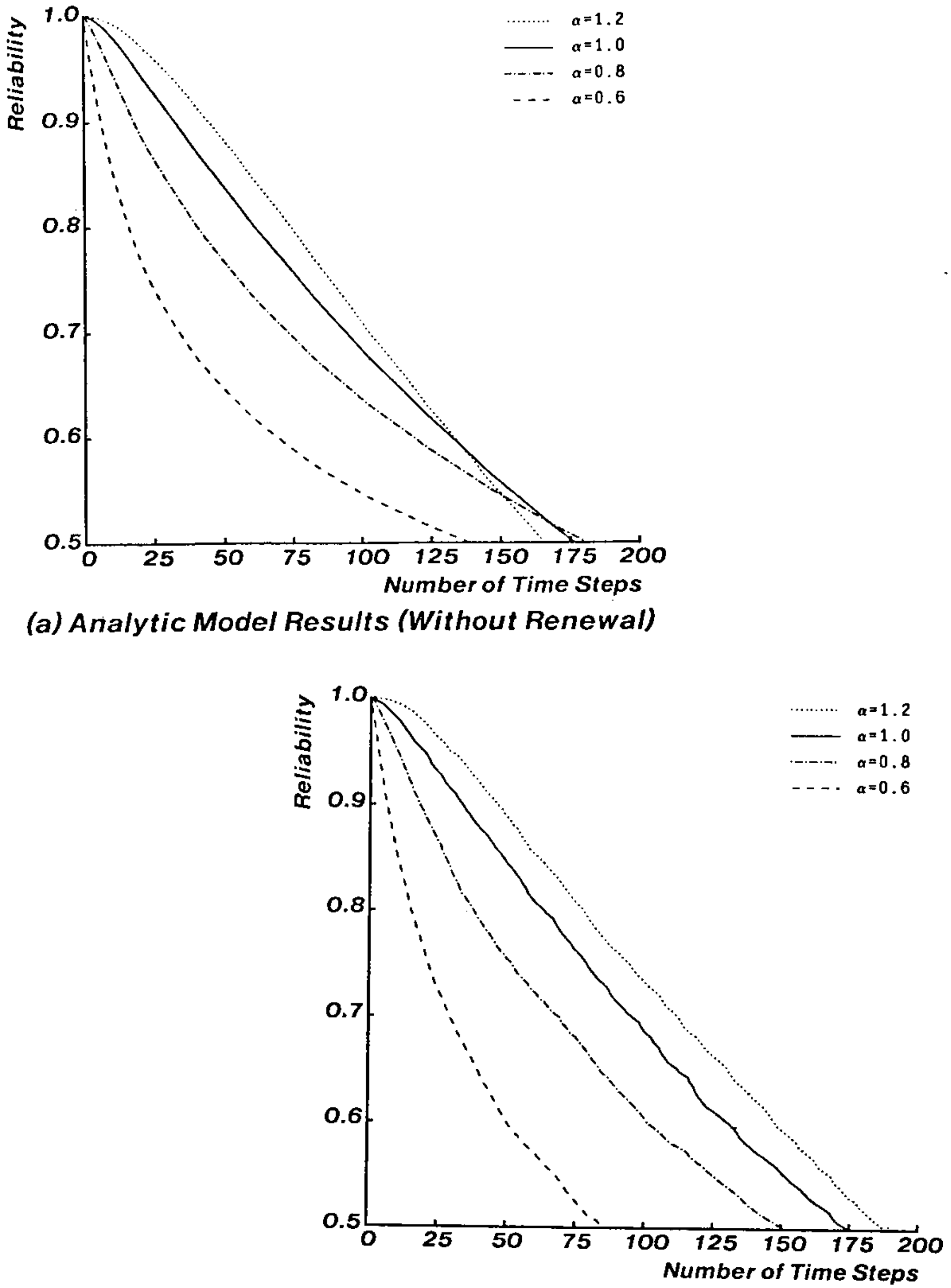
Figure 5-1: Simplex System Reliability



(a) Analytic Model Results (Without Renewal)

(b) Simulation Model Results (With Renewal)

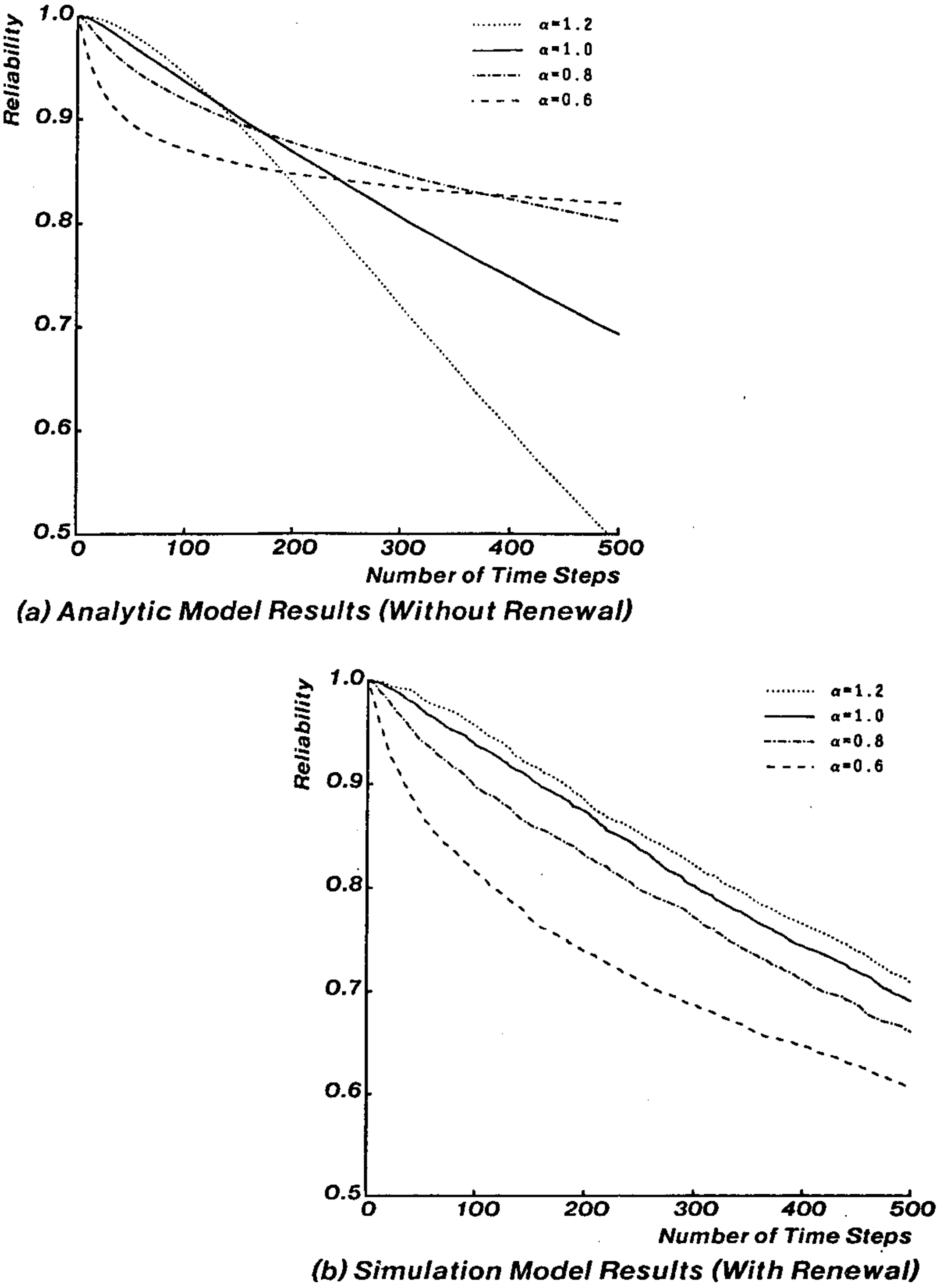
Figure 5-2: Duplex System Reliability



(a) Analytic Model Results (Without Renewal)

(b) Simulation Model Results (With Renewal)

Figure 5-3: TMR System Reliability



(a) Analytic Model Results (Without Renewal)

(b) Simulation Model Results (With Renewal)

Figure 5-4: Hybrid System Reliability

hazard rates occur with overestimation of reliability for decreasing hazard rate systems and underestimation of reliability for increasing hazard rate systems.

Comparing the results of the two types of redundant system models, in general for the the models with error process renewal, the initial differences are slightly greater while the later differences are much smaller. For example, the TMR system without error process renewal has maximum reliability differences of about -0.20 and +0.22 for $\alpha_e = 0.6$. The same system and parameter values with error process renewals show maximum differences of roughly -0.24 and +0.02.

Something which is obvious from all the figures shown in this section thus far is that the magnitude of the reliability difference is directly related to the distance α_e is from one. For example, in Figure 5-7, the reliability difference for $\alpha_e = 0.8$ ranges between -0.08 and +0.11; while for $\alpha_e = 0.6$, it ranges between -0.20 and +0.22. These figures are for the TMR model without error process renewal. Similar results hold for the other models, including those with error process renewal.

5.4 Mission Time Improvement Relative to Constant Hazard Rate Systems

For systems with stringent reliability requirements, the mission time function MT(r) is often used. The relationship between reliability R(t) and mission time MT(r) is given by

$$R[MT(r)] = r \, . \, MT[R(t)] = t$$

For comparing two systems, the mission time improvement MTI(r) is defined as

$$MTI(r) = \frac{MT(r)}{MT_b(r)}$$

where $MT_b(r)$ is the baseline system mission time. Figures 5-9 through 5-12 plot mission time improvement for the different systems. In each case, the baseline system is the corresponding structure with $\alpha_e = 1.0$ (constant hazard rate).

The data used to generate the curves of Figure 5-9 was calculated by inverting the Weibull reliability function to find the values of t corresponding to the desired reliability levels. The curves for the other systems without error process renewal are based on the reliability curve points, using linear interpolation to obtain non-integer mission times. The mission times on which the error process renewal model curves are were derived by sorting the generated simulation times, and taking the kth entry, for the reliability level equal to (1 - k/3000).

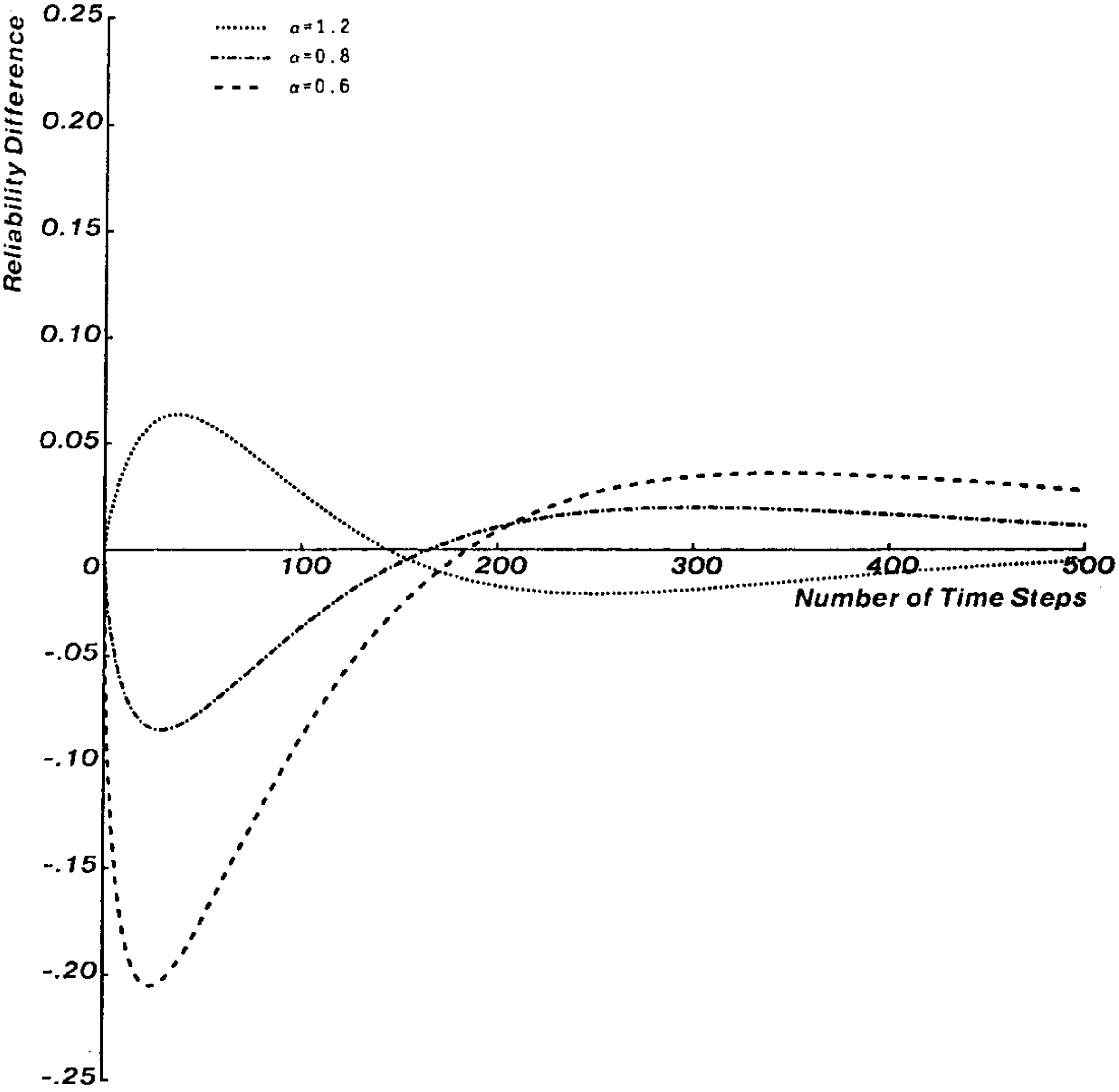


Figure 5-5: Reliability Difference of Simplex System Relative to $\alpha_c = 1.0$

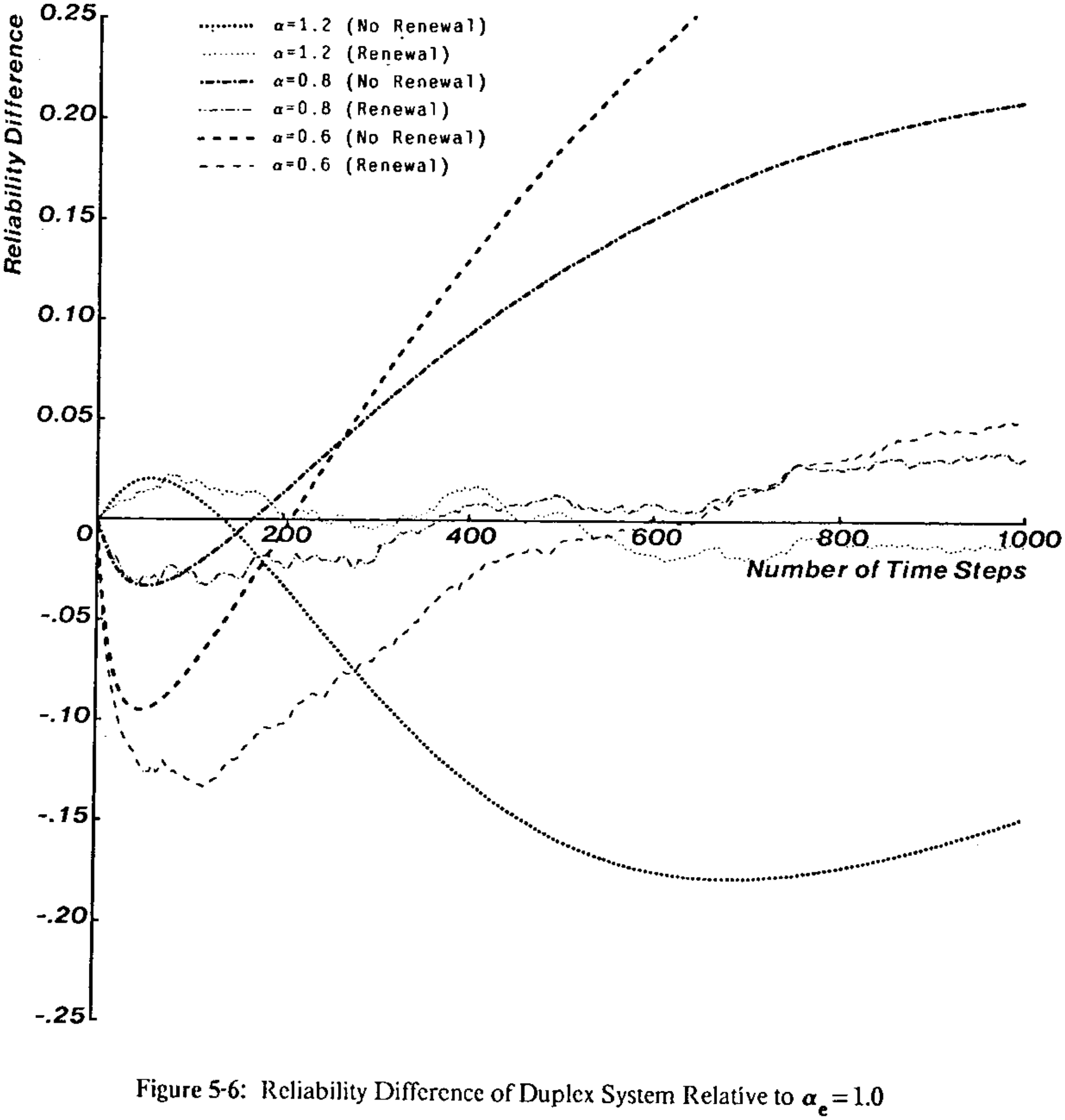


Figure 5-6: Reliability Difference of Duplex System Relative to $\alpha_e = 1.0$

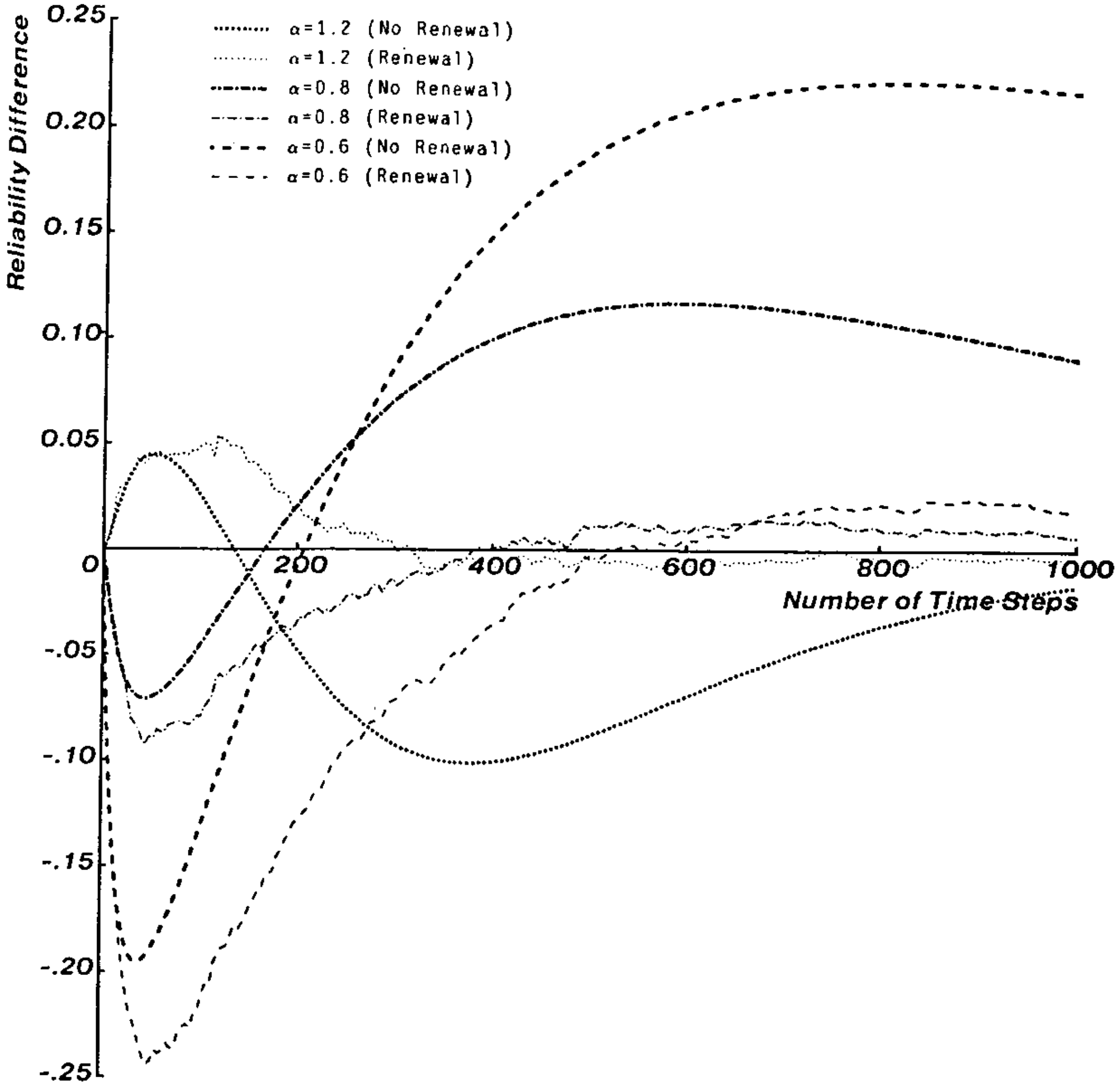


Figure 5-7: Reliability Difference of TMR System Relative to $\alpha_e = 1.0$

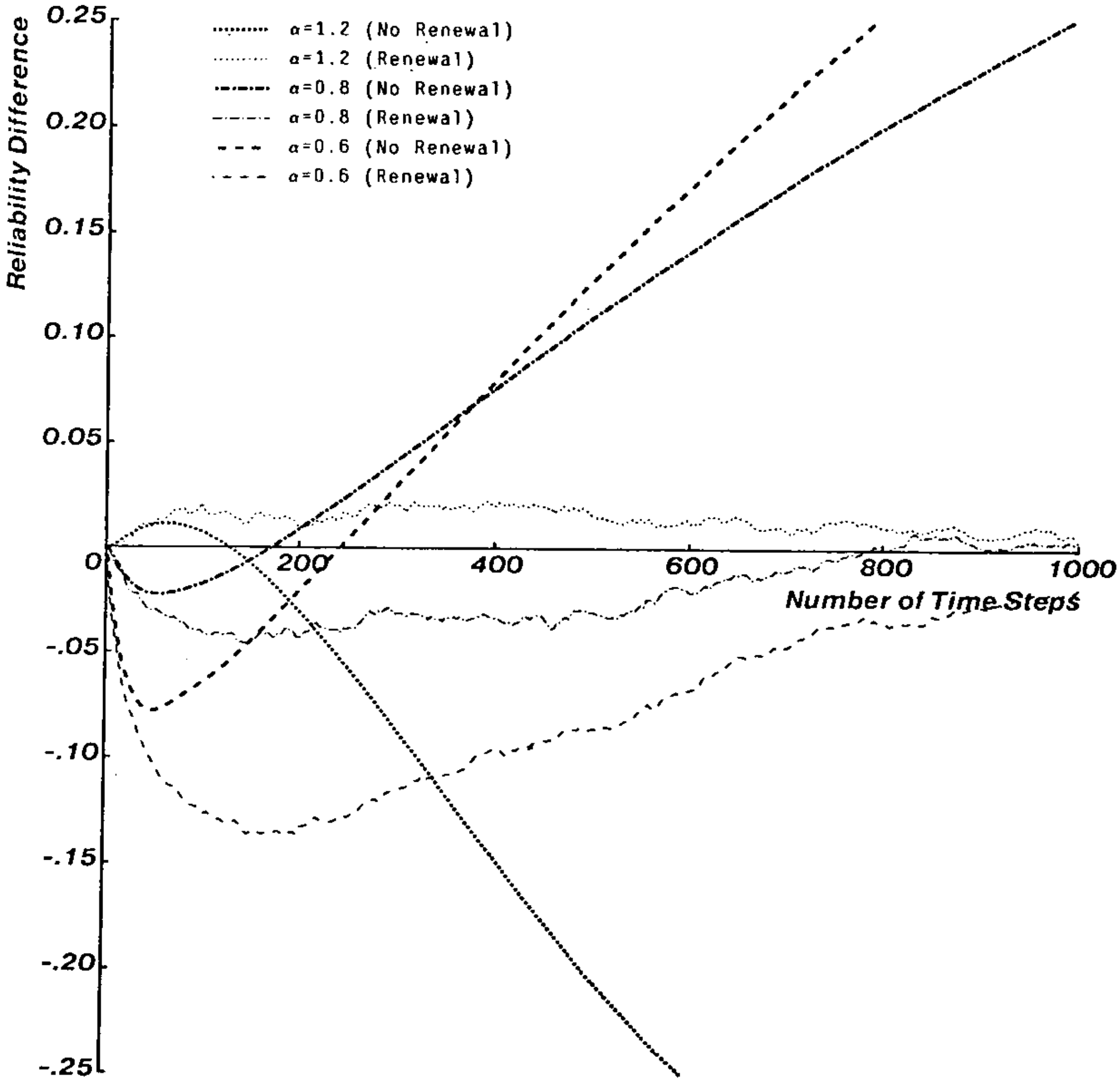


Figure 5-8: Reliability Difference of Hybrid System Relative to $\alpha_e = 1.0$

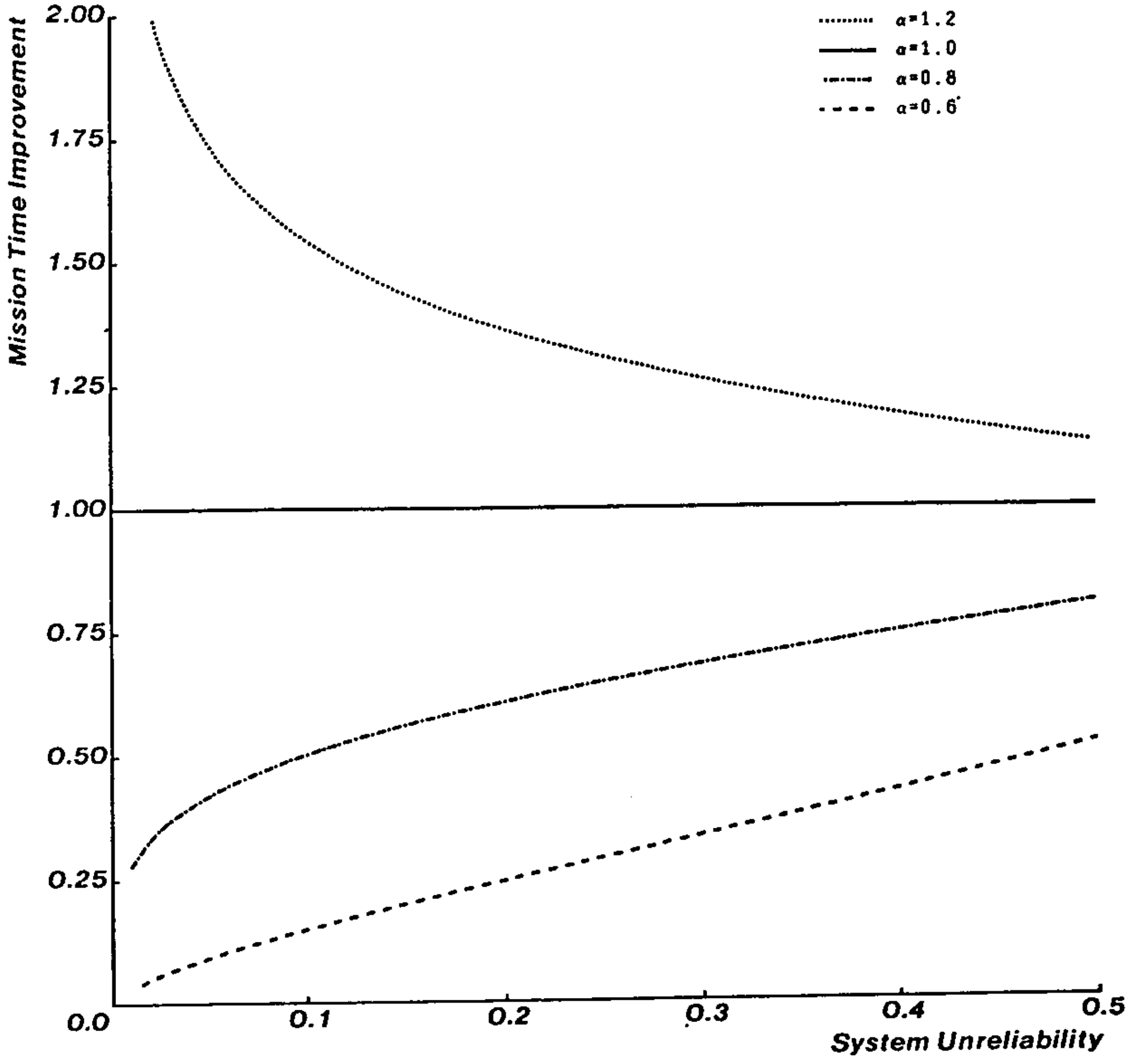


Figure 5-9: Mission Time Improvement of Simplex System Relative to $\alpha_e=1.0$

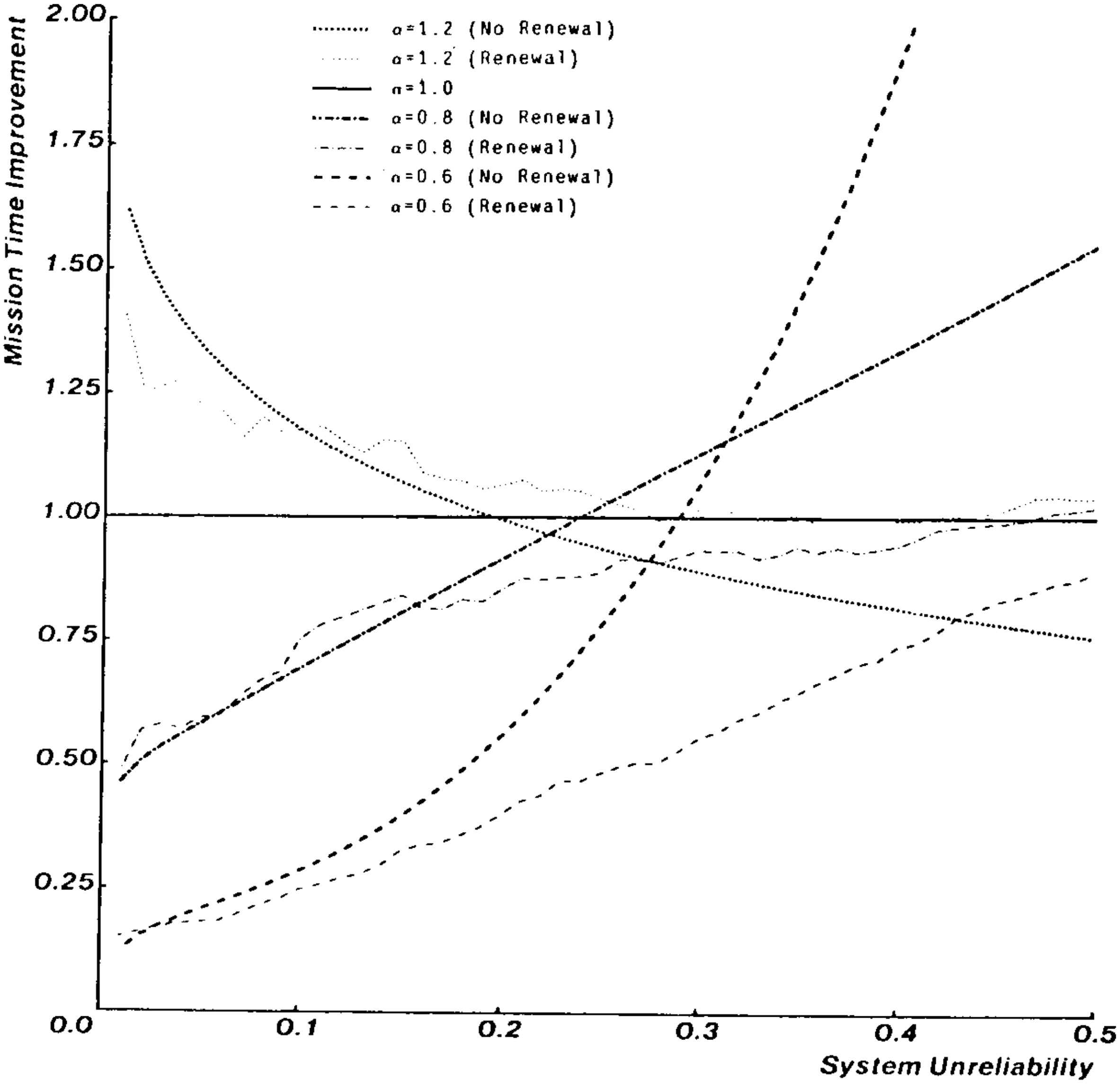


Figure 5-10: Mission Time Improvement of Duplex System Relative to $\alpha_c = 1.0$

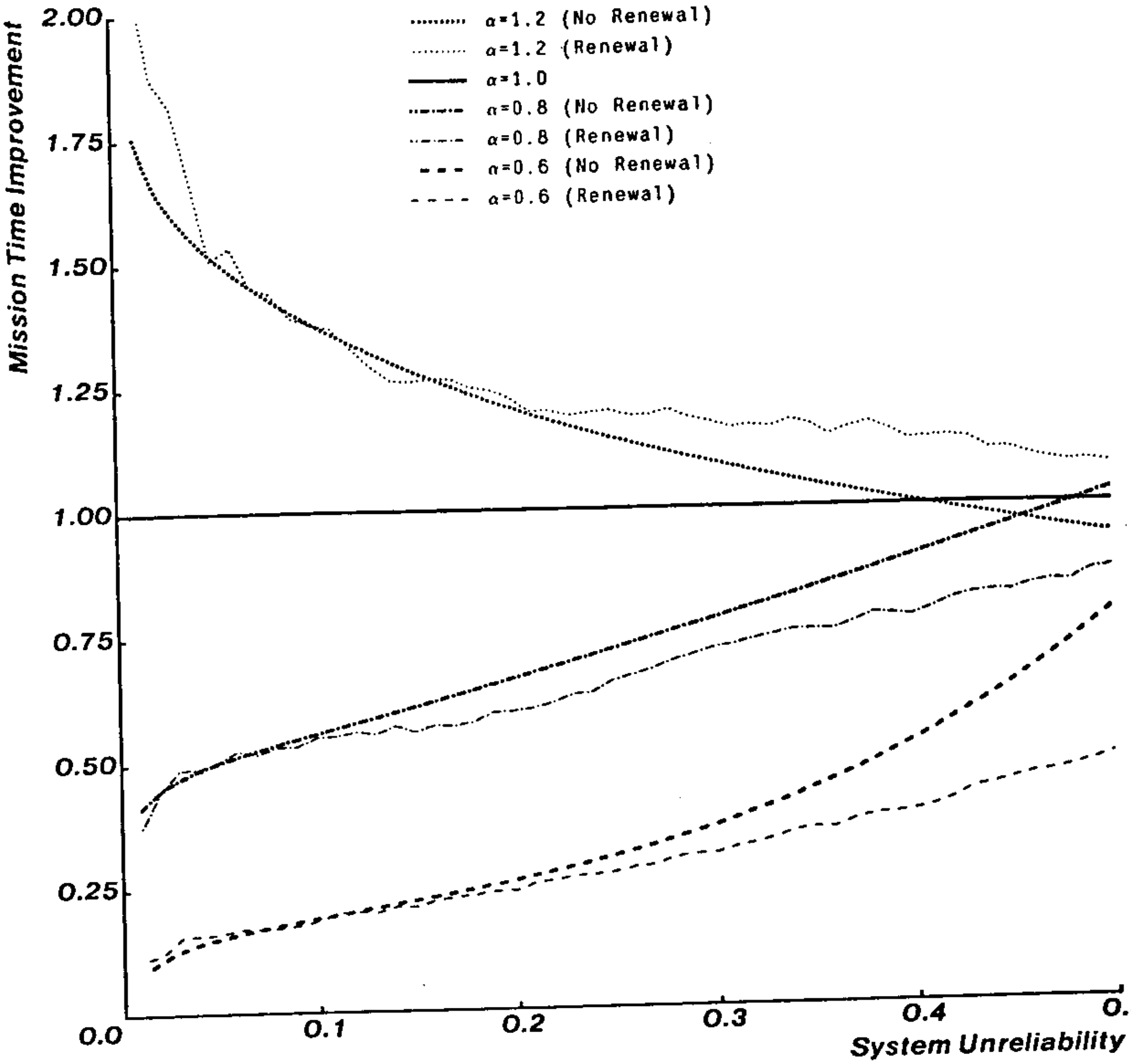


Figure 5-11: Mission Time Improvement of TMR System Relative to $\alpha_e = 1.0$

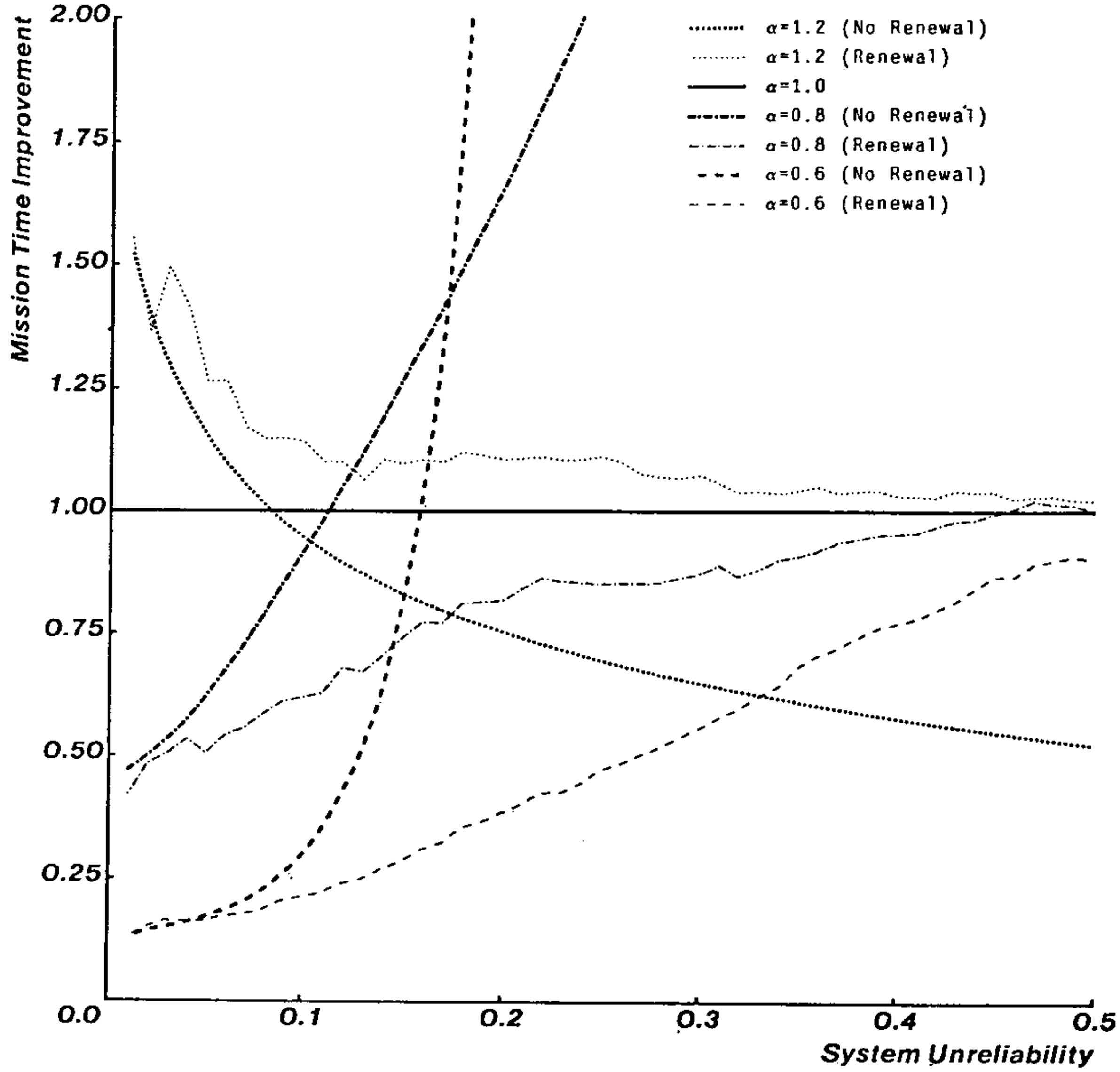


Figure 5-12: Mission Time Improvement of Hybrid System Relative to $\alpha_e = 1.0$

The trends observable in the mission time improvement plots are what would be expected from examining the reliability graphs. For $\alpha_e < 1$ (i.e., decreasing hazard rates), the mission time improvement relative to the constant hazard rate starts out very small for high reliability levels, and increases monotonically as the reliability goes down. An opposite effect occurs for increasing hazard rates. The important thing to note is that for both redundant and nonredundant systems, the obtainable mission times for high mission reliabilities are *much* smaller for $\alpha_e < 1$ than for $\alpha_e \geq 1$. This is true even for redundant structures and for α_e only a little less than one.

5.5 Reliability Difference Relative to Simplex System

In addition to the direct effect on predicted reliability of decreasing ($\alpha_e < 1$) or increasing ($\alpha_e > 1$) hazard rates, the indirect effect on the reliability difference between redundant and simplex systems is also of interest. Figures 5-13 through 5-15 display the reliability differences between the redundant systems and the simplex system. The range of values for α_e given earlier is shown for each of the three redundant structures: duplex, TMR, and hybrid. For each curve plotted in these graphs, the baseline system is the simplex system with the same value of α_e.

For the system models without error process renewal, the interesting feature of these graphs is that changing the shape parameter does not significantly affect the peak magnitude of the reliability difference for the range of parameter values shown. The main effect seems to be that the smaller the value of α_e, the slower the decline from the peak reliability difference. This is true for all three types of redundant structure. On the other hand, for the system models *with* error process renewal, the peak magnitude of the reliability difference increases noticeably with increasing values for α_e, but the reliability differences all decline quickly from that peak value to converge to a relatively small range of values.

An important fact gleaned from Figures 5-13 through 5-15 is that, for every value of α_e, the hybrid redundant system has the highest reliability, followed by the duplex system and the TMR system respectively. Thus, the relative improvement in reliability due to one redundancy technique compared to another is not sensitive to the value of α_e. This should be good news for system designers.

The reason that the TMR system, with three modules, is less reliable than the duplex system, with only two modules, proceeds as follows. Both the duplex and TMR systems can survive only one module crash. In fact, the duplex system has a small probability (1% throughout this thesis) of not

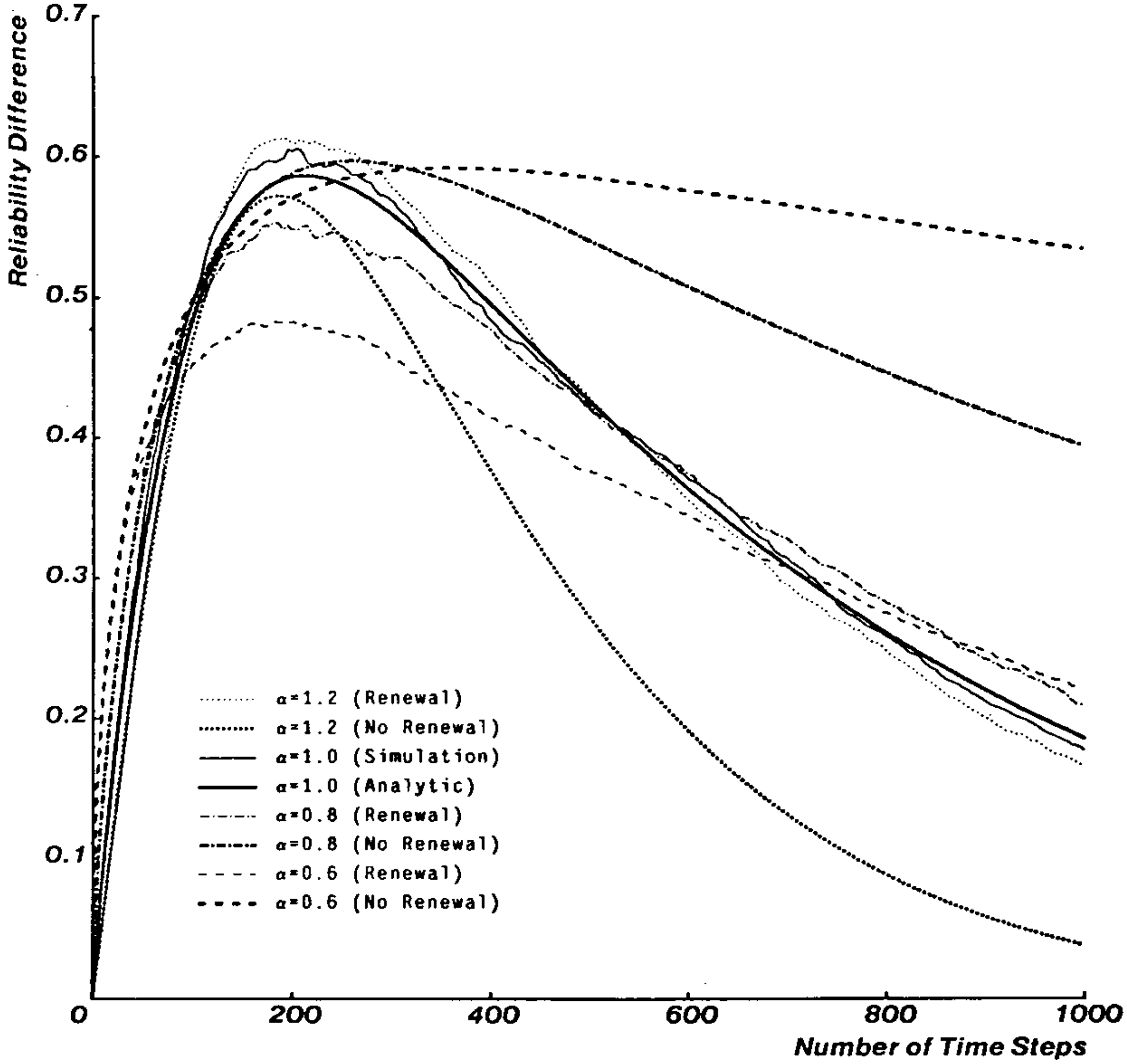


Figure 5-13: Reliability Difference Comparing Duplex to Simplex System

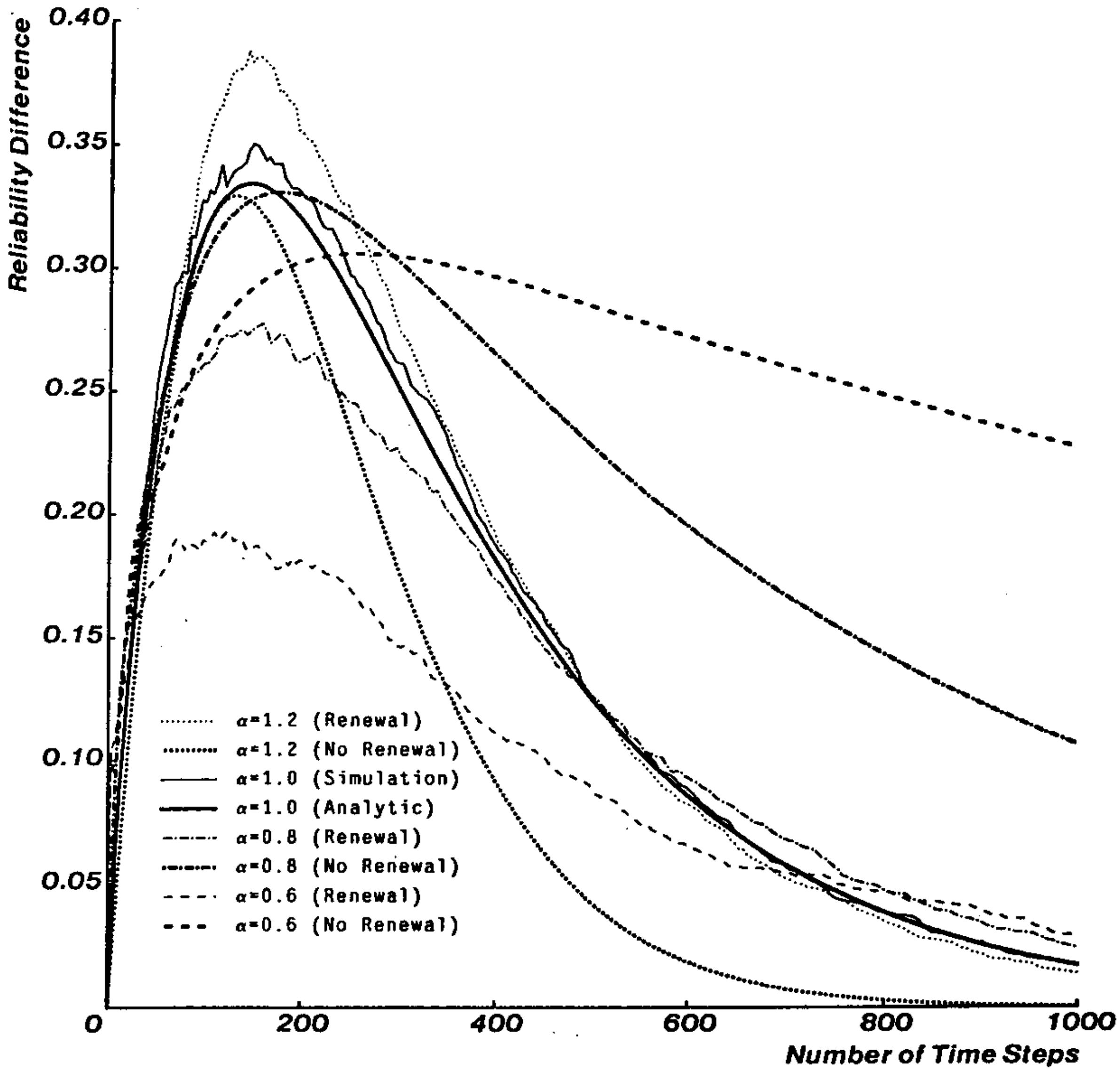


Figure 5-14: Reliability Difference Comparing TMR to Simplex System

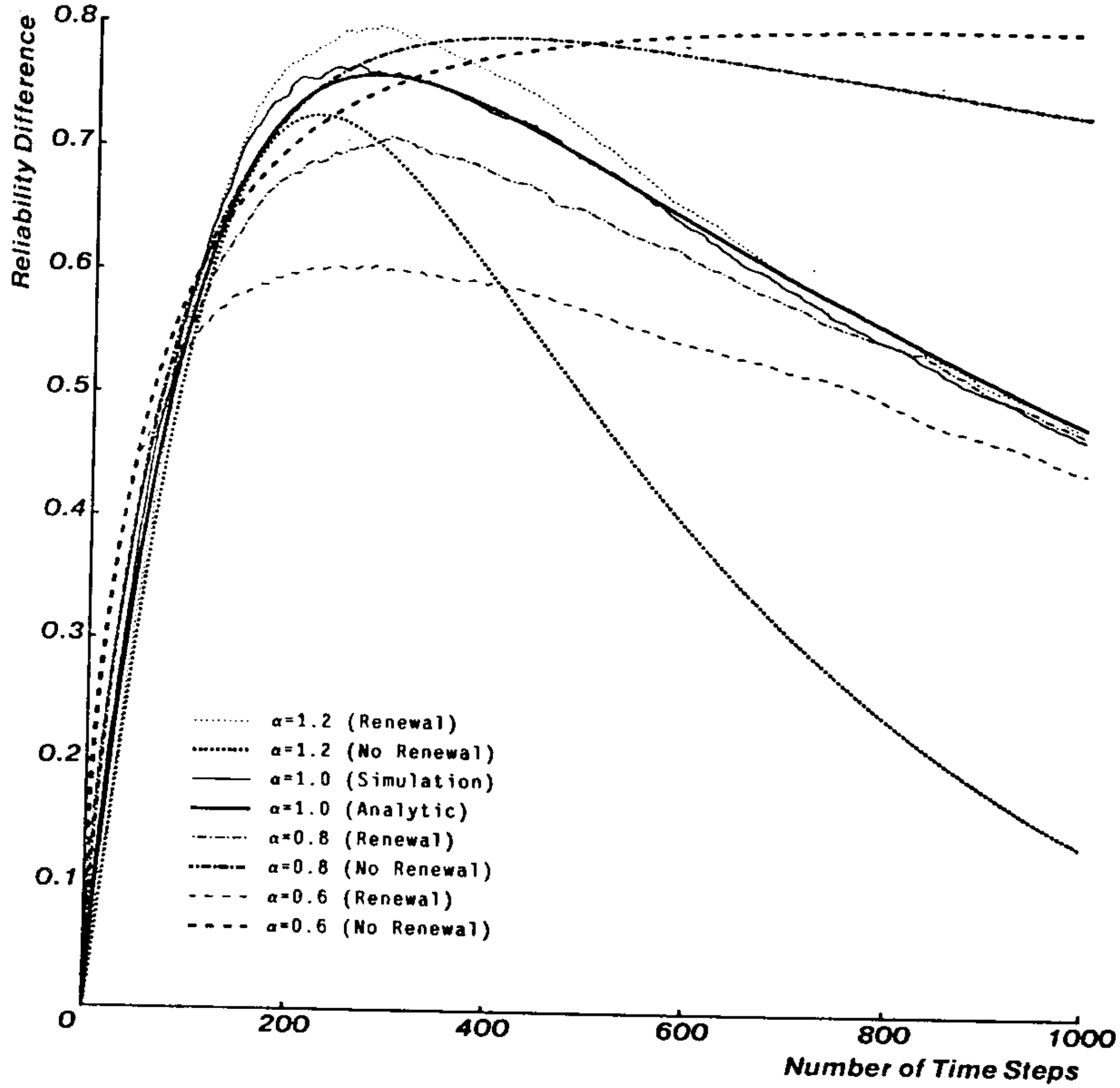


Figure 5-15: Reliability Difference Comparing Hybrid to Simplex System

surviving even one module error. Consider therefore the situation of fully operational systems. The duplex system has two modules contributing to the system error rate while the TMR system has three. One would expect therefore that the TMR system would have an average system error rate 50% greater than the duplex system. This imbalance is even larger for the situation with one module crashed: the TMR system still has two modules contributing to the system error rate while the duplex system has only one. Thus one would expect the TMR system to average twice the error rate of the duplex system, when each system has one module already crashed. The coverage factor (probability of successfully detecting an error and reconfiguring the system) for the duplex system directly affects the comparative system error rates of the the duplex and TMR systems when all modules are operational. However, the TMR structure is inherently much less reliable when one module has already crashed.

5.6 Mission Time Improvement Relative to Simplex System

Mission time improvement for each of the three redundant systems is plotted in Figures 5-16 through 5-18. For these graphs, the baseline for comparison is always the simplex system with the same value of α_e. As would be hoped, the redundant systems all exhibit much greater mission times than the simplex system.

Figures 5-16 through 5-18 show very different patterns for the system models with and without error process renewal. For the models without error process renewal, the mission time improvement decreases monotonically for $\alpha_e \geq 1$; and for $\alpha_e < 1$, the mission time improvement falls to a minimum and then starts rising again. Opposite trends are apparent (but much less definite) for the system models with error process renewal: the mission time improvement is generally decreasing with reliability for $\alpha_e \leq 1$, but has an inflection point with a local minimum for $\alpha_e > 1$.

One aspect of all these mission time improvement curves holds true regardless of system structure or error process renewal. For high levels of mission reliability, the mission time improvement increases as the value of the shape parameter α_e decreases. Thus, even though the absolute mission time attainable for a system with $\alpha_e < 1$ may be very small, the relative gain achieved by redundancy is still very much worthwhile.

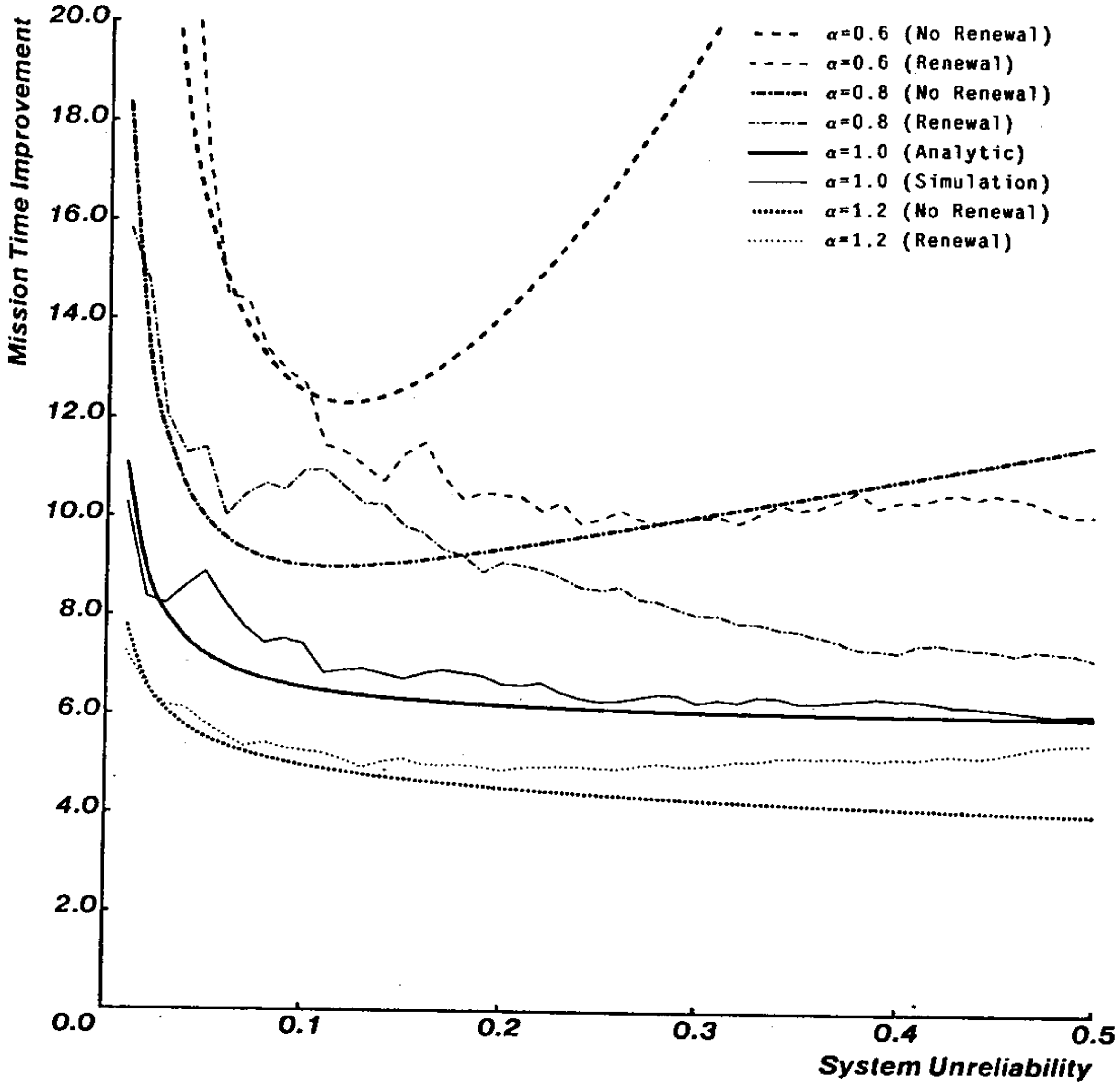


Figure 5-16: Mission Time Improvement Comparing Duplex to Simplex System

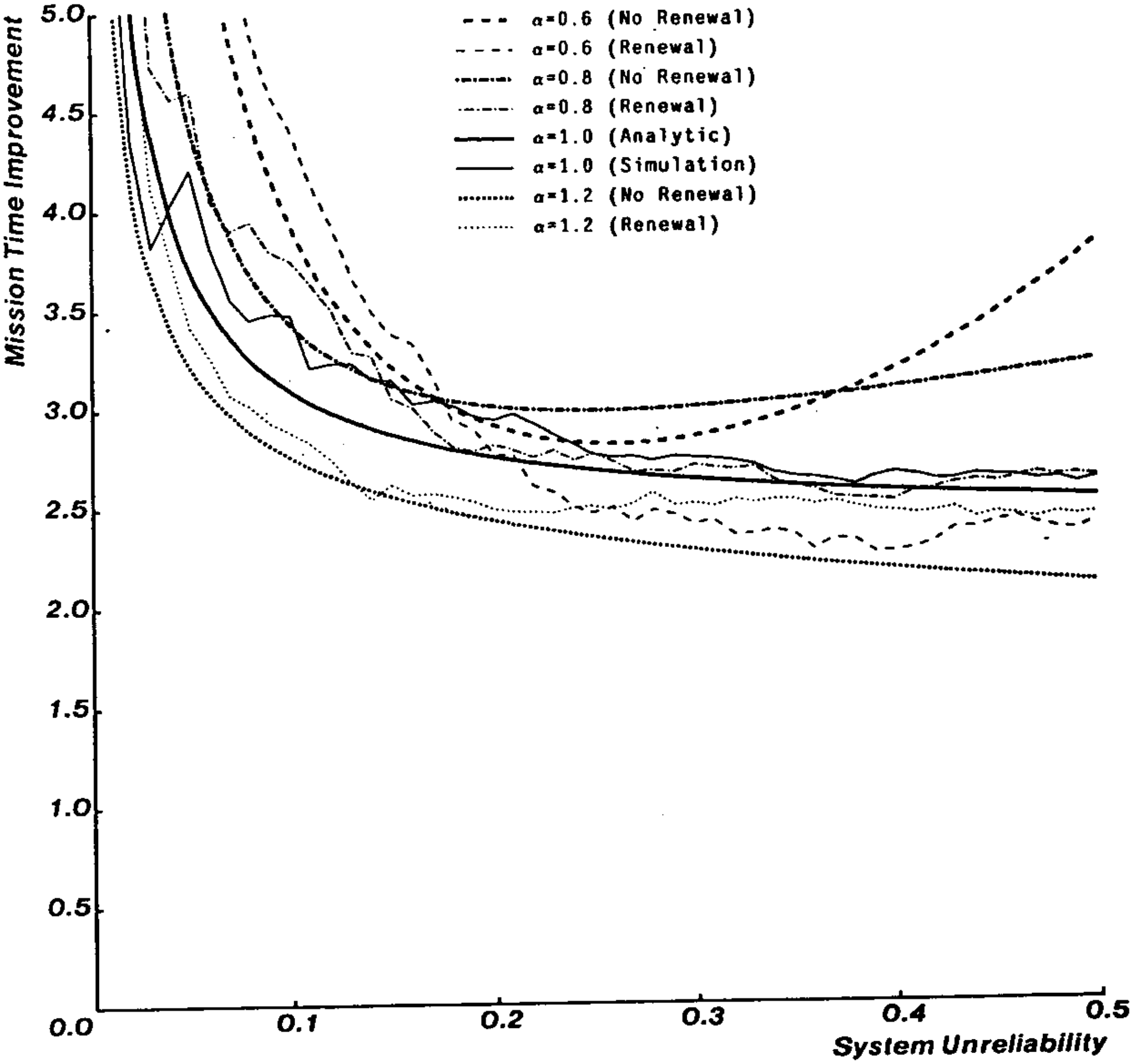


Figure 5-17: Mission Time Improvement Comparing TMR to Simplex System

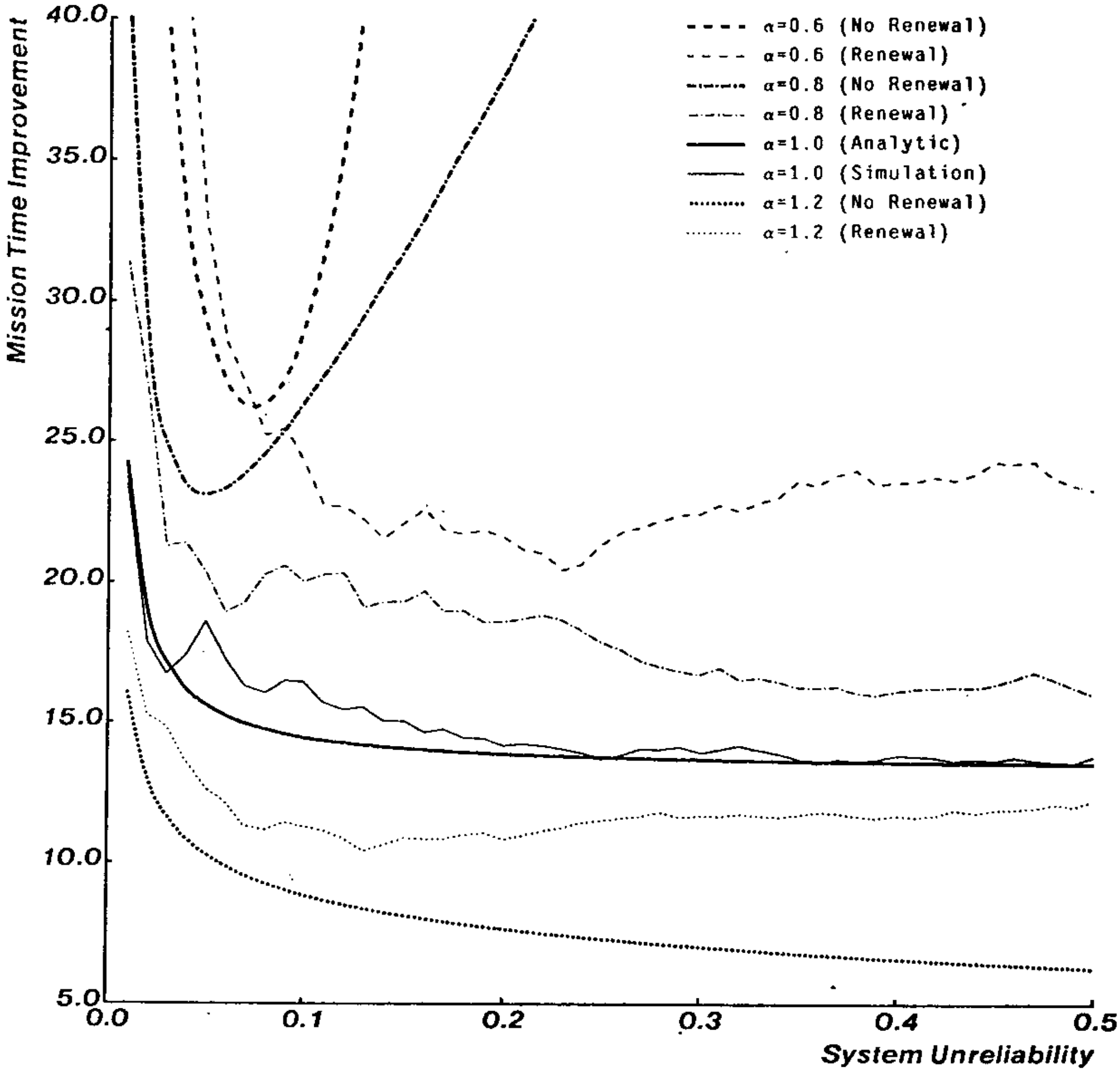


Figure 5-18: Mission Time Improvement Comparing Hybrid to Simplex System

Chapter 6
Conclusions and Comments

The focus of this paper has been a study of transient errors in digital computers. Data collected from several different systems has been analyzed with a variety of parameter estimation techniques and goodness of fit tests. In every case, the data showed a much better fit to a decreasing hazard rate Weibull distribution than to the constant hazard rate (exponential) distribution. Maximum likelihood estimates of the Weibull distribution shape parameter for the collected sets of data range between 0.6 and 0.8. Any value for the shape parameter less than one indicates a decreasing hazard rate. Thus, it is only reasonable to conclude that transient errors follow a decreasing hazard rate distribution rather than the constant hazard rate distribution usually assumed in the past.

Reliability models of both redundant and nonredundant systems have been developed using the decreasing hazard rate Weibull distribution for the module error processes. Constant and increasing hazard rate distributions were used as well for purposes of comparison. An initial set of transient error reliability models was developed under the assumption that the hazard rates decrease (or increase) monotonically from the time that the system begins operation. Discrete time Markov processes were used to solve these models for system reliability. A change in assumptions produced a second set of models. This second set assumes that module error processes are *renewed* (reset to time zero) whenever a module recovers from a transient error. This assumption is a more realistic basis for extending the nonredundant system results to redundant system models. Unfortunately, this requires the use of Monte Carlo simulations to obtain approximate solutions for system reliability. It is more realistic because the data analysis (based primarily on nonredundant systems) follows the assumption that the error process is renewed at the time of recovery.

Large variations are noted in the reliability differences and mission time improvements resulting from relatively small changes in the value of the shape parameter α_e. This is true whether the comparisons are made with respect to the same structure having a constant hazard rate ($\alpha_e = 1$), or with respect to the simplex system having the same value for the shape parameter. Consider the duplex system model without error process renewal, for $\alpha_e = 0.8$ (decreasing hazard rate). Comparing the reliability of this system with that of the same system for $\alpha_e = 1.0$ (constant hazard rate), the 20% change in the value of α_e produces a maximum reliability difference of about 0.22, and a 50% decrease in mission time for a mission reliability of 0.99. Comparing the two duplex system models ($\alpha_e = 0.8$ and $\alpha_e = 1.0$) to the corresponding simplex system models, the reliability differences are very close for the first 200 time steps (twice the module mean time to error), after which they diverge until, at 1000 time steps, the decreasing hazard rate ($\alpha_e = 0.8$) system shows twice as large a reliability difference as does the constant hazard rate system. For a mission reliability of 0.99, the decreasing hazard rate duplex system shows a 65% increase in mission time improvement over the the constant hazard rate duplex system (both systems being compared to the corresponding simplex system).

The models with error process renewal produce somewhat different results than those without error process renewal. In comparing the former to the latter, the reliabilities of nonconstant hazard rate systems converge much more quickly to the reliability of the corresponding constant hazard rate system. Despite this long term convergence, the initial peak magnitude for the reliability difference due to changes in the value of the shape parameter appears to increase somewhat for the error process renewal assumption. Both of these tendencies can be explained by remembering that the renewal process serves to limit the range of values which the module hazard function can take. In long term averages, the hazard function which is (randomly) renewed periodically resembles a constant hazard rate, albeit with randomly varying fluctuations. Over shorter time intervals, the periodic (random) renewals of time varying hazard functions emphasize the initial short term values of the given hazard function. Thus, the initial high error rate for decreasing hazard rate processes is emphasized even more in the effects on a redundant system model; and the initial low error rate for increasing hazard rate processes is also emphasized in redundant system models. The major impact of varying hazard rates remains the same regardless of error process renewals: attainable mission times for high levels of reliability are severely limited if the error process follows a decreasing hazard rate.

Two lessons for system designers can be drawn from this study of transient errors. First and most important, designers have an additional parameter for sensitivity analysis: the shape parameter α_e. This is perhaps best illustrated by an example.

Consider a nonredundant system with a mean time to transient error of ten hours and a six minute mean time to recovery from same. Assume that a mission time of sixteen hours is desired, with a 99%

probability of success. Table 6-1 shows the mission reliabilities for three sets of systems: $\alpha_e = 1.0$, $\alpha_e = 0.8$, and $\alpha_e = 0.6$. Note that any of the redundant systems meet the desired mission goals for $\alpha_e = 1.0$, but that only the hybrid system achieves the desired reliability for $\alpha_e < 1.0$. (If rounding off to two decimal places is allowed, the duplex system barely meets the mission reliability requirements for $\alpha_e = 0.8$.) This highlights the utility of redundancy with spares for dealing with transient errors, which leads to the next point.

System	$\alpha_e = 1.0$	$\alpha_e = 0.8$	$\alpha_e = 0.6$
Simplex	0.8521	0.7748	0.6534
Duplex	0.9940	0.9878	0.9708
TMR	0.9914	0.9788	0.9396
Hybrid	0.9998	0.9992	0.9958

Table 6-1: Mission Reliabilities for Design Example

The second lesson for system designers is that, with transient errors being a prominent cause of system failure, it is best to design systems with error *recovery* in mind. For instance, a system which switches in a spare module when a failure is apparently found should be designed such that the module which is switched out is added to the pool of spares (with lowest priority). This allows transient errors with massive effects to be flushed out of the system without unnecessarily discarding hardware which is undamaged. All models in this paper assume a structure which can recycle modules that have been replaced by a spare.

Although much remains to be learned about transient errors in digital computers, a solid foundation has been laid in this research. Analysis tools have been developed (see [McConnel 81]) which can facilitate future studies of transient errors.

References

[Almassy 79] G. Almassy.
Limits of Models in Reliability Engineering.
In *Proceedings, Annual Reliability and Maintainability Symposium*, pages 364-367. IEEE Reliability Society, 1979.

[Avizienis 77] A. Avizienis.
Fault-Tolerant Computing--Progress, Problems, and Prospects.
In B. Gilchrist (editor), *Information Processing 77*, pages 405-420. IFIP, North-Holland Publishing Company, New York, August, 1977.

[Bell et al. 78] C. G. Bell, A. Kotok, T. N. Hastings, and R. Hill.
The Evolution of the DECsystem 10.
Communications of the Association for Computing Machinery 21(1):44-63, January, 1978.

[DEC 78] Digital Equipment Corporation.
TOPS-10 and TOPS-20 SYSERR Manual.
Software Documentation No. AA-D533A-TK.

[DoD 74] Department of Defense.
Military Standardization Handbook: Reliability Prediction of Electronic Equipment.
MIL-STD-HDBK-217B.

[Easterling 76] R. G. Easterling.
Goodness of Fit and Parameter Estimation.
Technometrics 18(1):1-9, February, 1976.

[Fuller & Harbison 78]
S. H. Fuller and S. P. Harbison.
The C.mmp Multiprocessor.
Technical Report CMU-CS-78-146, Carnegie-Mellon University, October, 1978.

[Geilhufe 79] M. Geilhufe.
Soft Errors in Semiconductor Memories.
In *Digest of Papers, Compcon Spring 79*, pages 210-216. IEEE Computer Society, 1979.

[Howard 71] R. A. Howard.
Dynamic Probabilistic Systems (2 vols.).
John Wiley & Sons, Inc., New York, 1971.

[Knuth 69] D. E. Knuth.
The Art of Computer Programming, Volume 2: Seminumerical Algorithms.
Addison-Wesley Publishing Company, Reading, Massachusetts, 1969.

[McConnel et al. 79]
S. R. McConnel, D. P. Siewiorek, and M. M. Tsao.
The Measurement and Analysis of Transient Errors in Digital Computer Systems.
In *Digest of Papers, Ninth Annual International Symposium on Fault-Tolerant Computing*, pages 67-70. IEEE Computer Society, 1979.

[McConnel 81] S. R. McConnel.
Analysis and Modeling of Transient Errors in Digital Computers.
PhD thesis, Carnegie-Mellon University, June, 1981.

[Miller & Freund 65]
I. Miller and J. Freund.
Probability and Statistics for Engineers.
Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965.

[Morganti et al. 78]
M. Morganti, G. Coppadoro, and S. Ceru.
UDET 7116--Common Control for PCM Telephone Exchange: Diagnostic Software Design and Availability Evaluation.
In *Digest of Papers, Eighth Annual International Conference on Fault-Tolerant Computing*, pages 16-23. IEEE Computer Society, 1978.

[Morganti 78] M. Morganti.
private communication.
1978.

[Nakagawa & Osaki 75]
T. Nakagawa and S. Osaki.
The Discrete Weibull Distribution.
IEEE Transactions on Reliability R-24(5):300-301, December, 1975.

[Ohm 79] V. J. Ohm.
Reliability Considerations for Semiconductor Memories.
In *Digest of Papers, Compcon Spring 79*, pages 207-209. IEEE Computer Society, 1979.

[Salvia 79] A. A. Salvia.
Consonance Sets for 2-Parameter Weibull and Exponential Distributions.
IEEE Transactions on Reliability R-28(4):300-302, October, 1979.

[Salvia 80] A. A. Salvia.
Some Fundamental Properties of Kolmogorov-Smirnov Consonance Sets.
Technometrics 22(1):109-111, February, 1980.

[Scelza 77] D. Scelza.
Cm Host User's Manual.*
Technical Report, Carnegie-Mellon University, September, 1977.

[Siewiorek et al. 78a]
D. P. Siewiorek, V. Kini, H. Mashburn, S. R. McConnel, and M. M. Tsao.
A Case Study of C.mmp, Cm*, and C.vmp: Part I -- Experiences with Fault Tolerance in Multiprocessor Systems.
Proceedings of the IEEE 66(10):1178-1199, October, 1978.

[Siewiorek et al. 78b]
D. P. Siewiorek, V. Kini, R. Joobbani, and H. Bellis.
A Case Study of C.mmp, Cm*, and C.vmp: Part II--Predicting and Calibrating Reliability of Multiprocessor Systems.
Proceedings of the IEEE 66(10):1200-1220, October, 1978.

[Spillman 77] R. J. Spillman.
A Markov Model of Intermittent Faults in Digital Systems.
In *Proceedings of the Seventh Annual International Symposium on Fault-Tolerant Computing*, pages 157-161. IEEE Computer Society, 1977.

[Stiffler et al. 79] J. J. Stiffler, L. A. Bryant, and L. Guccione.
CARE III Final Report Phase I (2 Vols.).
NASA Contractor Reports 159122-159123.

[Su et al. 78] S. Y. H. Su, I. Koren, and Y. K. Malaiya.
A Continuous-Parameter Markov Model and Detection Procedures for Intermittent Faults.
IEEE Transactions on Computers C-27(6):567-570, June, 1978.

[Swan et al. 77] R. J. Swan, S. H. Fuller, and D. P. Siewiorek.
Cm*--A modular, multi-microprocessor.
In *AFIPS Conference Proceedings*, pages 637-644. AFIPS, 1977.

[Swan 78] R. J. Swan.
The Switching Structure and Addressing Architecture of an Extensible Multiprocessor: Cm.*
PhD thesis, Carnegie-Mellon University, August, 1978.

[Thoman et al. 69]
D. R. Thoman, L. J. Bain, and C. E. Antle.
Inferences on the Parameters of the Weibull Distribution.
Technometrics 11(3):445-460, August, 1969.

[Tsao 78] M. M. Tsao.
A Study of Transient Errors on Cm*.
Master's thesis, Carnegie-Mellon University, December, 1978.

[Wakerly 75] J. F. Wakerly.
Transient Failures in Triple Modular Redundancy Systems with Sequential Modules.
IEEE Transactions on Computers C-24(5):570-573, May, 1975.

[Yakowitz 77] S. J. Yakowitz.
Computational Probability and Simulation.
Addison-Wesley Publishing Company, Reading, Massachusetts, 1977.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CMU-CS-81-128	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TRANSIENT ERROR RELIABILITY MODELS BASED ON DATA ANALYSIS	5. TYPE OF REPORT & PERIOD COVERED Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephen R. McConnel Daniel P. Siewiorek	8. CONTRACT OR GRANT NUMBER(s) NR-048-645
9. PERFORMING ORGANIZATION NAME AND ADDRESS Carnegie-Mellon University Computer Science Department Pittsburgh, PA. 15213	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE June 1981
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; distribution unlimited

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)