PLG-0206
Revised

RELIABILITY ANALYSIS OF SAFETY INJECTION SYSTEM MODIFICATION SAN ONOFRE NUCLEAR GENERATING STATION — UNIT 1

by
Dennis C. Bley
Lincoln G. H. Sarmanian
Daniel W. Stillwell

Prepared for
SOUTHERN CALIFORNIA EDISON COMPANY
Rosemead, California
October 2, 1981

PICKARD, LOWE AND GARRICK, INC.
CONSULTANTS — ELECTRIC POWER
IRVINE, CALIFORNIA WASHINGTON, D.C.

TABLE OF CONTENTS

LIST OF FIGURES

1. STATEMENT OF PURPOSE

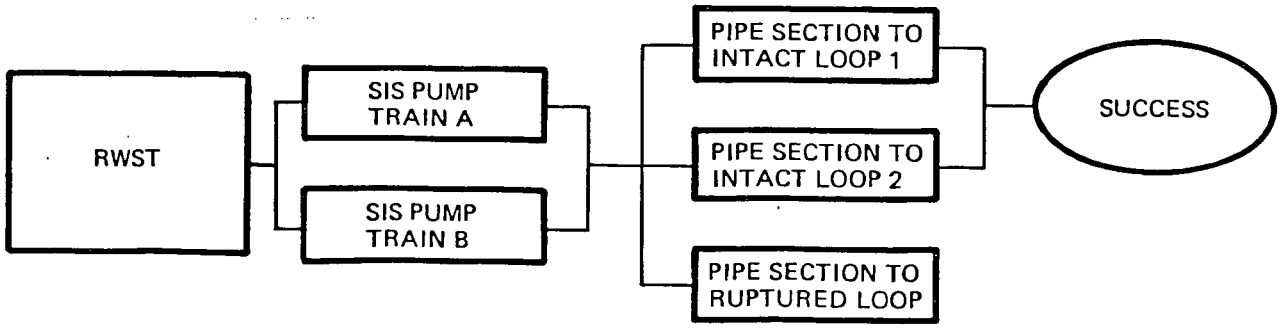
The purpose of this report is to evaluate the reliability of the San Onofre Nuclear Generating Station Unit 1 (SONGS-1) safety injection system (SIS) for the Southern California Edison Company (SCE) following proposed modifications and to consider the effect of SIS reliability in the broader context of plant risk.

2. BACKGROUND

The safety injection system meets the reactor coolant system makeup requirements under a variety of loss of coolant accident (LOCA) conditions up to and including the large LOCA design basis accident (a complete severing of the cold leg piping). The SIS can also serve as a functional backup to the auxiliary feedwater system; bleed and feed cooling to remove decay heat can take the place of secondary cooling via the steam generators.

At SONGS-1, a single safety injection system (see Figure 2-1) supplies makeup water for all breaks greater than about 5/8-inch (smaller leaks can be handled by the normal makeup capability of the charging pumps and are not classed as LOCAs). The system operates in the following manner:

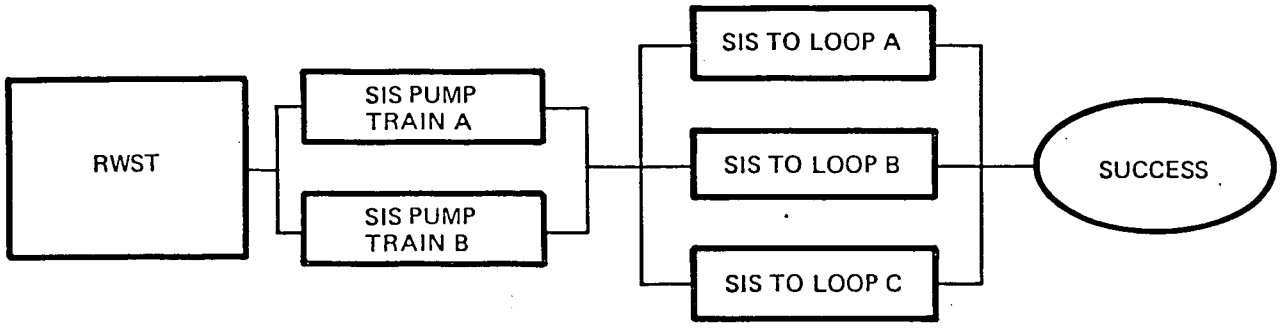
- The two main feedwater pumps are isolated from the main feed system (suction valves HV-854A and HV-854B and discharge valves HV-852A and HV-852B are shut) and are realigned into the SIS (suction valves HV-853A and HV-853B and discharge valves HV-850A and HV850B are opened).
- The two safety injection pumps take suction on the refueling water storage tank (RWST) supplying water separately to each of the main feedwater pumps.
- The two SIS pump trains are cross-connected downstream of the HV-852A and HV-852B valves forming a single header.
- The SIS header supplies makeup water to the cold leg of each reactor coolant system (RCS) loop (A, B, and C) via valves MOV-850A, MOV-850B, and MOV-850C.
- For a large LOCA, injection water must be supplied from one pump train to one of two intact RCS loops, i.e., it is a one-out-of-two-twice arrangement:



Reliability Block Diagram

If the core has operated long enough to generate nearly its maximum decay heat, flow must reach the RCS within about 25 seconds to prevent core damage (Reference 1). If flow is delayed, core damage is expected; but as long as flow arrives within about 30 minutes, core degradation can be arrested, the debris can be contained within the reactor vessel, and severe offsite releases can be prevented (Reference 2).

- For a more likely, very small LOCA (about 2 inches or less), injection water can be supplied from one pump train to any one RCS loop, i.e., it is a one out of two and one out of three arrangement:



Reliability Block Diagram

Furthermore, with no SIS, core damage would not begin for at least 1 hour. Supplying SIS flow within several hours (about 3 to 4) would arrest the sequence in-vessel and prevent serious offsite releases.

The design at SONGS-1 is unique. Most other PWRs have two systems: a low flow, high pressure injection system for small LOCAs and a high flow, low pressure injection system for large LOCAs and long term recirculation cooling. SONGS-1 has the single injection system described above and a separate long term recirculation cooling system. Nevertheless, the functions are similar and comparison with the more standard systems as analyzed in the Reactor Safety Study, WASH-1400 (Reference 3), may be helpful. Point value unavailability estimates for the PWR analyzed in WASH-1400 were:

System	Point Estimate Unavailability in WASH-1400
Low Pressure Injection	3.1×10^{-3}
High Pressure Injection	4.4×10^{-3}

These values included the effects of supporting systems but were dominated by the following effects:

System	Leading Contributors to Unavailability in WASH-1400
Low Pressure Injection	Single valve failures (hardware and human error) and maintenance.
High Pressure Injection	Failures of redundant valves to operate, failure of BIT heaters or BIT piping heat tracing along with failure to defect the heater failure, and single valve failures.

The performance of the SISs in WASH-1400 was considered in the context of demands placed on the SIS by the following accidents:

Initiating Event	Frequency Point Estimate	SIS Requirements
Large LOCA	1×10^{-4}	LPIS and Accumulators
Medium LOCA	3×10^{-4}	LPIS and HPIS
Small LOCA	1×10^{-3}	HPIS

While no LOCA demands for the SIS have occurred at SONGS-1, an actuation signal in response to plant low pressure was generated in September 1981. The SIS should have operated, but it failed. Both sets of feedwater isolation valves closed properly and the RWST was lined up to both main feedwater pumps via the safety injection pumps. The system failed due to a common problem: both 14-inch SIS discharge valves HV-851A and HV-851B failed to open. These double disk gate valves were bound shut by high differential pressure across the downstream disks. Binding problems (due to galling) have been previously identified (References 4, 5, and 6) but were believed solved. The original motor operators had been replaced by faster acting hydraulic operators in 1976 to meet specific licensing requirements (References 7, 8, and 9). Several possible changes to system design and testing are being considered by SCE. They are described and analyzed in Sections 3 and 4 (References 10 and 11).

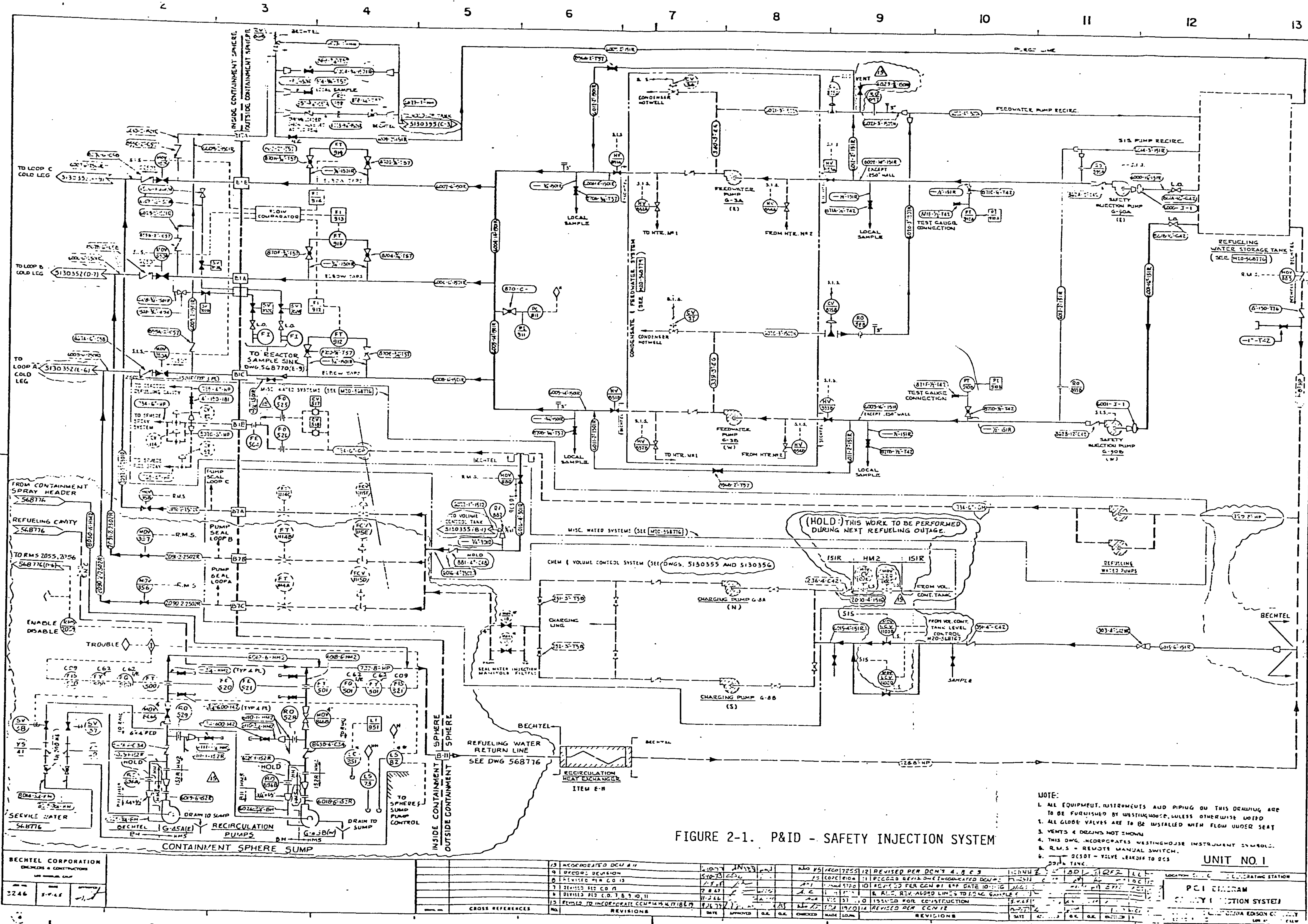


FIGURE 2-1. P&ID - SAFETY INJECTION SYSTEM

3. MODIFICATION OPTIONS

Three modification options are quantified in this analysis.

- Option 1. The preferred option consists basically of a main feedwater pump trip upon receipt of a safety injection signal and an immediate open signal to valves HV-851A and HV-851B. Valves HV-851A and HV-851B are interlocked with their respective main feedwater normal suction valves HV-854A and HV-854B which prevents the safety injection signal from opening HV-851A or HV-851B until HV-854A or HV-854B is fully closed. Eleven seconds later, the main feedwater pumps are started and valves MOV-850A, MOV-850B, and MOV-850C are opened (Reference 10). In addition to the main feed pump trip and restart, a vent is installed which will depressurize the between disc area of HV-851A and HV-851B. This vent line contains a DC solenoid valve which opens in response to a valve-open signal.

  The main feedwater pump safety injection suction valves, HV-853A and HV-853B, are also vented to depressurize the between-disc area through a normally open manual valve in a bonnet vent line. To aid in depressurization of the area between the main feedwater pump discharge and the HV-851 valves, the discharge check valve disc for the main feedwater pumps has been notched.

- Option 2. The second option would install a bypass line around HV-851A and HV-851B. This bypass would serve to vent the area between the valve disks and equalize the pressure around the valve. The bypass line would contain two DC powered solenoid valves in series which would receive an open signal when a safety injection signal is received. In this option, a running main feedwater pump would not be signaled to stop by the safety injection signal (Reference 11).

- Option 3. A third option was analyzed as a possible long term modification to the system design. This option replaces the hydraulic valve operators presently installed on HV-851A and HV-851B with fast acting motor operators. With this option, the main feedwater pumps will continue to operate while the valves are cycling.

4. RELIABILITY OF THE MODIFIED SONGS-1 SAFETY INJECTION SYSTEM

4.1 METHODOLOGY

The methodology used here for the system and data analysis follows that described in Reference 12, which is attached to this report as Appendix A. While the appendix discusses all aspects of risk assessment, Sections 0.9 and 0.12 through 0.16 describe the methods used here. Briefly, we model to the level of component functional failure modes to determine the failure logic for the system. Then we build a table of all possible causes for each failure mode and quantify unavailability based on these causes. Data analysis uses three basic kinds of information:

- General engineering knowledge of the design and the manufacture of equipment in question.
- Past experience in the plant being studied (SONGS-1).
- Historical performance in applications similar to the one in question.

Bayes' theorem is used to specialize generic data to plant specific distributions. Finally, unavailability is presented in probability of frequency format to completely display the underlying uncertainties in the results.

4.2 SYSTEM MODELS AND ANALYSIS

The system reliability models presented below are based on the safety analyses of Reference 1 and detailed plant information (Reference 13).

4.2.1 Assumptions and Model

Figure 4-1 presents a simplified system schematic for the safety injection system. Figure 4-2 is a simplified block diagram of the safety injection system which aids in defining the success criteria for the safety injection system. For all of the options considered, system success is at least one pump train (safety injection pump and feedwater pump) injecting to at least one intact reactor coolant system cold leg.

Figure 4-3 is the system fault tree which depicts those component failures which must occur to cause system failure. This fault tree, because of its construction, is applicable to all options considered.

The following assumptions concerning the safety injection system were made to assist in the quantification of the various options:

1. No operator action during safety injection. Operator action to recover from individual or train failures is not considered in this analysis for the following reasons:

   a. The short time which is available under large break loss of coolant accident scenarios to prevent core damage with no flow from the safety injection system.

   b. Failures in the feedwater pump suction could lead to damage of the feedwater pump in a short period of time.

2. A large LOCA has occurred. The large LOCA was assumed since its demands on the SIS are the most severe.

3. Electric power. All electric power required for successful operation of the safety injection is assumed to be available for the following reasons:

   a. This analysis is only concerned with the probability of the frequency of failure of the safety injection system.

   b. Failure of the electric power sources which supply the safety injection system affect all modifications equally.

   c. While analysis of the electric power system is beyond the scope of this analysis, failure of electric power has been found to be only a small contributor to SIS failure in other analyses (References 3 and 14).

4. Main feedwater pumps. The main feedwater pumps are assumed to be operating prior to the initiating event. Technical specifications require plant shutdown within 1 hour if one of these pumps is unavailable.

5. Operator errors prior to the initiating event. Operator errors during testing or maintenance of the safety injection pumps were considered for all options to make this analysis consistent with other analyses of similar systems.

6. Refueling water storage tank. The refueling water storage tank is assumed to be available for all options considered for the following reasons:

   a. Failure of the refueling water storage tank affects all options equally.

   b. The frequency of failure of the refueling water storage tank is insignificant in relation to other causes of system failure (References 3 and 14).

7. Safety injection signal. The safety injection signal is assumed to be present for both trains of safety injection for the reasons noted in item 3 above.

8. The safety injection system must start and operate for 1 hour for system success.

The following paragraphs describe the differences between the options and the data and methods used to quantify these differences.

4.2.2 Data

4.2.2.1 Generic Data. The generic data used for this report is based upon a review of various data sources and distributions performed in support of several probabilistic reliability analyses presently being performed. The data used and the source documents for the data are presented below.

	Mean	Variance
Motor-Operated Pump		
Failure to operate on demand* (Reference 15)	2.1×10^{-3}	2.5×10^{-6}
Failure during operation per hour* (Reference 15)	2.6×10^{-5}	3.9×10^{-10}
Hydraulically Operated Valve		
Failure to operate on demand* (Reference 16)	3.4×10^{-3}	9.7×10^{-5}
Motor-Operated Valve		
Failure to operate on demand* (Reference 16)	7.0×10^{-3}	2.8×10^{-5}
Check Valve		
Failure to open on demand (Reference 16)	3.0×10^{-4}	5.4×10^{-7}
Breaker (Greater than 480V)		
Failure to open on demand* (Reference 17)	7.9×10^{-4}	3.8×10^{-6}
Manual Valve		
Fails to remain open per hour (Reference 16)	3.4×10^{-8}	6.9×10^{-15}

*Failure data for these components includes associated control circuits and breakers if applicable.

	Mean	Variance
Solenoid Operated Valve		
Failure to open on demand* (Reference 3)	2.4×10^{-3}	3.3×10^{-5}
Relay		
Failure to operate on demand (Reference 3)	2.4×10^{-4}	3.3×10^{-7}

4.2.2.2 Updated Data. A review of plant maintenance, test, and operating records was performed to collect plant specific data for the main feedwater pumps and the hydraulically operated valves. This review covered the period from 1976 when the hydraulically operated valves were installed to September 1981.

Raw data from this review is summarized for the failure modes of interest below:

- Main Feedwater Pump
  - Failure to operate on demand. One failure in 105 demands.
  - Failure to continue operating per hour. Zero failures in 76,251 hours.
- Hydraulically Operated Valve (HV-851 through HV-854)
  - Failure to operate on demand (zero dp across the valve). One failure in 84 demands.**

A Bayesian update as described in Section 0.14.2 of Appendix A was performed using this data to generate a posterior distribution for the failure modes described above. The results of these updates are presented in Figures 4-4, 4-5, and 4-6. The mean values and variances of these distributions are presented on the following table.

*Failure data for these components includes associated control circuits and breakers if applicable.

**The failure on September 3, 1981, and those in the subsequent test program that provided the basis for identification of the systematic failure modes are excluded. The systematic failure modes due to operation under high differential pressure (galling and small mechanical margin of the operator) should be eliminated by the proposed modifications. Therefore, this analysis is applicable only when it is demonstrated that Options 1 and 2 actually eliminate excessive valve loads.

	Mean	Variance
Main Feedwater Pump		
Failure on demand	2.8×10^{-3}	3.5×10^{-6}
Failure to continue operating, per hour	1.5×10^{-5}	7.0×10^{-11}
Hydraulically Operated Valve		
Failure to operate on demand	5.6×10^{-3}	3.8×10^{-5}

4.2.3 Option 1 - Main Feedwater Pump Trip

4.2.3.1 Random Failures. Under this option, the main feedwater pumps are tripped upon receipt of a safety injection signal. Valves HV-851A, HV-851B, HV-853A, and HV-853B are sent opening signals and valves HV-852A, HV-852B, HV-854A, and HV-854B are sent closing signals. Note: HV-851A and HV-851B are interlocked with valves HV-854A and HV-854B, respectively, in such a manner that HV-851A and HV-851B will not open until their associated valves, HV-854A and HV-854B, are fully closed. This interlock exists for all options.

To ensure venting of the between-disc area of the closed hydraulic valves, HV-851A, HV-851B, HV-853A, and HV-853B, bonnet vents have been added to these valves.

1. HV-851A and HV-851B are vented to the upstream side of the valve (main feedwater pump discharge). A normally deenergized closed DC solenoid valve is included in each vent line. This DC solenoid valve is energized open upon receipt of a valve-open signal for its associated hydraulically operated valve.

2. HV-853A and HV-853B are vented to their respective downstream sides through a manual, locked open, bonnet vent valve. This change is applicable to all options.

Eleven seconds after the safety injection signal, the main feedwater pumps are signaled to start. This time delay is accomplished by a time delay relay in each pump's breaker control circuit. Eleven seconds after the safety injection signal, valves MOV-850A, MOV-850B, and MOV-850C are signaled to open. Their time delay is also accomplished by a time delay relay in each valve control circuit.

System failure for Option 1 can now be quantified. The failure data for this option and all other options is discussed in the data section which precedes this section.

Referring to Figure 4-3, system failure results from failure of safety injection train A and safety injection train B (where a safety injection train includes all components from the suction of the safety injection pump to the discharge of the main feedwater pump) or failure of all injection paths to the reactor coolant system.

Failure of train A of the safety injection system can be caused by the following failures:

1. Failure of the suction isolation valve to pump G-50A to remain open.

2. Failure of the safety injection pump, G-50A, to start and operate for 1 hour.

3. Failure of the discharge check valve for pump G-50A to open and remain open.

4. Failure of HV-853A to open and remain open.

5. Failure of HV-854A to close.

6. Failure of the main feedwater pump, G-3A, to trip or restart and operate.

7. Failure of the discharge check valve for pump G-3A to open and remain open.

8. Failure of HV-852A to close.

9. Failure of HV-851A to open and remain open.

The frequency of failure on demand for the hydraulically operated valves from the data section is 5.6×10^{-3} (Note: this value and all values reported in this section are mean values). The frequency of failure of valve HV-851A is increased to 8.0×10^{-3} due to addition of the DC solenoid valve which must open to vent the pressure between the valve disks for HV-851A.

The frequency of failure to start and operate for 1 hour for the safety injection pump is 2.1×10^{-3}.

The frequency of failure of the main feedwater pump is 3.8×10^{-3} (stop, start, and operate for 1 hour).

Check valve failure to open on demand is 3.0×10^{-4}. For two check valves this value is 6.0×10^{-4}

Failure of the safety injection pump suction valve to remain open is determined as follows:

The safety injection pumps are tested monthly. Failure of this valve to remain open would be detected at this time. The frequency of failure per hour for a manual valve failing to remain open is 3.4×10^{-8}. Using one-half of the test interval as the mean detection time for this failure, the frequency of failure on demand of this valve is 1.2×10^{-5}.

For a single safety injection train, the frequency of failure on demand is the sum of the individual failure frequencies and is 3.1×10^{-2}. Train B is similar to train A.

System failure due to random failure of both safety injection trains is 1.6×10^{-3}.

Failure of a single injection MOV to open on demand is 7.0×10^{-3}. This value is increased to 7.2×10^{-3} per demand due to the additional time delay relay added as a result of this modification.

Failure of a single injection line is the sum of the MOV failure to operate on demand and the check valve failure to open on demand and is 7.5×10^{-3}.

For the purpose of quantification, one injection line is assumed to be discharging to a failed reactor coolant loop. System success requires injection via one of the two remaining paths. System failure due to random failure of the injection paths is now 8.3×10^{-5}.

The frequency of system failure due to random failures of equipment is the sum of the failures due to injection pump train failures and injection path failures and is 1.6×10^{-3}.

4.2.3.2 Test and Maintenance. Failure of the safety injection system while maintenance is being performed on a safety injection pump is negligible when compared with other failure mechanisms for the following reasons:

1. Based upon the plant technical specifications, a single safety injection pump may only be out of service for 1 hour before the plant must be shut down.

2. The frequency of occurrence of maintenance during plant operation for these pumps is less than one act per year.

Human errors after maintenance or during maintenance are precluded by the post-maintenance and prestartup testing performed on the pumps.

System testing does not contribute to system failure for the following reasons:

1. Testing of the safety injection pumps is performed monthly by starting and operating each pump on recirculation flow. No system valve or switch lineups are made to perform this testing.

2. Testing of the other components is performed during plant cold shutdowns (refueling). Errors in testing will be detected during plant startup.

4.2.3.3 Other Causes of System Failure. A review of plant records was performed to determine: (a) potential common cause failure mechanisms, and (b) quantify these mechanisms, if found.

The safety injection pumps are tested monthly during plant operation. To date, there have been no reported failures of these pumps to operate on demand.

The modifications made to the main feedwater pump breaker control circuit consist of the addition of a time delay relay. This relay is common to the nuclear industry. Failure of this pump to restart after a safety injection signal due to control circuit modifications is not significantly affected. No failure of two pumps to restart due to common cause failure mechanisms in their control circuits was quantified for this reason.

Failure of these pumps due to hydraulic instabilities in the suction or discharge piping (flashing, etc.) is significantly reduced because the pump is no longer operating during the switchover to the refueling water storage tank.

A potential for common cause failure of the main feedwater pump motors due to the stop/start transient does exist; however, industry data for similar equipment (4,160V motors driving essential pumps under a loss of power condition) indicate that the frequency of occurrence of multiple failures due to this condition is low.

The modifications performed on the control circuits of the hydraulically operated valves consist of changing the safety injection signal timed input. No change in the operation of the valve due to this modification is expected.

The addition of the vent line to the bonnet of valves HV-851A and HV-851B will not increase the frequency of failure to operate on demand and should decrease this frequency. Common cause failure of the DC solenoid in this vent line is quantitatively insignificant.

Failure of HV-851A and HV-851B to open on demand due to mechanical effects (such as differential pressure) is an identified common mode failure mechanism. The modifications that are installed are designed to reduce the frequency of occurrence of this event. Plant testing with zero differential pressure across the valve disks has shown that the frequency of failure of the valves is low (5.6×10^{-3} per demand) and of a random nature.

For the reasons identified above, a beta factor of 0.01 is subjectively assigned to the frequency of occurrence of multiple safety injection train failure due to common cause mechanisms for this option. This results in a frequency of occurrence of system failure due to all causes of pump train failures of 1.9×10^{-3}.

System testing prior to plant startup under the expected operating conditions (i.e., with a main feedwater pump operating, a safety injection signal is simulated and system response verified) is necessary to validate this beta factor.

Common cause failure of the motor-operated valves in the injection lines was also considered. A review of industry experience with motor-operated valves was conducted resulting in a beta factor of 0.05 for multiple motor-operated valve failures. This beta factor is applicable to the injection MOVs for all options considered. The frequency of occurrence of injection line failure due to all causes of failure is now 4.6×10^{-4}.

The total frequency of failure of the safety injection system on demand is now 2.3×10^{-3} and displayed as a frequency distribution for Option 1 in Figure 4-7.

4.2.4 Option 2 - HV-851 Bypass

4.2.4.1 Random Failures. Under this option, the main feedwater pumps remain running upon receipt of a safety injection signal. A bypass line consisting of a bonnet vent connection, an upstream and downstream main line connection, and two DC operated solenoid valves in each bypass line, is used to equalize the differential pressure across HV-851A and HV-851B.

Valve interlocks for the HV-845 series and HV-851 series valves remain as stated in Section 4.2.3.1.

System success requirements remain as stated previously.

Failure of a safety injection train due to main feedwater pump failure is now caused by a main feedwater pump failing to remain in operation with a mean frequency of failure per hour of 1.5×10^{-5}.

Failure of HV-851A or HV-851B to open on demand is caused by mechanical failure of the valve or failure of either DC solenoid valve to operate with a mean frequency of failure on demand of 1.0×10^{-2}.

All other component random failures remain as previously quantified.

The frequency of train failure is now 3.0×10^{-2}.

System failure due to random failures in both safety injection trains is now 1.5×10^{-3}.

Total system random failure is the sum of the safety injection train failures and the injection line failures, and is 1.6×10^{-3}.

4.2.4.2 Other Causes of System Failure. For this option, hydraulic instabilities leading to main feedwater pump failure during the suction line switchover is a possible source of common cause failure. During the switchover, the supply line to the main feedwater pump from the condensate system is isolated while the suction from the refueling water storage tank is opening. During the period of time while the supply is shifting over, the main feedwater pump continues to run. The mode of operation is similar to the mode of operation which existed prior to the modification which is being analyzed and no failures were reported for this mode of operation.

In addition, because the main feedwater pumps remain in operation for this option, the differential pressure across the HV-851 series valves may be more severe for this option. If the pressure across HV-851A and HV-851B is not equalized prior to the opening of the injection path motor-operated valves, failure of HV-851A and HV-851B to open is likely to occur.

Because of the concerns stated above, a beta factor of 0.01 is subjectively assigned to the frequency of occurrence of multiple pump train failures due to common cause mechanisms for this option which results in a frequency of failure of the pump trains due to all causes of 1.8×10^{-3}.

System testing prior to plant startup under the expected operating conditions (i.e., with a main feedwater pump operating, a safety injection signal is simulated and system response verified) is necessary to validate this beta factor.

The total frequency of failure of the safety injection system for the operating conditions specified is 2.3×10^{-3} and is displayed as a frequency distribution for Option 2 in Figure 4-7.

4.2.5 Option 3 - Motor Operators for Valves 851A and 851B

4.2.5.1 Random Failures. This option changes the operators on the HV-851A and HV-851B valves from hydraulic to motor. The new operators will be qualified to open under the differential pressure conditions possible at these valves. The main feedwater pumps will continue to operate through the safety injection suction switchover sequence as previously designed. The interlock between the HV-854 and HV-851 series valves will remain as previously stated.

The frequency of failure on demand for a single motor-operated valve is 7.0×10^{-3}. For one motor-operated valve and three hydraulically operated valves in series, the frequency of failure on demand is 2.4×10^{-2}. All other component frequencies of failure remain as previously stated.

For a single safety injection train, the frequency of failure is 2.6×10^{-2}. System failure due to random failures in both safety injection trains is 1.0×10^{-3}.

Total system random failure is 1.1×10^{-3}.

4.2.5.2 Other Causes of System Failure. The statements in Section 4.2.3.2 concerning the operation of the main feedwater pumps remain valid for this case also.

Common cause failures of motor-operated valves is now a factor because of the increased complexity of the valve control circuitry. As stated previously, a beta factor of 0.05 for multiple motor-operated valve failures is used. The pump train frequency of failure for the operating conditions specified is 2.3×10^{-3}.

The frequency of failure of the system for this option is 2.8×10^{-3} and is displayed as a frequency distribution for Option 3 in Figure 4-7.

4.2.6 Comparison of the Results

The table shown below presents the results of the analyses performed for the various options with the pump train beta factors assumed for each case.

Option	Random Failure	Pump Train Beta Factor	Total
1	1.6×10^{-3}	.01	2.3×10^{-3}
2	1.6×10^{-3}	.01	2.3×10^{-3}
3	1.1×10^{-3}	.05	2.8×10^{-3}

The slight differences between Option 1, Option 2, and Option 3 are due primarily to the beta factor assumed for common cause failure of multiple safety injection pump trains. Although one may argue for a higher beta factor to be used in each case, the reasons stated for the difference in the beta factors remain the same and the relative difference should remain the same.

The percentiles associated with the total frequency of system failure for all the options analyzed are presented below.

Option	Mean	5th	Median	95th
1	2.3×10^{-3}	4.6×10^{-4}	1.3×10^{-3}	7.4×10^{-3}
2	2.3×10^{-3}	3.7×10^{-4}	1.2×10^{-3}	7.2×10^{-3}
3	2.8×10^{-3}	8.3×10^{-4}	1.9×10^{-3}	7.1×10^{-3}

Several points should be emphasized concerning this analysis.

1. All three options compare favorably with the injection systems analyzed in WASH-1400. No active single point failures were identified. The only single point failures, piping failure or failure of the refueling water storage tank, do not affect the results of this analysis.

2. On the basis of reliability, there is little to recommend one option over the others. The hydraulic operators used in Options 1 and 2 will only work if these options sufficiently unload the valve differential pressure. Therefore, prestartup testing at expected operating conditions is crucial if Option 1 or 2 is selected. Option 3 offers valve operators that are slightly less reliable under unloaded valve conditions, but which are much more capable of opening the valves if high differential pressure exists.

3. Because of the 1-hour limit on pump downtime during plant operation, maintenance does not significantly affect system frequency of failure on demand.

4. Tests performed during plant operation do not require valve lineup or pump lineup changes and therefore do not contribute to system frequency of failure on demand.

In order to explicitly identify the "degradation in reliability" introduced by the pump trip and additional complexity of the individual tuning relays of Option 1, additional calculations have been made. For a sensible comparison, we must turn to Option 3, the only option using valves qualified to operate under full differential pressure with the main feed pumps running. We have requalified Option 3 with the pump trip and restart of Option 1 and the tuning relays included. The results shown in Figure 4-8 are summarized here:

Option	Mean	5th	Median	95th
3	2.8×10^{-3}	8.3×10^{-4}	1.9×10^{-3}	7.1×10^{-3}
3' (pump trip)	3.2×10^{-3}	1.1×10^{-3}	2.3×10^{-3}	7.9×10^{-3}

The pump trip then, while making use of the hydraulically operated valves viable, does introduce a small (about 20%) degradation in realiability.

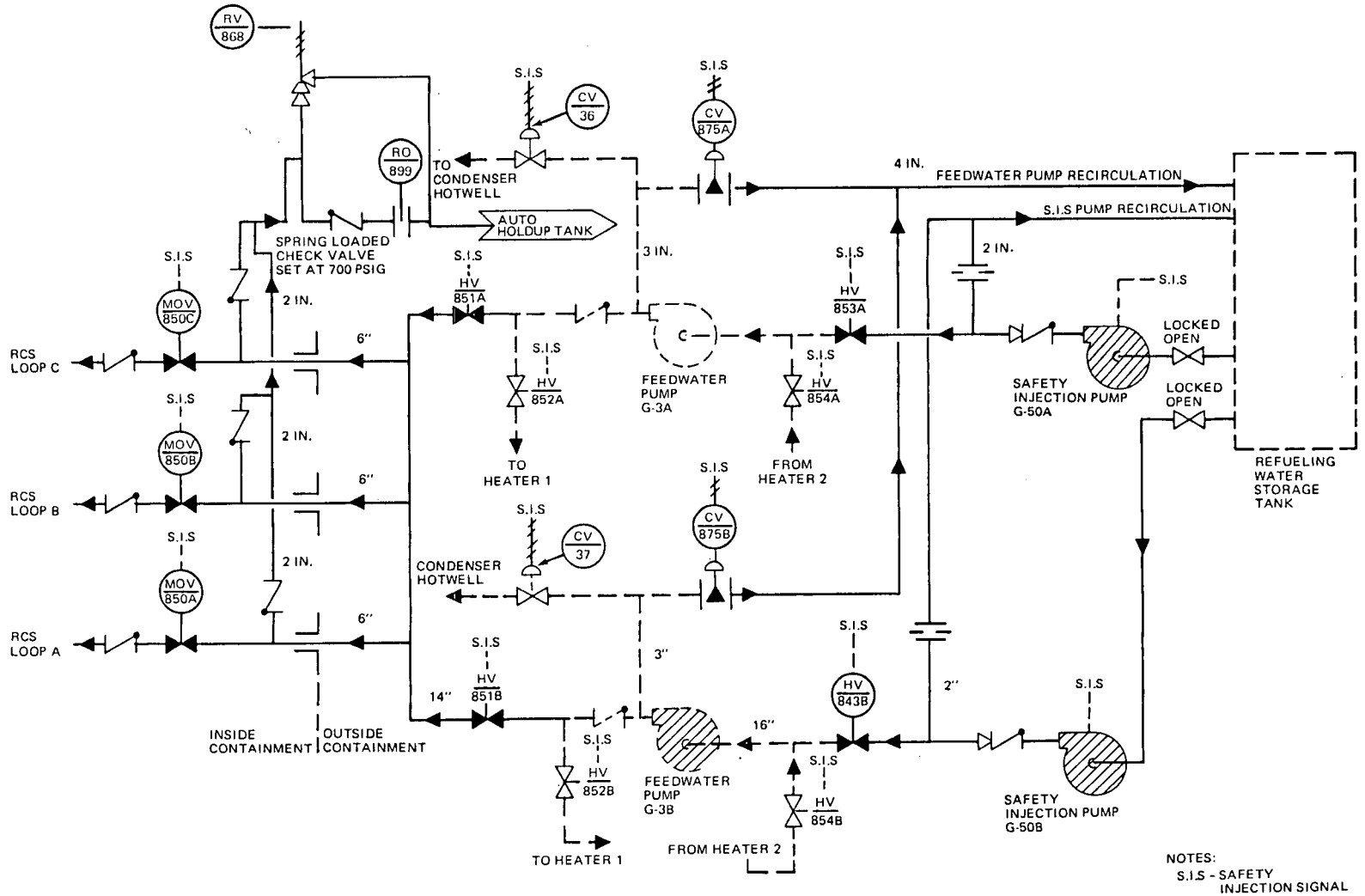


FIGURE 4-1. SIMPLIFIED SCHEMATIC OF SONGS-1 SAFETY INJECTION SYSTEM

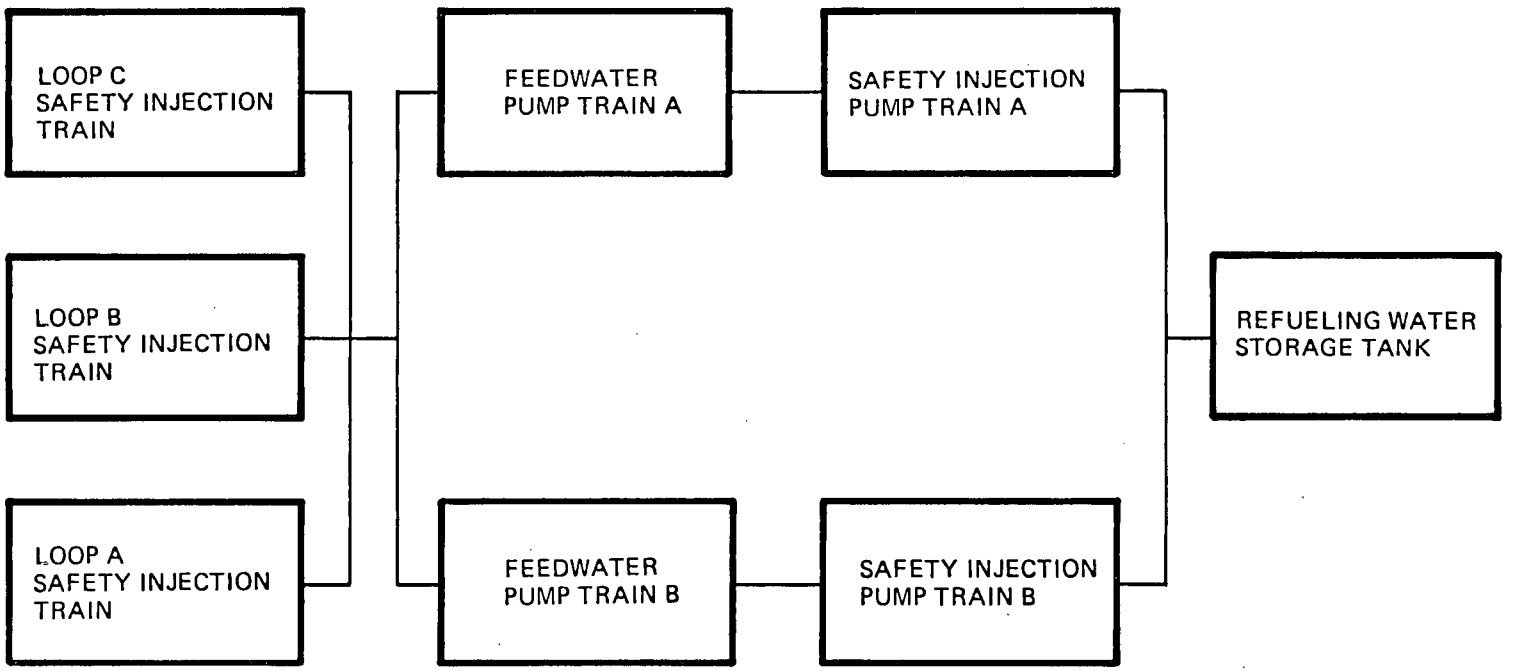


FIGURE 4-2. SIMPLIFIED RELIABILITY BLOCK DIAGRAM
OF THE SONGS-1 SAFETY INJECTION SYSTEM

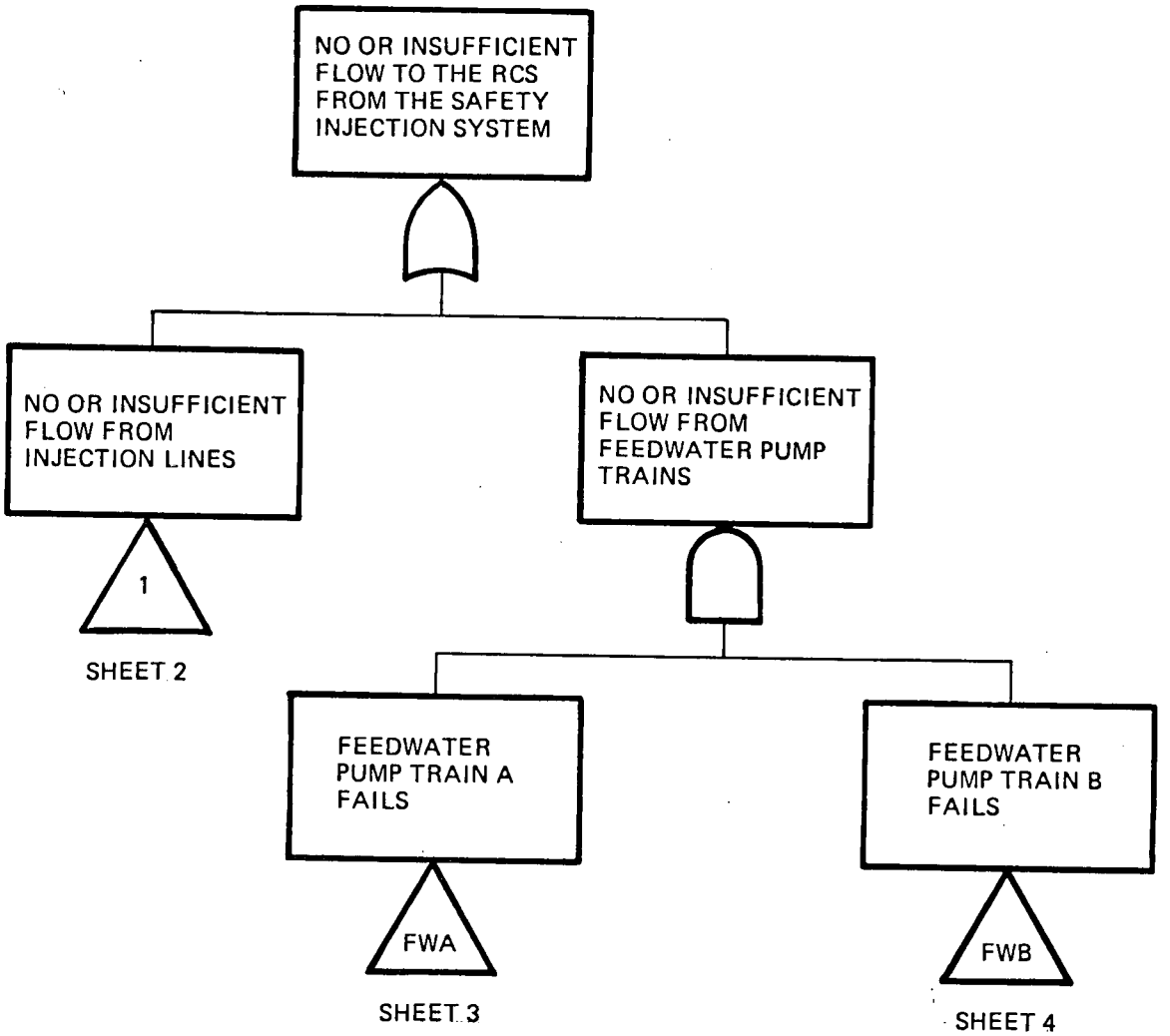


FIGURE 4-3. FAULT TREE FOR THE SONGS-1 SAFETY INJECTION SYSTEM (Large Loca Conditions) (sheet 1 of 10)

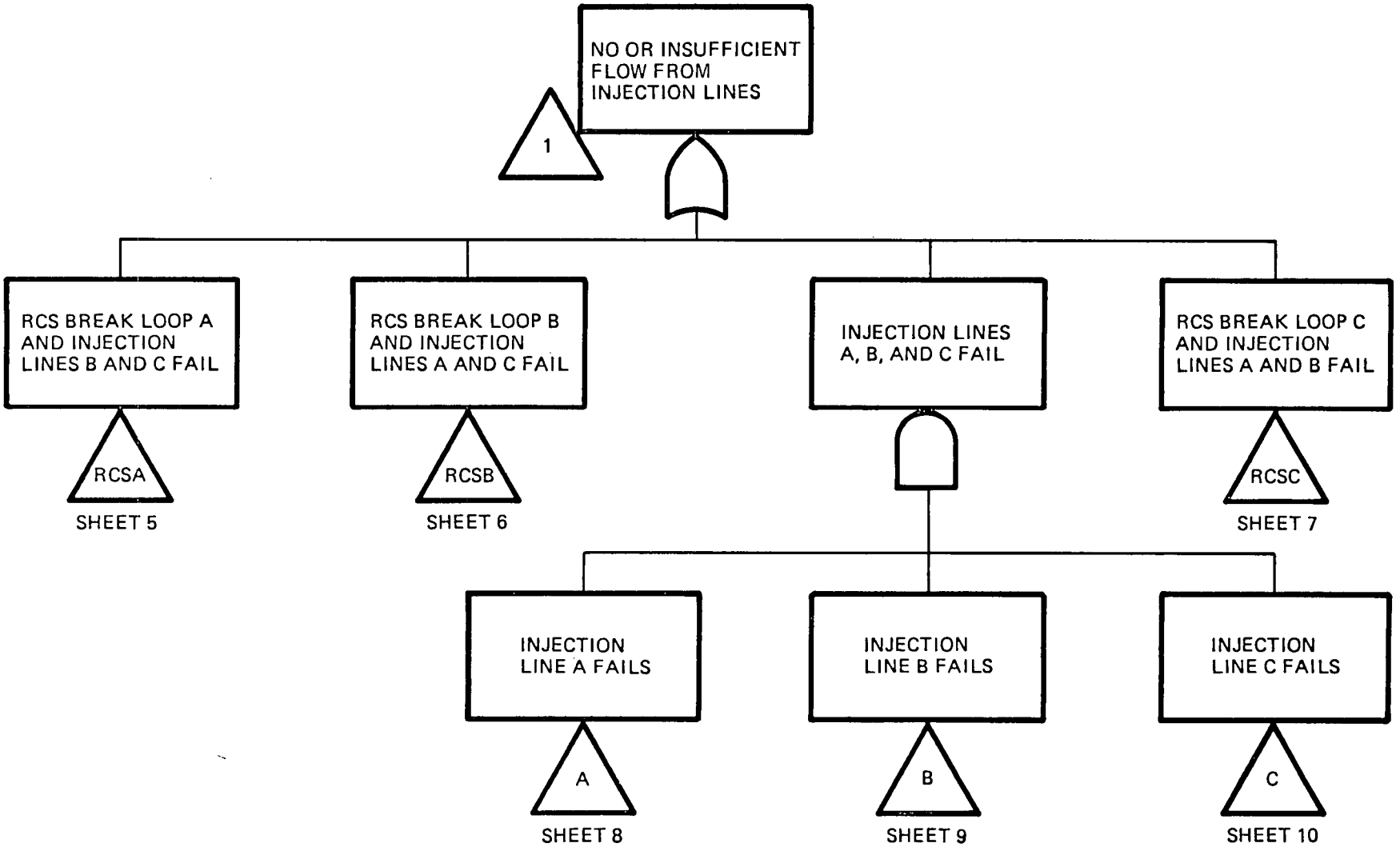


FIGURE 4-3 (sheet 2 of 10)

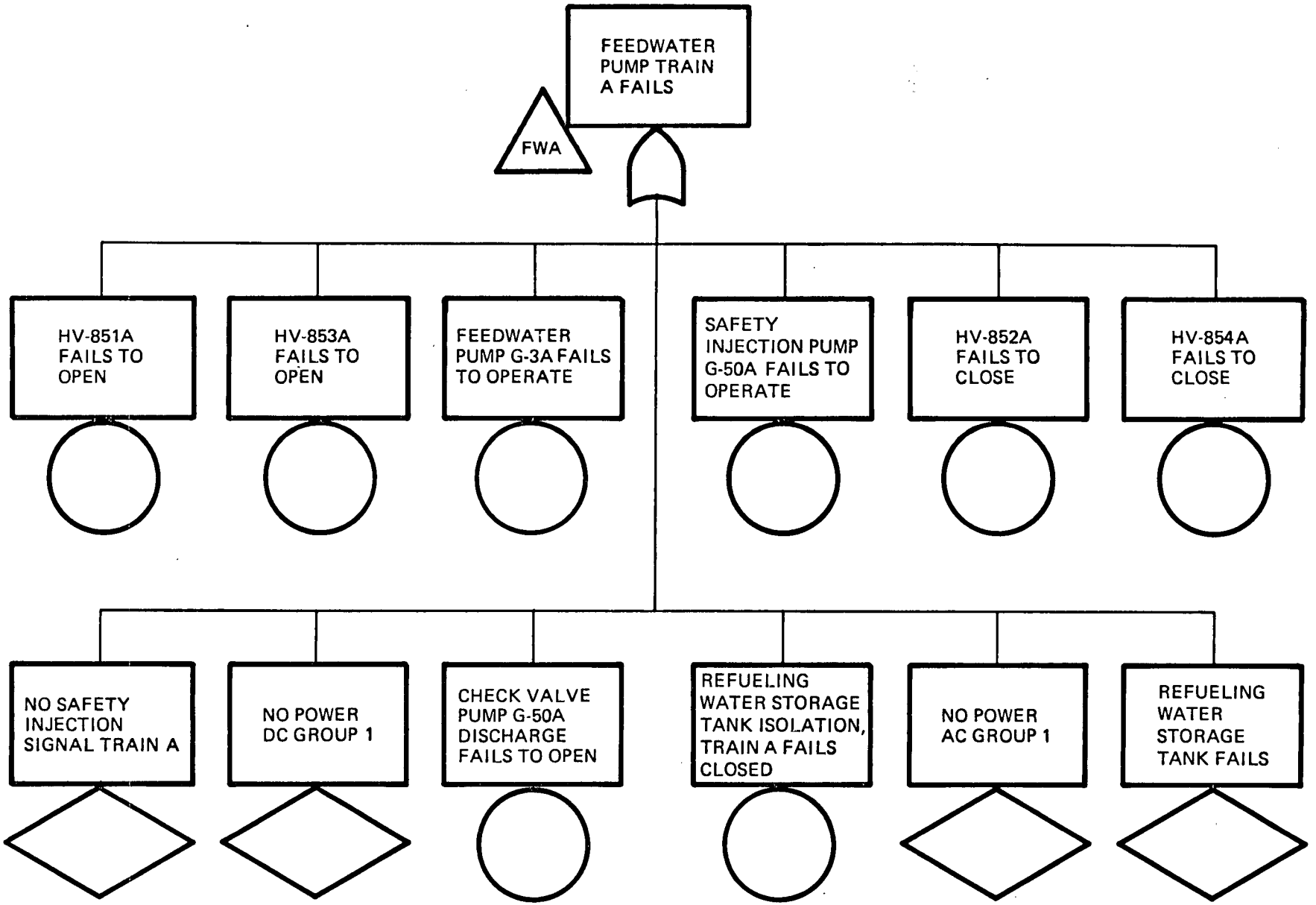


FIGURE 4-3 (sheet 3 of 10)

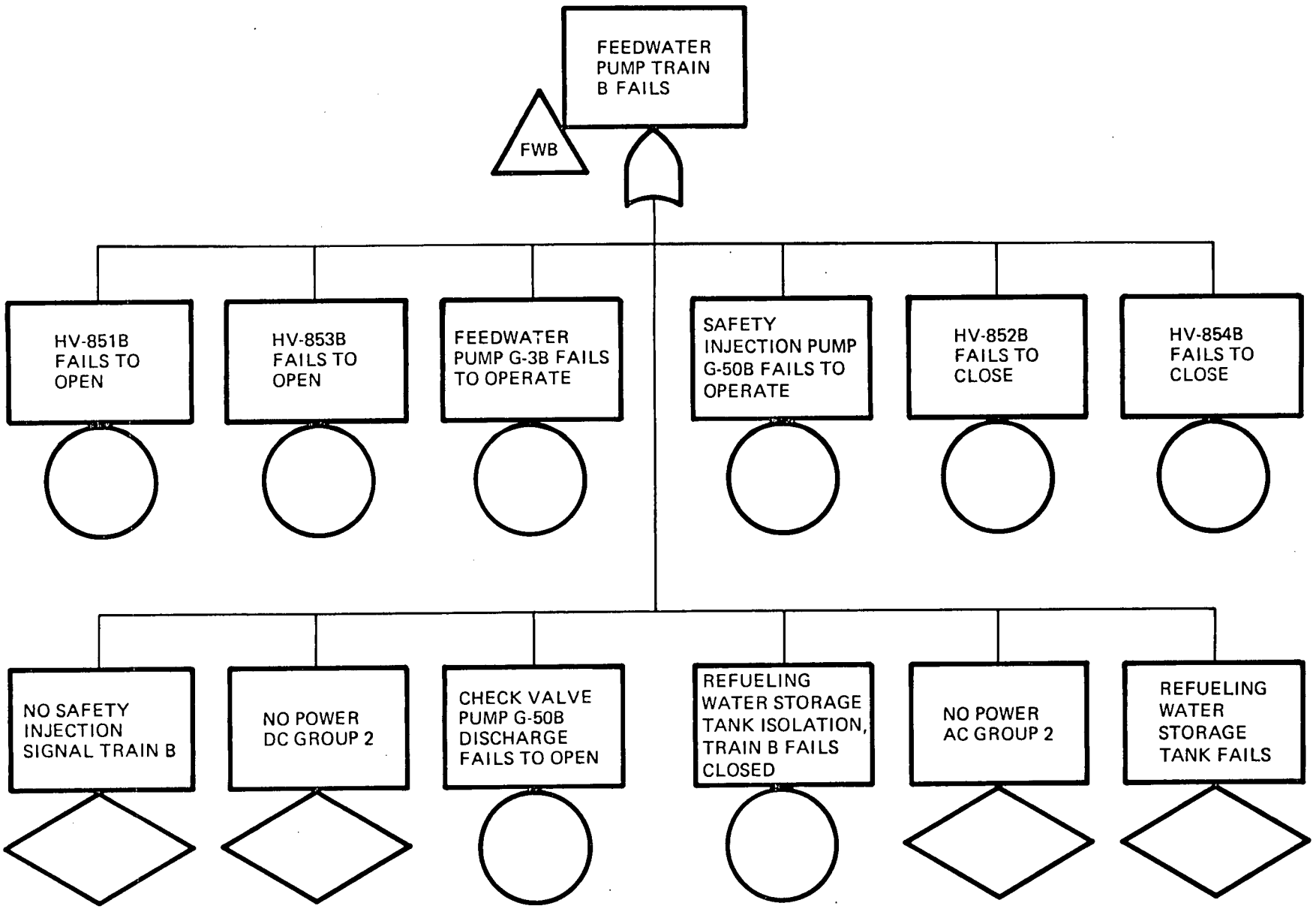


FIGURE 4-3 (sheet 4 of 10)

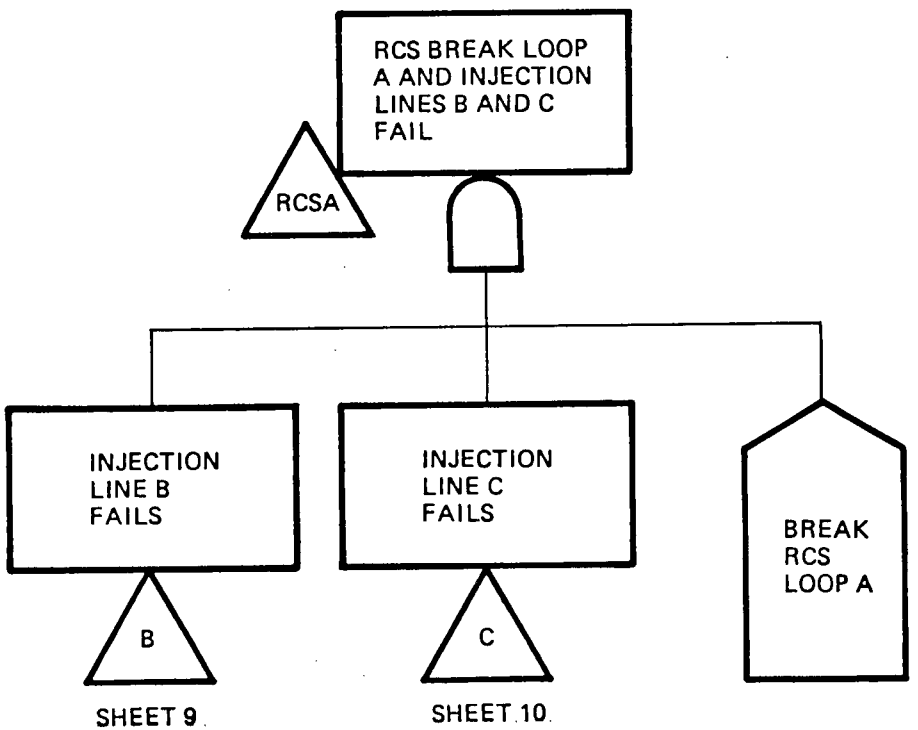


FIGURE 4-3 (sheet 5 of 10)

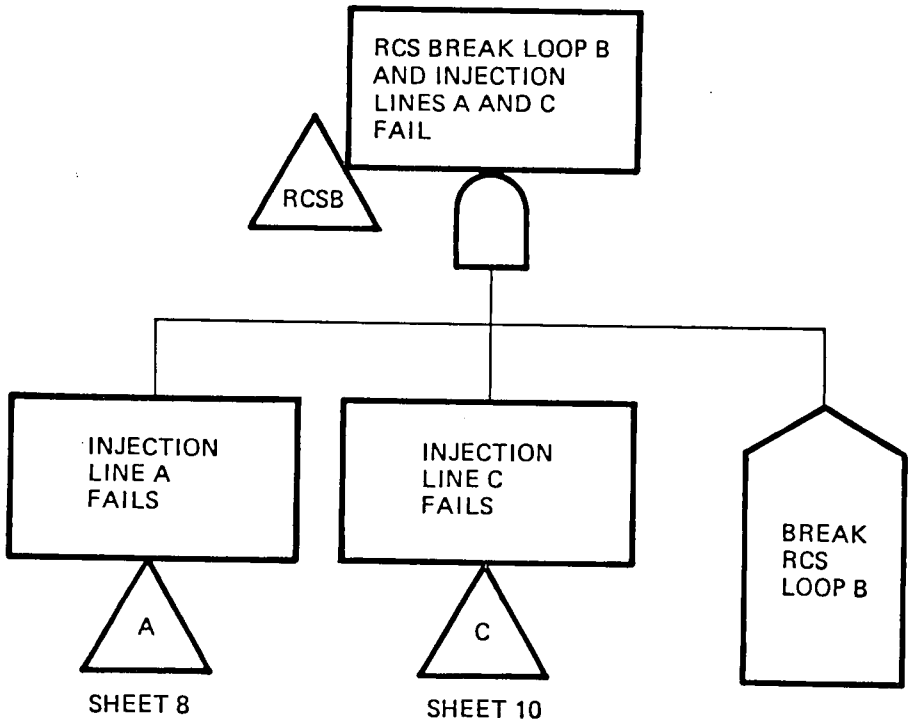


FIGURE 4-3 (sheet 6 of 10)

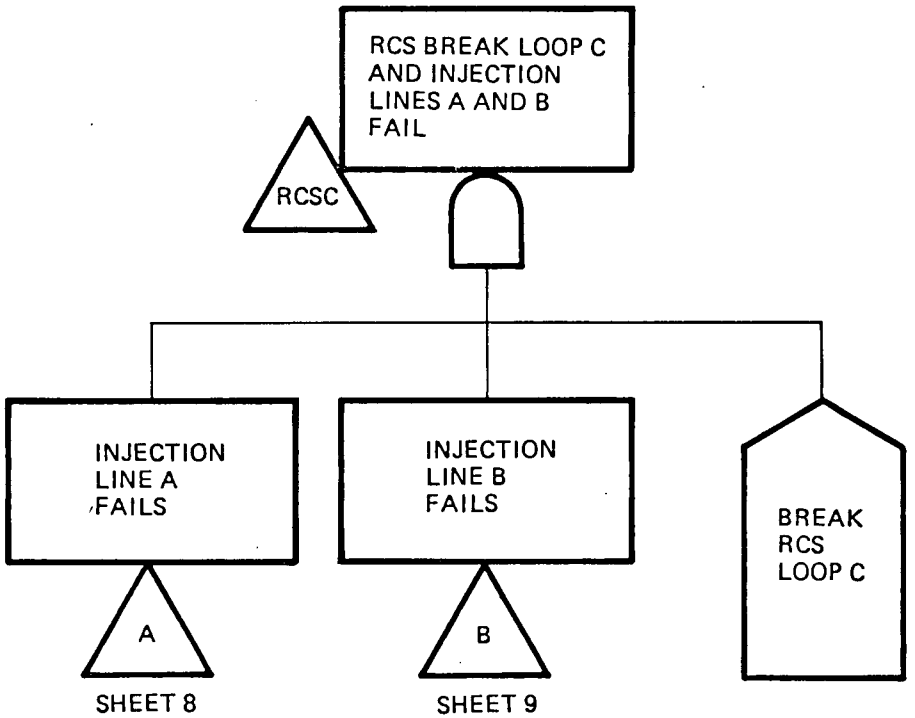


FIGURE 4-3 (sheet 7 of 10)

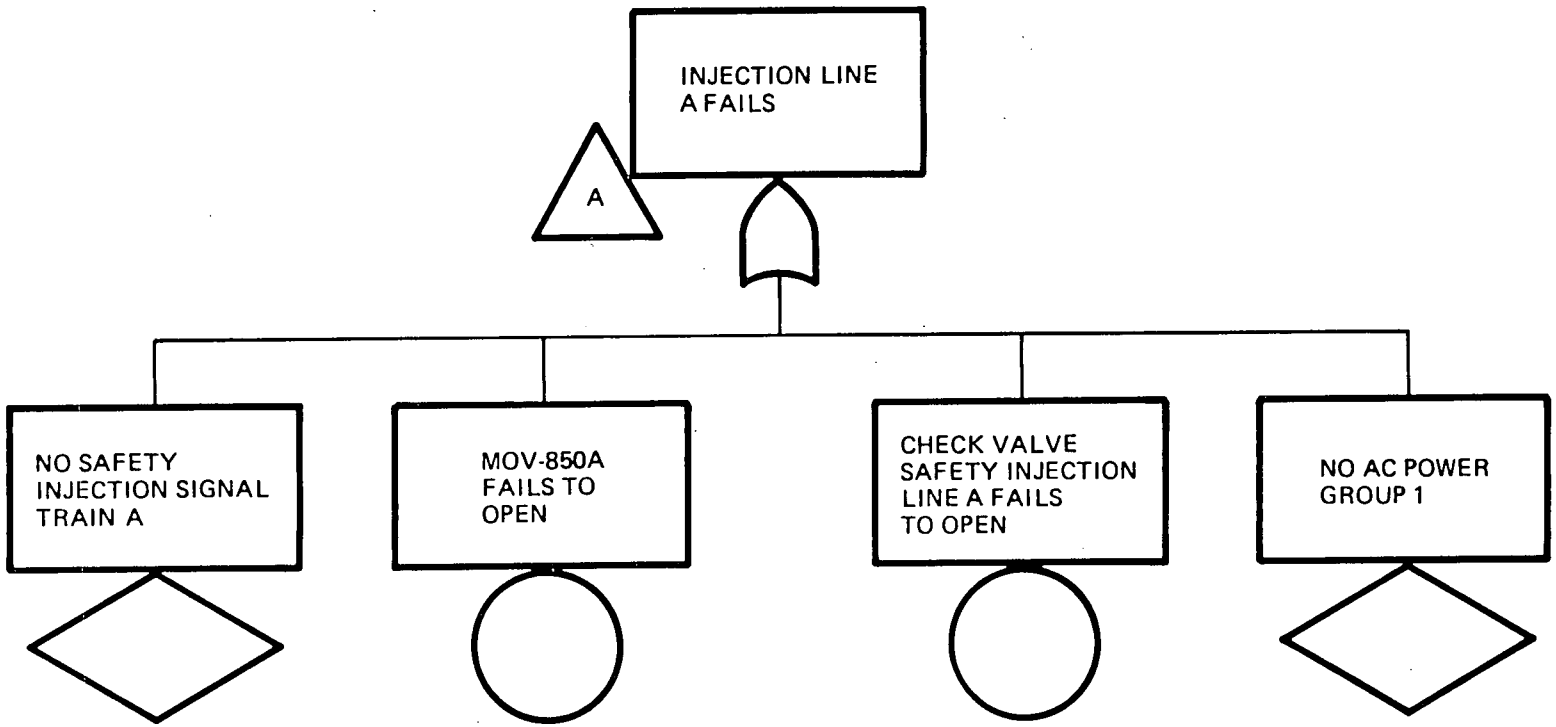


FIGURE 4-3 (sheet 8 of 10)

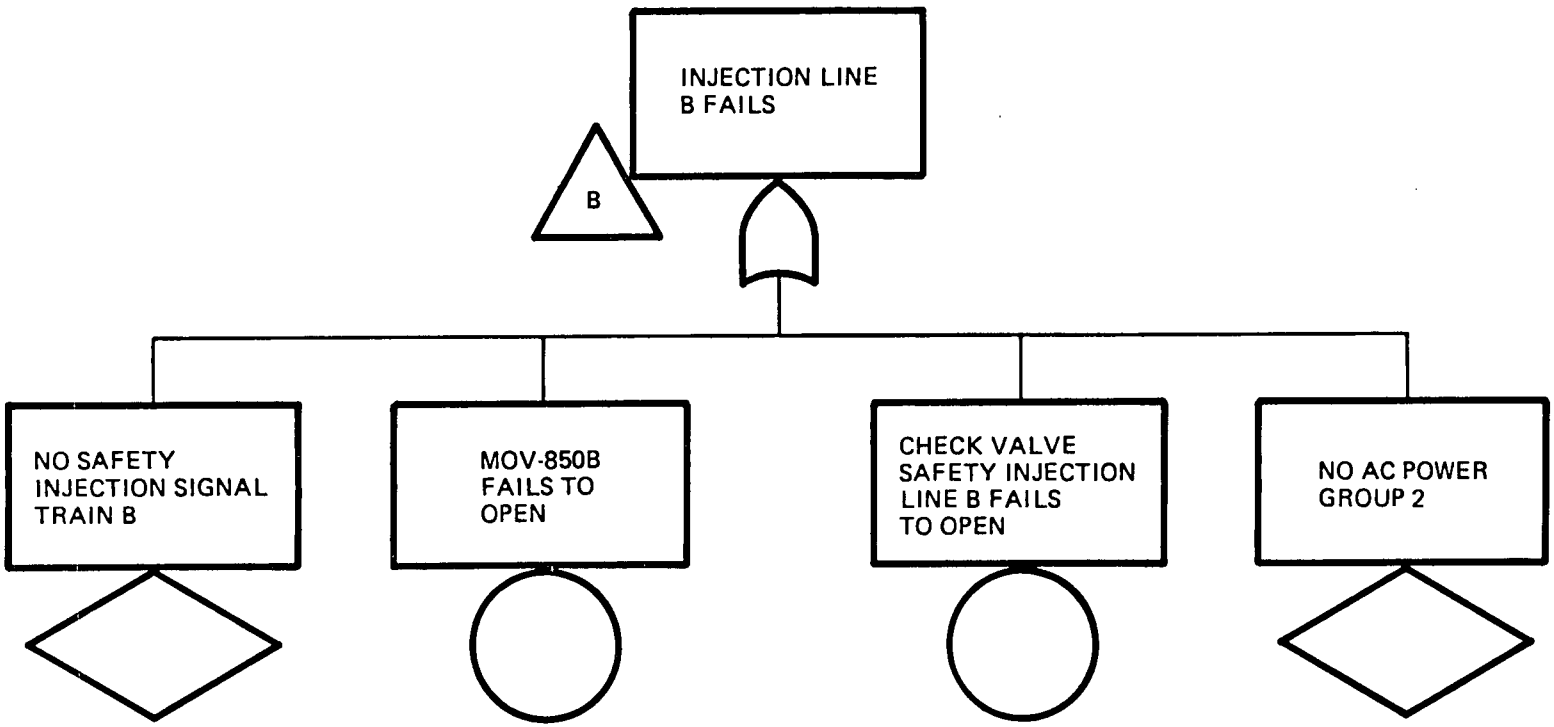


FIGURE 4-3 (sheet 9 of 10)

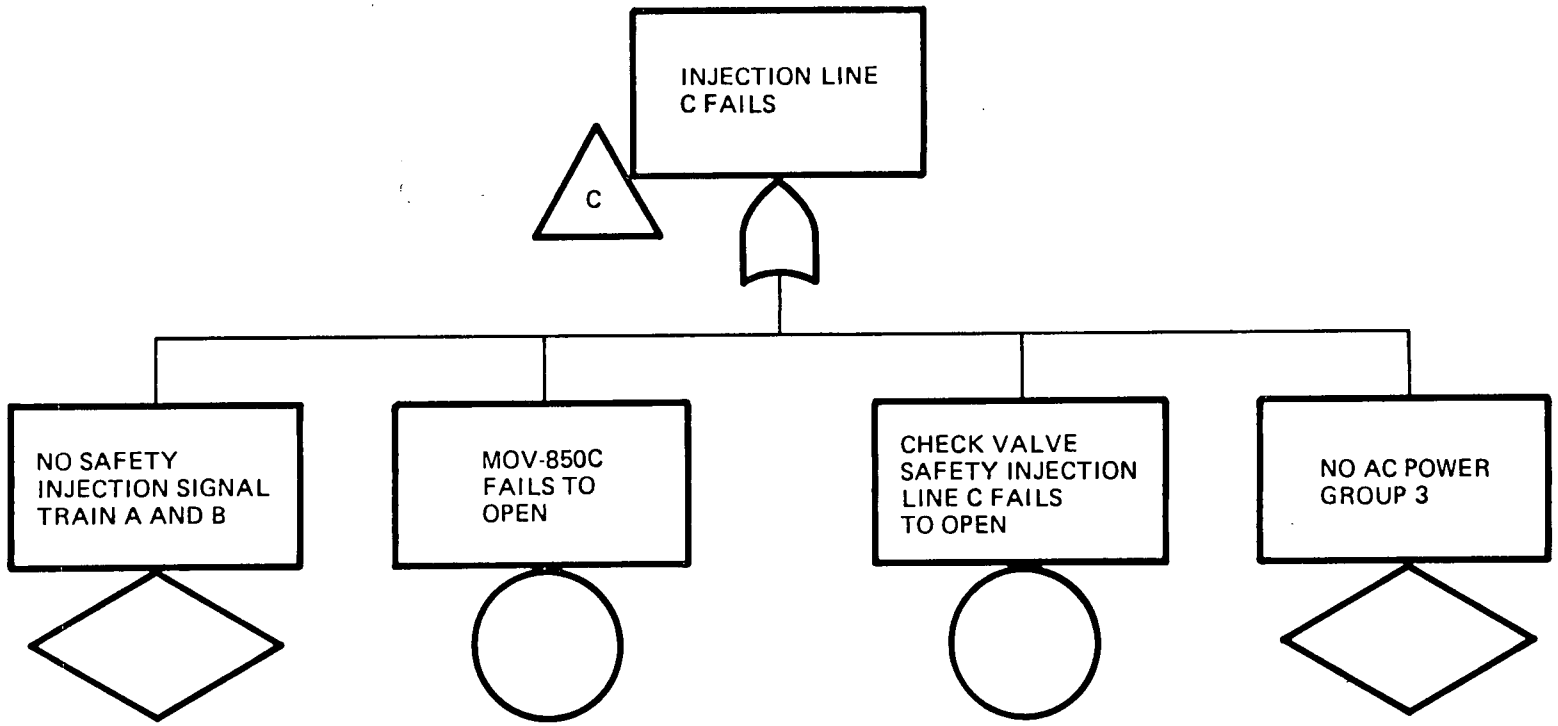


FIGURE 4-3 (sheet 10 of 10)

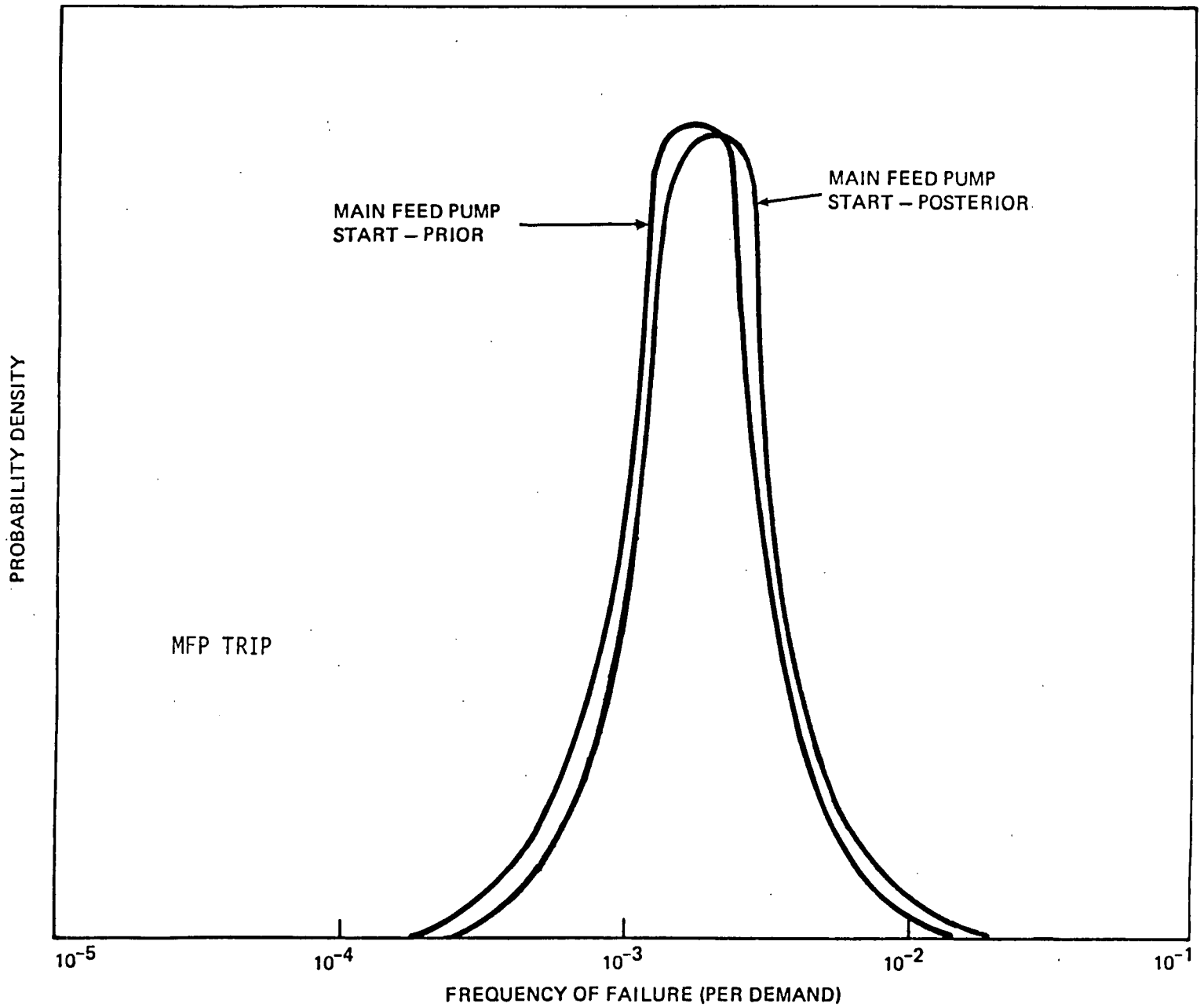


FIGURE 4-4. MAIN FEEDWATER PUMP PROBABILITY OF FREQUENCY OF FAILURE TO OPERATE ON DEMAND

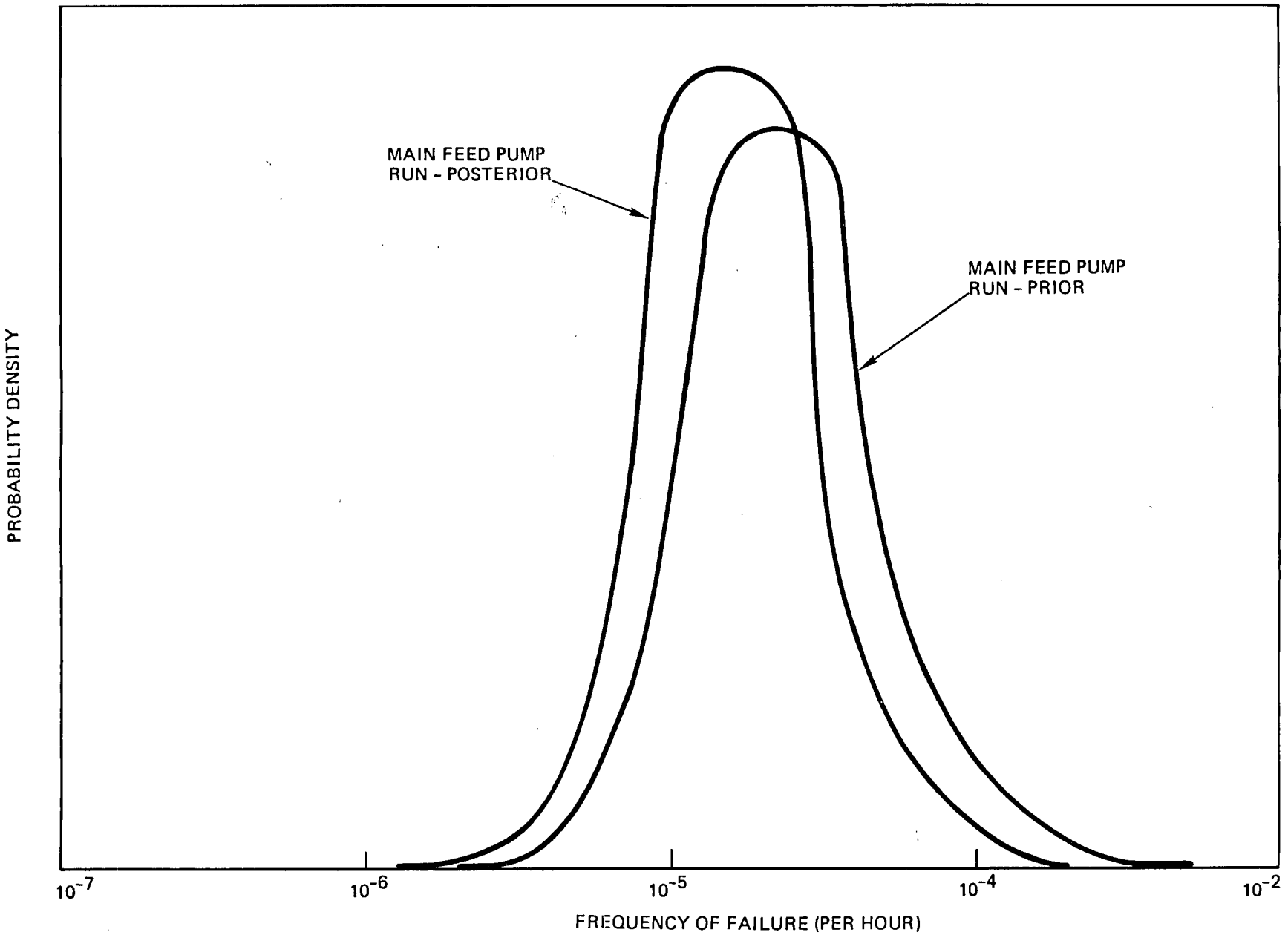


FIGURE 4-5. MAIN FEEDWATER PUMP PROBABILITY OF FREQUENCY OF FAILURE TO CONTINUE OPERATING

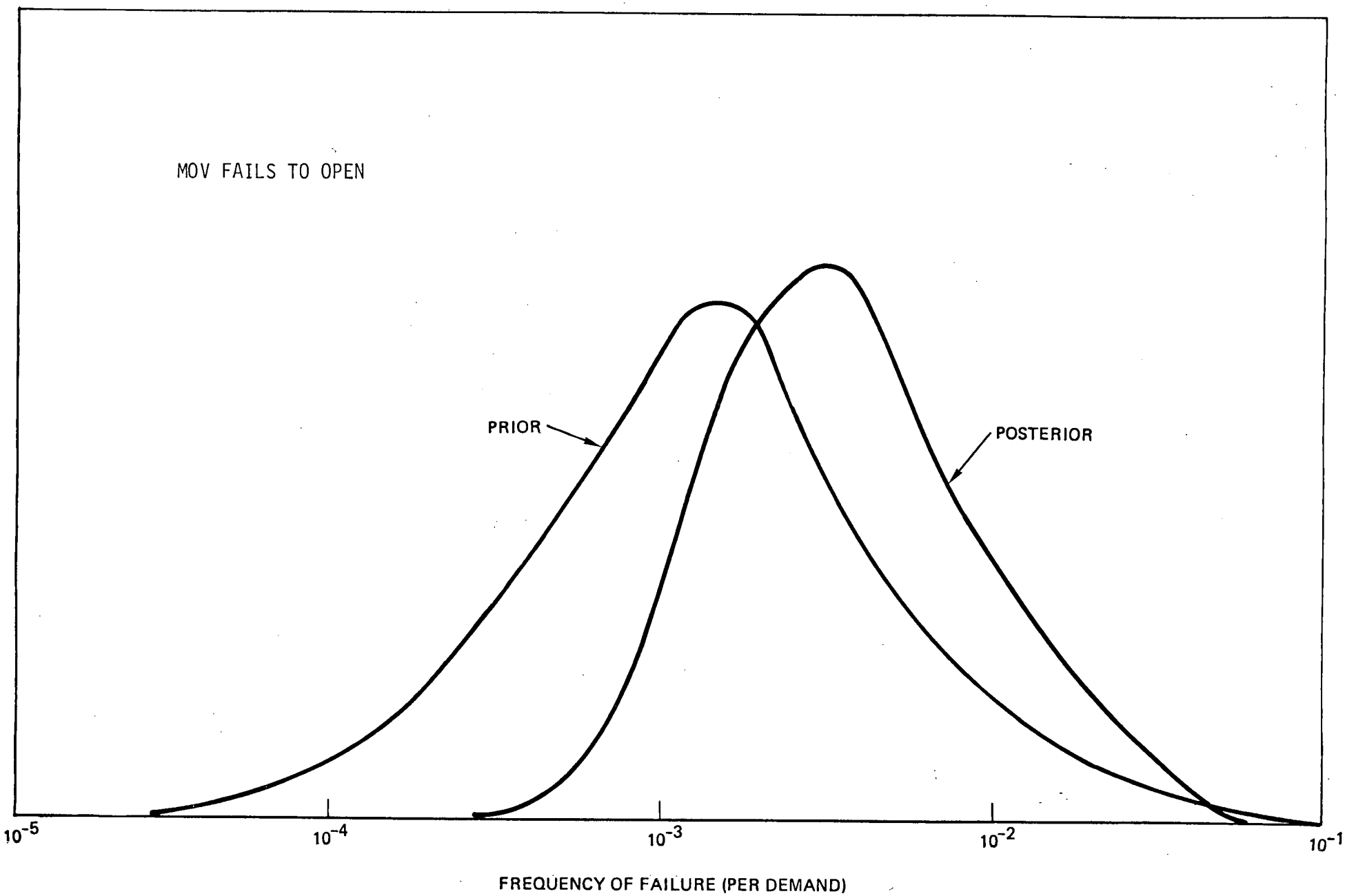


FIGURE 4-6. HYDAULICALLY OPERATED VALVES PROBABILITY OF FREQUENCY TO OPERATE ON DEMAND

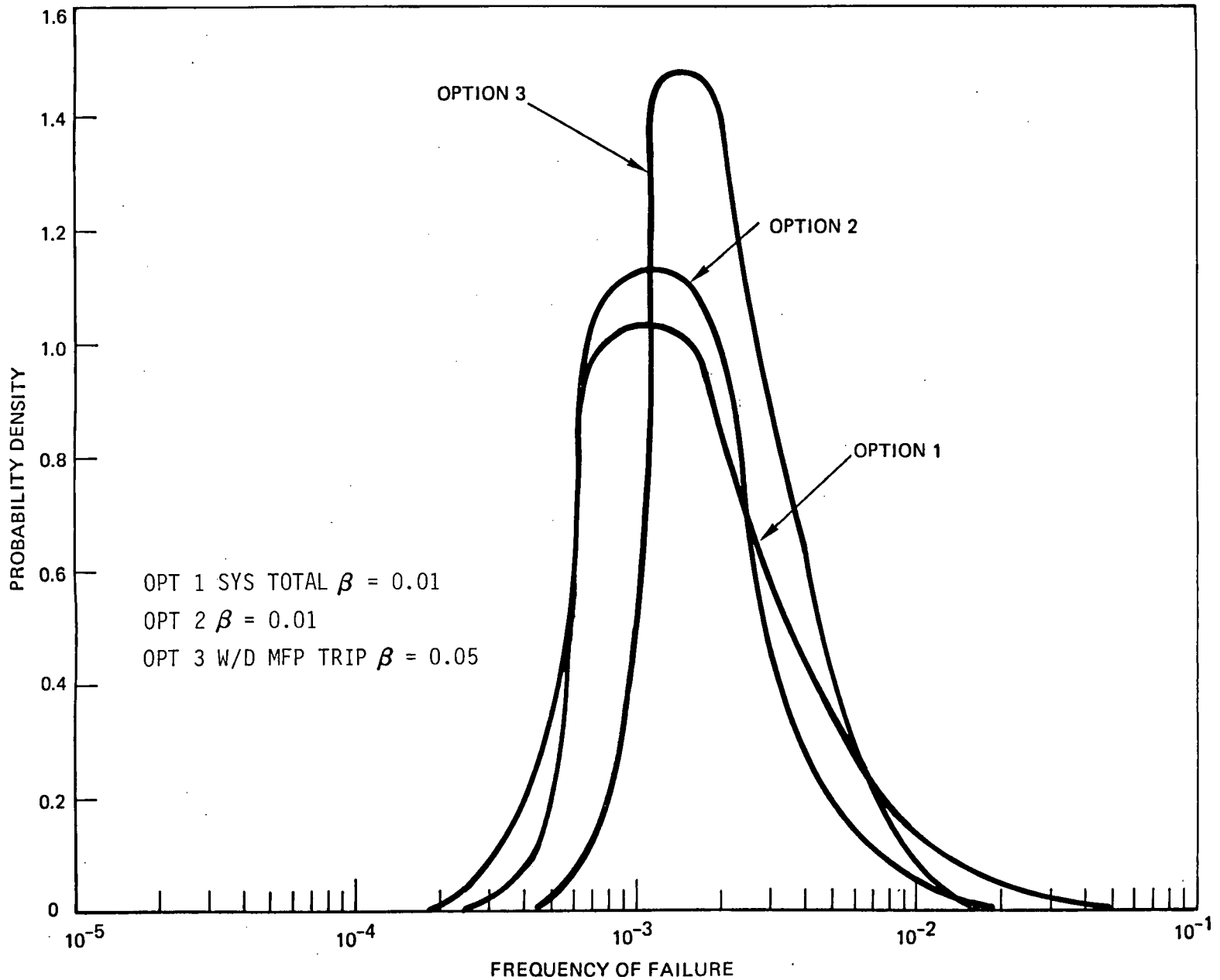


FIGURE 4-7. FREQUENCY OF FAILURE - SI SYSTEM - ALL OPTIONS

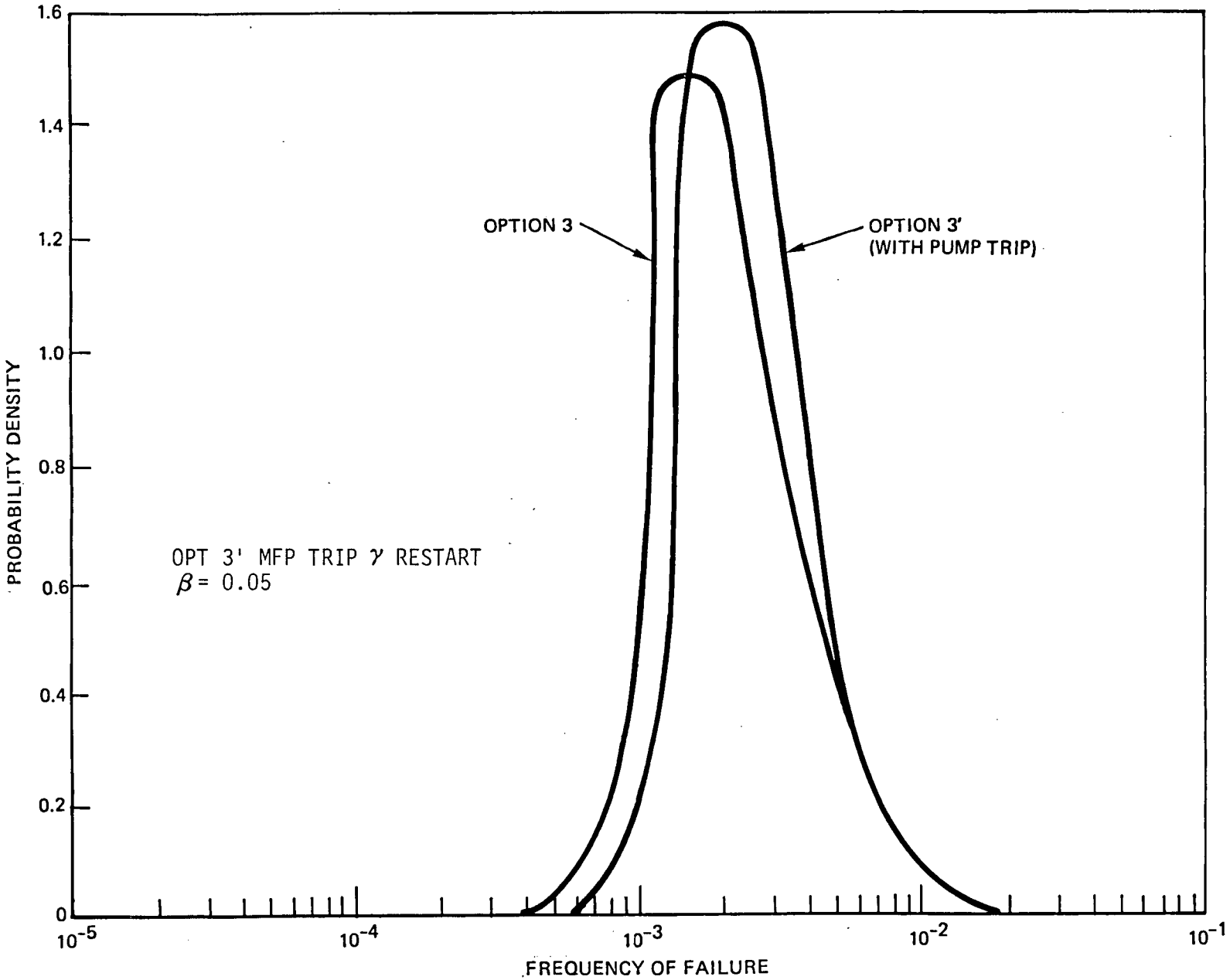


FIGURE 4-8. EFFECT OF PUMP TRIP AND TIMING RELAYS ON SYSTEM FREQUENCY OF FAILURE

5. REFERENCES

1. Southern California Edison Company, "Final Safety Analysis, San Onofre Unit 1."

2. "Zion Station Probabilistic Safety Study," Section 3, "Degraded Core Phenomena," submitted to the USNRC, September 1981.

3. U.S. Nuclear Regulatory Commission, "Reactor Safety Study," WASH-1400 (NUREG-75/014), October 1975.

4. Orlando, J. R., "Safety Injection Valve Testing," to SONGS-1 Startup Working Group, March 1, 1977.

5. Portanova, P. A., "Special Test Procedure for Cycling HV851 and HV854," memo to H. L. Richter, March 12, 1977.

6. Kaser, A., "Safety Injection Valve Testing (TP-14) SONGS-1 SPA," memo to H. L. Richter, March 28, 1977.

7. Baskin, K. P., "Docket No. 50-206, San Onofre Nuclear Generating Station Unit 1, NRC Questions Responded to," letter to K. R. Coller, USNRC, March 29, 1977.

8. SONGS-1 Nuclear Audit and Review Committee, minutes of meeting of June 1976, pp. 11-15, describing modification changing motor operators SIS valves to fast acting hydraulic operators.

9. Baskin, K. P., "Docket No. 50-206, San Onofre Nuclear Generating Station Unit 1, ECCS Performance Reanalysis," letter to D. Eisenhut, USNRC, May 30, 1978.

10. Moreton, B., "Safety Injection System Modification, Safety and Environmental Analysis, San Onofre Nuclear Generating Station Unit 1," Design Change 81-38, Bechtel Power Corporation, September 10, 1981.

11. Private communication with B. L. Curtis, Southern California Edison Company.

12. Pickard, Lowe and Garrick, Inc., "Zion Station Probabilistic Safety Study," Section 0, "Methodology," submitted to the USNRC, September 1981.

13. SONGS-1: P&IDs, maintenance and operating records, and discussions with plant personnel.

14. Pickard, Lowe and Garrick, Inc., "Zion Station Probabilistic Safety Study," Section 1, "Plant Analysis," submitted to the USNRC, September 1981.

15. "Data Summaries of Licensee Event Reports of Pumps at U.S. Commercial Nuclear Power Plants," NUREG/CR-1205, EG&G Idaho, Inc., 1980.

16. "Data Summaries of Licensee Event Reports of Valves at U.S. Commercial Nuclear Power Plants," NUREG/CR-1363, EG&G Idaho, Inc., June 1980.

17. "IEEE Guide to the Selection and Presentation of Electrical, Electronic and Sensing Component Reliability Data for Nuclear-Power Generating Stations," IEEE Std. 500-1977, June 1977.

APPENDIX A

ZION STATION PROBABILISTIC SAFETY STUDY

SECTION 0 - METHODOLOGY

September, 1981

SECTION 0
PROBABILISTIC RISK ASSESSMENT METHODOLOGY

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

0.0 PROBABILISTIC RISK ASSESSMENT METHODOLOGY

0.1 INTRODUCTION AND PURPOSE

The purpose of this section is to give an overall view of the basic methodology of risk analysis as used in this study. Various individual segments of this methodology are developed in greater depth in later sections. The emphasis here is on the overall structure and flow of the process and on how the various segments fit together. The section is divided into three major parts: Part 1, Definition of Risk; Part 2, Modeling and Analysis; Part 3, External Events.

To do a risk assessment, we obviously must agree upon the precise and usable definition for the word risk. This is the purpose of Part 1. This part begins (in Section 0.2) by discussing some qualitative aspects of the notion of risk as we use it in this study. It then proceeds, in Section 0.3, to give a quantitative definition in terms of a set of envisioned scenarios together with the probability and damage associated with each. This definition is called the "Level One" definition of risk. Section 0.4 explains the sense in which the word "probability" is used in this definition. For several reasons, given in this section, it is desirable to expand the Level One definition so that it may encompass some further subtleties of the idea of risk. Section 0.5 gives such an expansion and refers to it as the "Level Two" definition of risk. This latter definition then becomes the basis for the methodology of the study and the format for the presentation of the results.

Once the definition of risk is established, Part 2 then deals with the methods used to actually model and quantify the risk in a nuclear plant. Thus, with risk now defined fundamentally in terms of a list of scenarios, the next question is: "How does one identify and structure the scenarios on the list?" This question is addressed in Section 0.6. The key analytical device here is the "event tree" which is a structured presentation of the myriad of scenarios branching out of any given initiating event. Another key device is the notion of "pinch point" which allows the event trees to be sectioned into a "plant" segment, a "containment" segment, and a "site" segment.

With the scenarios identified and structured in terms of event trees, the next step is to determine frequencies of the various paths through the trees in terms of the "split fractions" at the branch points of the tree. This is discussed in Section 0.7.

Section 0.8 then addresses the question of assembling the information from this myriad of scenarios into a final presentation of the risk. The method chosen for this assembly takes maximum advantage of the organizational or structural properties of the list. Indeed, these properties and this method allow us to present the results in a very clean and compact matrix form which also provides much visibility into the performance of various parts of the plant. Thus any potential problems can be readily seen and the effects of proposed plant or procedure changes readily evaluated.

Sections 0.2 through 0.8 together, therefore, describe the definition of risk in terms of a list of scenarios; the identification, structuring, and quantifying of the list; and the assembly into a final presentation of risk curves. This much may be considered the "main stream" of the methodology. The remaining sections describe the numerous tributary flows into this stream.

Thus, Sections 0.9 and 0.10 describe the determination of the split fractions in terms of the frequencies of more basic "elemental" events. Section 0.11 describes the site modeling and consequences analysis given releases of radioactivity from the containment. Sections 0.12 and 0.13 review some of the basic mechanics of probability distributions and probabilistic calculations. Section 0.14 outlines the sources of information about the elemental events and the basic principle (Bayes' theorem) for combining these different types of information into probability distributions for the frequencies of elemental events. Section 0.15 discusses the treatment of an important type of elemental event, human error. Section 0.16 discusses some further aspects of the process of combining "elemental" probability distributions during the course of system analysis. Several important pitfalls are identified here relating to the dependence of probability distributions and the use of lognormal curves.

Finally, Part 3, External Events (Sections 0.17 and 0.18), provides detail on the methods used for seismic analysis and fire analysis.

PROBABILISTIC RISK ASSESSMENT

METHODOLOGY

Part 1

DEFINITION OF RISK

0.2 QUALITATIVE ASPECTS OF THE NOTION OF RISK

The subject of risk has become very popular in the last few years and is much talked about at all levels of industry and government. Correspondingly, the literature on the subject has grown very large. In this literature the word "risk" is used in many different senses and many different kinds of risk are discussed. Similarly, the words "hazard," "danger," "uncertainty," "probability," etc., are used in many different, intertwining senses. Therefore, in the interest of making this study more understandable, we should like, at the outset, to draw some distinctions in meaning between various of these words as we shall use them. We begin with "risk" and "uncertainty."

0.2.1 THE DISTINCTION BETWEEN RISK AND UNCERTAINTY

Suppose your rich relative had just died and named you as sole heir. The auditors are totaling up his assets. Until that is done you are not sure how much you will get after estate taxes. It may be one million or two. You would then certainly say you were in a state of uncertainty, but you would hardly say that you were facing risk. The notion of risk, therefore, involves both uncertainty and some kind of loss or damage that might be received. We could express this idea, compactly, in the form of a symbolic equation,

$$\text{Risk} = \text{Uncertainty} + \text{Damage}.$$

This equation thus expresses our first distinction, that between risk and uncertainty. As a second, it is very useful, especially in understanding the controversies surrounding energy facilities, to draw a distinction between the ideas of risk and hazard. This is the subject of the next section.

0.2.2 THE DISTINCTION BETWEEN RISK AND HAZARD

In the dictionary (Reference 0-1) we find hazard defined as "a source of danger." Risk is the "possibility of loss or injury" and the "degree of probability of such loss." Hazard, therefore, simply exists as a source. Risk includes the likelihood of conversion of that source into actual delivery of loss, injury, or some form of damage. This is the sense in which we use the words. As an example, the ocean can be said to be a hazard. If we attempt to cross in a rowboat, we undergo great risk. If we use the Queen Elizabeth, the risk is small. The Queen Elizabeth thus is a device that we use to safeguard us against the hazard, resulting in small risk. As in Section 0.2.1, we express this idea symbolically in the form of an equation,

$$\text{Risk} = \frac{\text{Hazard}}{\text{Safeguards}} .$$

This equation further brings out the thought that we may make risk as small as we like by increasing the safeguards but may never, as a matter of principle, bring it to zero. Risk is never zero, but it can be small.

In a similar way the radioactivity contained in the core of a nuclear reactor is a hazard as defined above and, were it to be spread through the surrounding population, it could do much damage. The same could be said of the water in Lake Michigan. The question in both cases is: Given the hazard, what is the risk? Following this come additional questions: Shall we institute further safeguards? How much effort is required to implement these safeguards? How much do they reduce risk? Could this effort be more productively applied to gain greater risk reduction or benefit creation in some other way or in some other application?

To answer such questions in a rational way, and to have any hope of reaching a workable consensus on them, it is obvious that we need a way to define the idea of risk in a precise, quantitative, and reproducible way. This is the topic of the next section in which we follow generally the approach of Reference 0-35.

0.3 THE QUANTITATIVE DEFINITION OF RISK (LEVEL ONE)

0.3.1 THE "SET OF TRIPLETS IDEA"

In analyzing risk we are attempting to envision how the future will turn out if we undertake a certain course of action (or inaction). Fundamentally, therefore, a risk analysis consists of an answer to the following three questions:

- What can happen, i.e., what can go wrong?
- How likely is it that this will happen?
- If it does happen, what are the consequences?

To answer these questions we would make a list of outcomes or "scenarios" as suggested in Table 0.3-1.

TABLE 0.3-1

SCENARIO LIST

Scenario	Likelihood	Consequence
s_1	p_1	x_1
s_2	p_2	x_2
.	.	.
.	.	.
.	.	.
s_N	p_N	x_N

The ith line in this table can be thought of as a triplet

$$\langle s_i, p_i, x_i \rangle$$

where

s_i = a scenario identification or description,

p_i = the probability of that scenario, and

x_i = the consequence or evaluation measure of that scenario, i.e., the measure of damage.

If this table contains all the scenarios we can think of, we can then say that it (the table) is the answer to the questions and therefore is the risk. More formally, using braces, $\{\ \}$, to denote "set of" we can say that the risk, R, "is" the set of triplets,

$$R = \{\langle s_i, p_i, x_i \rangle\},\ i = 1, 2, \ldots N.$$

This definition of risk as a set of triplets is our first level definition. We shall refine and enlarge it later. For now, let us show how to give a pictorial representation of risk.

0.3.2 RISK CURVES

Imagine now, in Table 0.3-1, that the scenarios have been arranged in order of increasing severity of damage. That is to say, the damages x_i obey the ordering relationship

$$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_N.$$

Now add to the table a fourth column in which we write the cumulative probability, adding from the bottom (Table 0.3-2). The cumulative probability is represented by the upper case P as shown.

TABLE 0.3-2

SCENARIO LIST WITH CUMULATIVE PROBABILITY

Scenario	Likelihood	Consequences	Cumulative Probability
s_1	p_1	x_1	$P_1 = P_2 + p_1$
s_2	p_2	x_2	$P_2 = P_3 + p_2$
.	.	.	.
.	.	.	.
.	.	.	.
s_i	p_i	x_i	$P_i = P_{i+1} + p_i$
.	.	.	.
.	.	.	.
s_{N-1}	p^{N-1}	x_{N-1}	$P_{N-1} = P_N + p_{N-1}$
s_N	p_N	x_N	$P_N = p_N$

If we now plot the points $\langle x_i, P_i \rangle$, we obtain the staircase function shown as a dashed line in Figure 0.3-1.

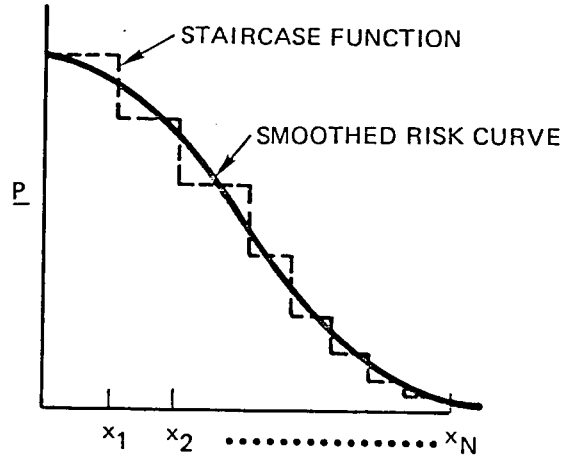


Figure 0.3-1. Risk Curve

Let us next note that what we called "scenarios" in Table 0.3-1 are really categories of scenarios. Thus, for example, the scenario "pipe break" actually includes a whole category of different kinds and sizes of breaks that might be envisioned, each resulting in a slightly different damage, x.

Thus, we can argue that the staircase function should be regarded as a discrete approximation to a continuous reality. Thus, if we draw in a smoothed curve through the staircase, we can regard that curve as representing the actual risk. Hence we call it the "risk curve." When a risk curve is plotted on a log-log scale it takes on the characteristic concave downward shape shown in Figure 0.3-2. In this case, the asymptotes have the interpretation of "maximum possible damage" and "probability of any damage at all."

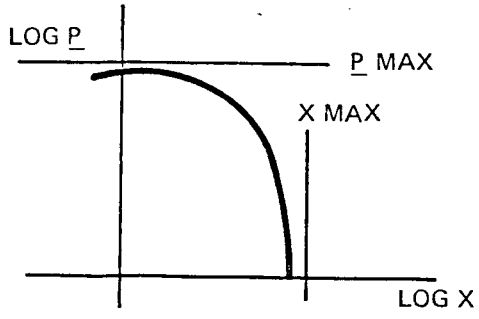


Figure 0.3-2. Risk Curve on a Log-Log Scale

0.3.3 COMMENTS ON THE DEFINITION

One often hears it said that "risk is probability times consequence." We find this definition misleading and prefer instead, in keeping with the set of triplets idea, to say that "risk is probability and consequence." In the case of a single scenario the probability times consequence viewpoint would equate a low probability high damage scenario with a high probability low damage scenario; clearly not the same thing at all.

In the case of multiple scenarios the probability times consequence view would correspond to saying that the risk is the expected value of damage, i.e., the mean of the risk curve--a single number. The point of view taken in this study is that it is not the mean of the curve but the curve itself which is the risk. A single number is not a large enough concept to communicate the idea of risk. It takes a whole curve.

Now actually a curve is not a large enough concept either. It takes a whole family of curves to fully communicate the idea of risk. This is the basis for the Level 2 definition to which we shall come shortly. First we pick up some further points in connection with the Level 1 definition.

0.3.4 MULTIDIMENSIONAL DAMAGE

In the case of a nuclear plant, as in many other applications of risk analysis, it is appropriate to identify different types of damage, e.g., loss of life and loss of property. In these cases, the damage, x, can be regarded as a multidimensional or vector quantity rather than a single scalar. The risk curve then can be thought of as a risk surface over the multidimensional space as suggested in Figure 0.3-3.

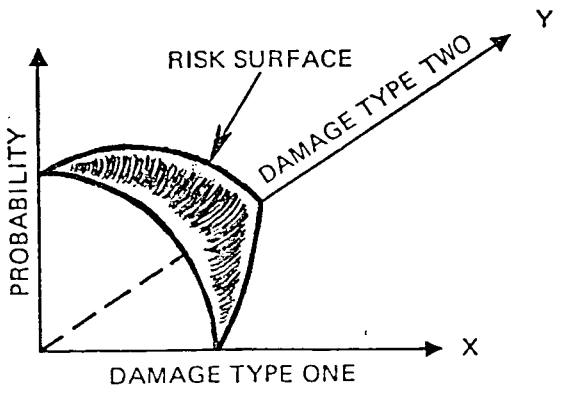


Figure 0.3-3. Risk Surface

Alternatively, the risk can be expressed with a series of figures like Figure 0.3-2, one for each type of damage. This is the procedure adopted in WASH-1400. We shall also adopt it here.

0.3.5 COMPLETION OF THE SCENARIO LIST

One of the criticisms that has been made of the Reactor Safety Study (Reference 0-2) may be paraphrased essentially as follows:

> "A risk analysis is essentially a listing of scenarios. In reality, the list is infinite. Your analysis, and any analysis, is perforce finite, hence incomplete and inherently limited. Therefore, no matter how thoroughly and carefully you have done your work, I am not going to trust your results. I'm not worried about the scenarios you have identified, but about those you haven't thought of. Thus I am never going to be satisfied."

The critic here has a valid point about risk analysis. The implied conclusion, that we should not build nuclear reactors, is not valid. For whatever course of action, or nonaction, is proposed in place of building reactors must also be subject to a risk analysis. That risk analysis will also have the same inherent limitation as the Reactor Safety Study. That limitation in itself, therefore, cannot be used to argue for one branch of the decision tree over another since it applies to all branches.

Nevertheless, the critic has made a good point about the risk analysis formalism. Let us see, therefore, what can be done to improve the formalism to address this point.

One tactic that comes to mind, in light of the fact that the s_i are categories of scenarios, is to include another category, s_{N+1}, to the list. We may call this category the "other" category. By definition, it contains all scenarios not otherwise included in the list. Correspondingly, we would now say that the risk is the set of triplets:

$$R = \{\langle s_i, p_i, x_i \rangle\}, \quad i = 1, 2, \ldots N+1$$

which includes all the scenarios we have thought of and also an allowance for those we have not thought of. Thus extended, the set of scenarios may be said to be logically complete.

It seems at first glance that what we have done here is simply a logical trick which does not address the fundamental objection. It is a little bit more than a trick, however. For one thing, it takes the argument out of the verbal realm and into the quantitative realm. Instead of the emotional question, "What about the things you haven't thought of?" We have, "What probability should we assign to the residual category s_{N+1}?"

Once the question has been phrased in this way, we can proceed in a rational manner, in the same way we do to assign any probability. We ask: What evidence do we have on this point? What knowledge, what relevant experience? In particular, we note that one piece of evidence is always present--namely, scenarios of the type s_{N+1} have not yet occurred, otherwise we would have included them elsewhere on the list.

How much is this piece of evidence worth? This is a question that can be answered rationally and objectively within the framework of the theory of probability, using Bayes' theorem. Within this framework the answer is not a matter of opinion but a matter of logic. We shall return to this point particularly, in Section 0.14.2, and more generally to the question of the "other" category in Section 0.19. Right, now, before going any further, it is important to explain the sense in which we are using the word "probability." This is the purpose of the next section.

0.4 PROBABILITY

People have been arguing about the meaning of probability for at least 200 years, since the time of Laplace and Bayes. The major polarization of the argument is between the "objectivist" or "frequentist" schools which view probability as something external, the result of repetitive experiments, and the "subjectivists" who view probability as an expression of an internal state--a state of knowledge or a state of confidence.

In this study we adopt the point of view that both schools are right; they are just talking about two different ideas. Unfortunately they both use the same word--which seems to be the source of most of the confusion. We need, therefore, to give each idea the dignity of its own name. We do this by calling one idea "frequency" and the other "probability." In the next section we shall carefully explain the distinction we make in this study between these two words. The reader should also pay careful attention because this distinction is fundamental to understanding the methodology, language, and numerical results of this study.

0.4.1 THE DEFINITION OF PROBABILITY AND THE DISTINCTION BETWEEN PROBABILITY AND FREQUENCY

What the objectivists are talking about we shall call "frequency." What the subjectivists are talking about we shall call "probability." Thus, "probability" as we shall use it is a numerical measure of a state of knowledge, a degree of belief, a state of confidence. "Frequency," on the other hand, refers to the outcome of an experiment of some kind involving repeated trials. Thus frequency is a "hard" measurable number. This is so even if the experiment is only a thought experiment or an experiment to be done in the future. At least in concept then, a frequency is a well-defined, objective, measurable number.

Probability, on the other hand, is a notion of a different kind. Defined as a number used to communicate a state of mind, it is thus inherently subjective and changeable as new information arrives. To make this notion useful, we must clearly define the correlation between the numbers and the state of mind.

This can be done in several ways. The most direct, however, is to use frequency in the following way. Suppose we have a lottery basket containing coupons numbered from 1 to 1,000. Suppose the basket is thoroughly mixed and that you are about to draw a coupon blindfolded.

We ask: Will you draw a coupon numbered 632 or less? With respect to this question you experience a certain state of confidence. Similarly, I experience a state of confidence with respect to this same question. Let us agree to call this state of confidence, "probability 0.632," equal to the frequency of such draws in an infinitely repeated experiment. Now we both know exactly what we mean by $p = 0.632$.

So if you now say that the probability of your latest business venture succeeding is 0.632, I know exactly what your experiential state of confidence is. We have communicated!

In the same way, we define or "calibrate" the entire probability scale, from zero to one, using frequency as a standard of reference. Note that the process used here is entirely parallel to the way by which we define "red," "chair," "seventeen," and all other words or symbols.

This method of definition shows the intimate connection between probability and frequency. This connection needs to be recognized always and at the same time not allowed to obscure the fundamental difference. Frequency is used to calibrate the probability scale in a "bureau of standards" sense. Once the calibration is established, we then use probability to discuss our state of confidence in areas where we are dealing with one-time events and have no frequency information at all.

In this way we liberate ourselves from the restrictions of the relative frequency school of thought (e.g., that only mass repetitive phenomena can be analyzed probabilistically) and instead create for ourselves a systematic, disciplined theory and language for dealing with rare events, for quantifying risks, for making decisions in the face of the uncertainties which are inevitably present in decision situations, for making those decisions and for taking the consequent actions with the knowledge that these are the best decisions and actions possible in light of all the information available to us.

This then is the definition adopted in this report. For additional insight, we quote the following paragraph from unpublished notes by E. T. Jaynes:

> "Probability theory is an extension of logic, which describes the inductive reasoning of an idealized being who represents degrees of plausibility by real numbers. The numerical value of any probability (A/B) will in general depend not only on A and B, but also on the entire background of other propositions that this being is taking into account. A probability assignment is "subjective' in the sense that it describes a state of knowledge rather than any property of the 'real' world; but it is completely 'objective' in the sense that it is independent of the personality of the user; two beings faced with the same total background of knowledge must assign the same probabilities."

and, as further elaboration, cite the following paragraph by A. DeMorgan.*

> "We have lower grades of knowledge, which we usually call degrees of belief, but they are really degrees of knowledge ... It may seem a strange thing to treat knowledge as a magnitude, in the same manner

*Further discussion of the foundations of the subjectivistic theory can be found in References 0-3 through 0-6.

as length, or weight, or surface. This is what all writers do who treat of probability, and what all their readers have done, long before they ever saw a book on the subject ... By degree of probability we really mean, or ought to mean, degree of belief ... Probability then, refers to and implies belief, more or less, and belief is but another name for imperfect knowledge, or it may be, expresses the mind in a state of imperfect knowledge."

0.4.2 THE DISTINCTION BETWEEN PROBABILITY AND STATISTICS

Corresponding to the above definitions of frequency and probability as numbers, we may say that statistics, as a subject, is the study of frequency type information. That is, it is the science of handling experimental data. On the other hand, probability as a subject we might say is the science of handling the lack of experimental data.

Thus, one often hears it said that we cannot use probability because we have insufficient data. In light of our current definitions, we see that this is a misunderstanding. When one has insufficient data, there is nothing else one can do but use probability.

0.4.3 COMMENTARY ON THE DEFINITIONS OF FREQUENCY AND PROBABILITY--AN EXAMPLE

We shall give a simple, tutorial example to further clarify the concept of probability and to indicate how we make the distinction between probability and frequency.

If I tossed a coin and asked you for the probability of it coming up heads, you will of course say .5. If I tell you that I have just tossed the coin ten times and the frequency of heads was .7, and now ask for the probability of a head on the next toss, you will still very likely say .5. If, however, I tell you that I have tossed it a hundred times and the frequency was .7, 70 heads, now you will begin to suspect that the coin is not equally balanced, and the probability you give for heads on the next toss may move up--say to .6.

If I tell you the frequency has been .7 in ten thousand trials, you will be convinced and will assign a probability of .7 to the next toss.

This example helps bring out the distinction between probability and frequency. The coin has not changed during this example, but your state of knowledge about the coin has--and this is reflected in your changing probabilities from .5 to .6 to .7. On the other hand, I knew the coin was unbalanced to begin with--and my probability was .7 all along.

Which of us was right? Both of us were right. Your probability reflected your state of knowledge and mine reflected mine. As such reflections, both were 100% accurate.

0.4.4 THE MEANING OF "THE" PROBABILITY--RELATION TO THE PHILOSOPHICAL BASIS FOR RISK ASSESSMENT

But what about "The" probability. Here our language plays tricks on us. There is no such thing as "The" probability--as if it were something external--there is only "your" probability, based on the evidence you have and "my" probability based on the evidence I have.

But, you say, suppose we toss the coin N times and plot the frequency of heads, $\phi(N)$, as a function of N. As N gets larger and larger, $\phi(N)$ will approach a limiting value. That value is "The" probability. Well, you could define it so. We find it more useful to call that limiting value "The" frequency--the frequency in an infinite experiment--and reserve the word probability to refer to the state of confidence at any moment.

There is another sense, however, in which it can be said that there is a "The" probability. This is in the sense of the last sentence of Jaynes' definition. Any two idealized beings, "rational" beings, given the same total background of evidence and experience must assign the same numerical value of probability to a given proposition. That value could be said to be "The" probability. It is independent of the personality of the user, hence "objective."

Thus, as idealized beings we have you and I, and we say that each of us is acting "rationally," "coherently" or "objectively" to the extent that our probability assignments follow the formalized rules of the theory of probability. Thus if we are both acting rationally and we both have the same information, we will both assign the same probabilities. If we do not assign the same probabilities, then either one or more of us is not coherent, or we do not have the same total background of evidence and information.

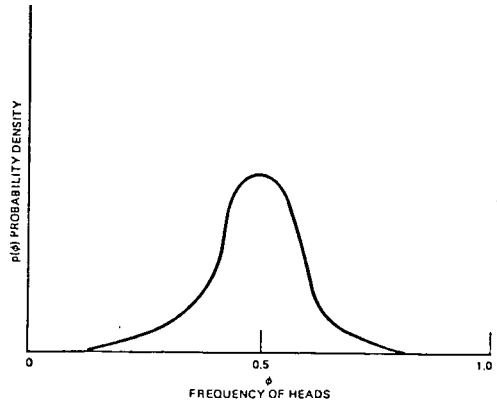
This idea of rationality or coherence is the philosophical cornerstone of our approach to risk assessment. For, if two rational beings, given the same body of information, will evaluate probabilities the same way, then by our definition, Section 0.3, they will also evaluate risk the same way. Thus a given specific body of information will imply and require a specific quantitative value of risk, and this value is objective and independent of who evaluates it.

0.4.5 TWO METHODS FOR DISCUSSING UNCERTAINTY: THE "PROBABILITY OF FREQUENCY" FRAMEWORK

There are two basic methods for quantifying uncertainty, corresponding to two different questions. We illustrate these in the context of another coin flipping example. In "Method One" we ask: What is the probability of a head on the next toss? Alternately, in "Method Two" we say: I am going to toss the coin 10,000 times. We then ask: What is the frequency, i.e., the percentage of heads going to be?

In Method One we answer simply with a number, our state of confidence on the prospect of a head on the next toss, as reflected for example in the odds we would take in a bet.

In Method Two we are asked to predict the outcome, ϕ, of an experiment to be done in the future. Since we do not know this outcome, we express our prediction in the form of a probability curve against it, e.g.:



Thus in the second method we are led to the notion of a probability curve against frequency as a way of, or a framework for, expressing our state of knowledge.

This notion of probability of frequency and this distinction between Method One and Method Two are central to the understanding of this study. They are discussed further in later sections and will be of use to us in the next section in extending the definition of risk. We therefore expand upon this notion in the following section.

0.4.6 REASONS FOR INTRODUCING THE NOTION OF PROBABILITY OF FREQUENCY

- Reason 1: First, let us note, coming back to the coins, that the answer to the first question can be derived from the answer to the second. Thus, having given the probability of frequency curve $p(\phi)$ we would then, consistent with that, express our probability of heads on the next try as

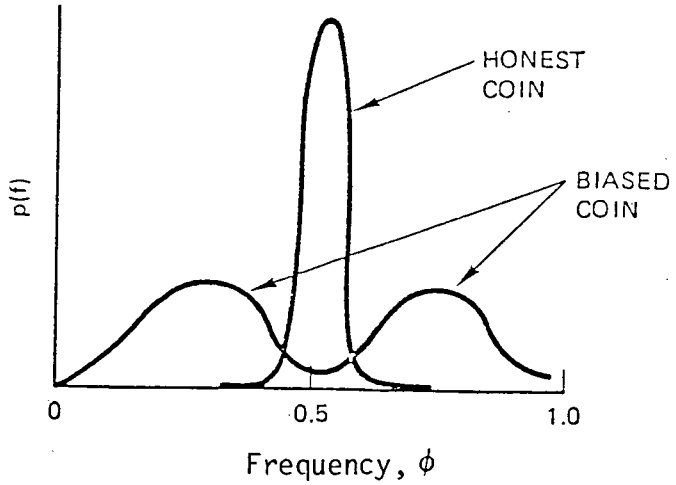
$$p(\text{head}) = \int_0^{1.0} \phi p(\phi) d\phi.$$

  That is, the expected frequency, in Method Two, is the probability in Method One. We see thus that the second method includes or encompasses the first. The reverse cannot be said. Thus Method Two is a fuller, more complete way of talking about uncertainty.

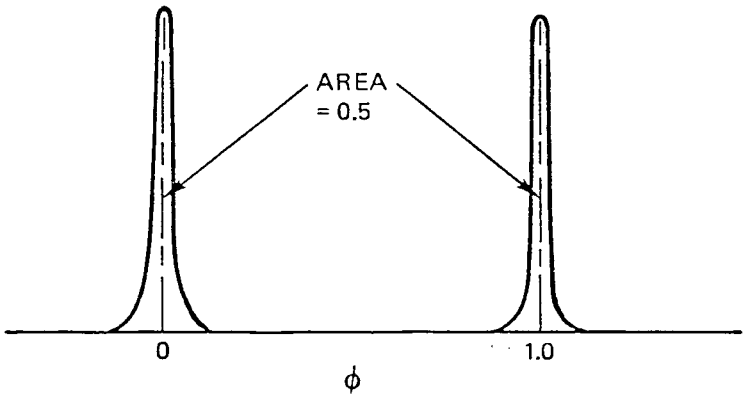
- Reason 2: Moreover, the need for a more complete way of talking is evident from a perusal of existing safety and probabilistic studies. Once a probability has been calculated, people inevitably ask: How accurate is that probability? How confident are you in that number? In response to such questions authors of probabilistic studies have been led to introduce such notions as confidence bounds on the probability and probability of probability, etc.

In the context of our definition, Section 0.4.1, phrases such as 'probability of probability' make no sense. It is not meaningful to ask how confident you are in your state of confidence. It is meaningful, however, to ask how confident you are in your prediction of the result of a repetitive experiment. Thus the probability of frequency notion provides a suitable framework within which the inevitable questions can be given a quantitative meaning.

- Reason 3: Another indication of the need for the Method Two approach becomes visible in the following situation: If I tell you this is an honest coin and ask for the probability of a head on the next flip, you will promptly answer .5. If I now tell you the coin is biased, but don't tell you in which direction, you will again answer .5 but this time after some hesitation. Your state of knowledge with respect to the outcome of the toss is the same--you would make the same bet--but somehow there is a difference in the way your belly feels. The probability of frequency notion is an appropriate one for bringing out this difference. For in the first case, the probability versus frequency curve would be strongly clustered about $\phi = .5$. In the biased coin case, the curve would be very broad--in fact, it would be bimodal with a strong dip at $\phi = .5$.



Let us take an extreme example of this. Suppose I tell you that the coin is the same on both sides--but don't tell whether it has two heads or two tails. Your state of knowledge, assuming of course that you are absolutely certain of my veracity, would be expressed precisely by a probability versus frequency curve consisting of two delta functions, located at $\phi = 0.0$ and $\phi = 1.0$, and each including an area of 0.5.

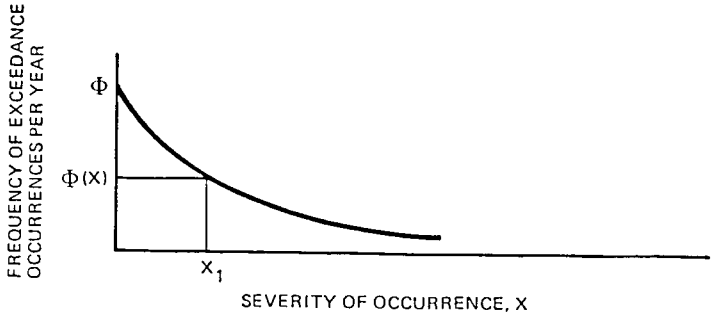


- Reason 4: In the course of this safety study, we commonly seek to calculate the likelihood that a system will fail from historical data on the frequency of failure of the system's components.

  Historical data available is often sparse or not fully relevant. The published compilations of component failure rates, therefore, usually quote these rates with "high" and "low" as well as "best estimate" values.

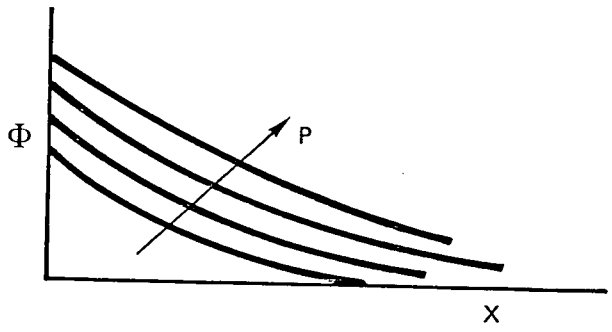
  Thus, standard industry compilations of failure data already express our degree of knowledge of component reliability in one way or another as a probability curve against frequency of failure. Given this form of component information, it is natural in this study to express the system reliability also in the form of probability of failure frequency.

- Reason 5: For events which are termed "external events" such as earthquakes, the frequency of occurrence is usually expressed in a "frequency of exceedance" diagram as follows:



  Here the ordinate, $\Phi(X_1)$ over the point X_1, is the frequency with which an event of severity X_1 or greater occurs. Severity would be measured in the case of earthquakes, for example, as peak ground accelerations.

Now, if we had a great deal of measured data, we would actually know the above frequency of exceedance curve, $\Phi(X)$. In actual fact, we have uncertainty about this curve and, therefore, express this uncertainty by putting forth a family of curves, with probability, P, as parameter.



Such a family of curves thus expresses our state of knowledge of the seismicity activity of a particular location. It is another example of usage of the probability of frequency concept.

Our probability curves are subject, of course, to change as new evidence is accumulated. This revision must be done coherently, and the appropriate tool for this is Bayes' theorem, which we shall describe later.

0.4.7 THE DISTINCTION BETWEEN FREQUENCY DISTRIBUTIONS AND PROBABILITY DISTRIBUTIONS

We now use our definitions to distinguish two further situations which are often badly confused.

Let x denote the height of an individual person selected at random from a population. If we now measure the height of each person we can draw a frequency distribution showing what fraction of the population falls in each height increment.

If the population is large, we can by a limiting process, express this distribution as a continuous curve, a frequency density distribution, $\phi_X(x)$, Figure 0.4-1.

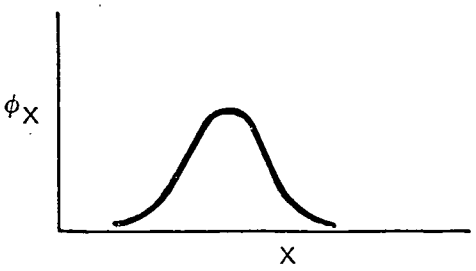


Figure 0.4-1. Population Variability Curve

The units of $\phi_X(x)$ are thus frequency per unit x, or fraction of population per unit height. This curve is an experimental quantity. It portrays the variability of the population--a measureable quantity. The value of x therefore varies with the individual selected. It is a truly fluctuating or random variable.

Now contrast the situation where we pick a specific individual (say Joe) in the population and ask what his height, x_{Joe}, is. Since we do not know his height for sure, we express our state of knowledge about it in the form of a probability density function as in Figure 0.4-2.

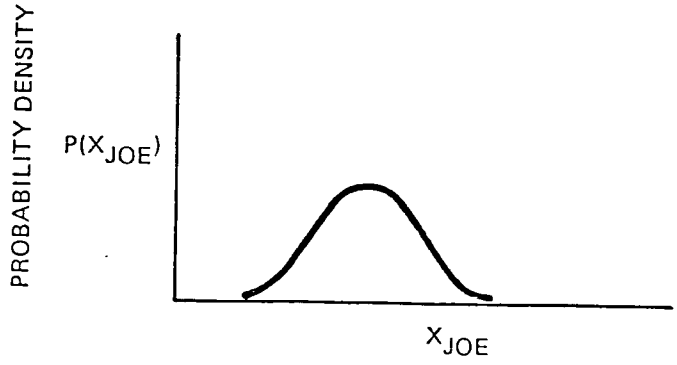


Figure 0.4-2. State-of-Knowledge Curve

The units here are probability per unit height. In this case x_{Joe} is not a random or fluctuating variable. x_{Joe} is a definite number. It is just that we don't know what it is. This is a very different situation from Figure 0.4-1.

Thus Figure 0.4-1 is the frequency distribution of a random, or fluctuating, variable. Figure 0.4-2 is the probability distribution for a fixed, nonfluctuating, but unknown quantity.

This distinction between population variability curves and state of knowledge curves must be made when we analyze data on failure rates and initiating event frequencies from our specific plant and from other plants and other industries.

0.5 THE LEVEL 2 DEFINITION OF RISK

Having defined the meaning of the word probability, as we shall use it, we now return to the definition of risk.

When one presents a risk curve as the result of an analysis, one of the things that invariably happens is that someone asks: "How confident are you in that curve?"* In view of our usage of the term probability, the risk curve already expresses our state of confidence. It appears thus as if the question is asking: "How confident are you in your state of confidence?" In this form the question seems like a nonsense question. However, there is a valid thought behind it. What we need to do, therefore, is to expand our framework somehow, in such a way that within the enlarged framework the question can be given a precise meaning and then be answered precisely.

0.5.1 RISK CURVES IN FREQUENCY FORMAT

For this purpose we make use of the Method Two, probability of frequency idea, in the following way. We imagine a thought experiment in which we undertake the proposed course of action, or inaction, many, many times. At the end of this experiment we will be able to look back at the records and ask: "How frequently did scenario s_i occur?" This frequency will then be an experimentally measured number. Let us denote it by ϕ_i. Its units are occurrences per trial.

At the end of the experiment, therefore, we will have the set of numbers, ϕ_i, and the set of triplets

$$\{\langle s_i, \phi_i, x_i \rangle\} \quad i=1,\ldots N+1.$$

As in Section 0.3.2, we could then compute the cumulative frequency:

$$\Phi_i = \sum_{x_j \geq x_i} \phi_j$$

(where the sum is over all scenarios having damage equal to or greater than x_i). Also as in Section 0.3.2, we could now plot Φ versus x, obtaining Figure 0.5-1 which we refer to as a risk curve in frequency format. This whole curve may be regarded as the outcome of our thought experiment.

*For example, the major criticism by the Lewis Committee (Reference 0-7) of the Reactor Safety Study (Reference 0-2) had to do with the uncertainty of the risk curves.

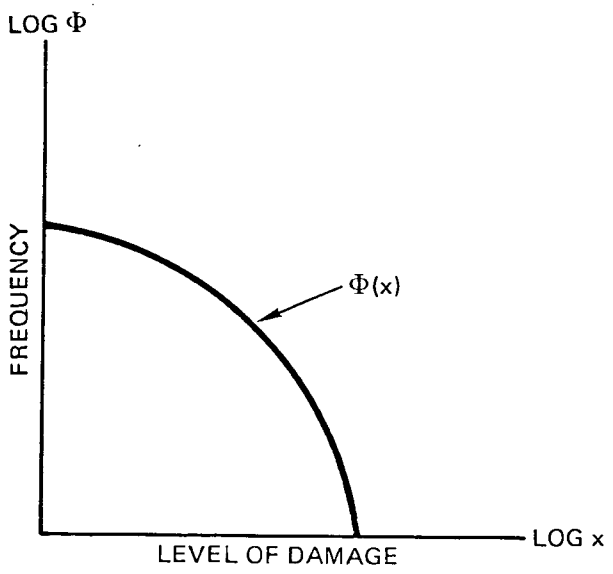


Figure 0.5-1. Risk Curve in Frequency Format

0.5.2 INCLUSION OF UNCERTAINTY

Now since we have not yet actually done the thought experiment of the previous section, we have uncertainty about what its outcome would be. The degree of uncertainty depends upon our total state of knowledge as of right now; upon all the evidence, data, information, and experience with similar courses of action in the past. We seek therefore to express this uncertainty using, naturally, the language of probability.

Since the thing we are uncertain about is a curve, $\Phi(x)$, we express the uncertainty by embedding this curve in a space of curves and erecting a probability distribution over this space.

Pictorially, this is represented by a diagram of the form of Figure 0.5-2.

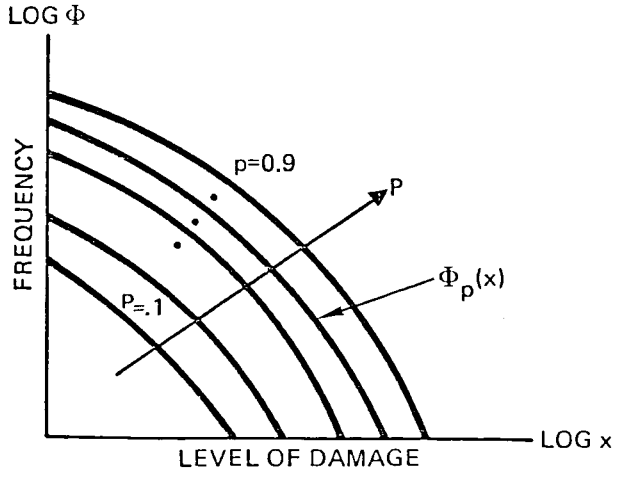


Figure 0.5-2. Risk Curve in Probability of Frequency Format

This figure is what we call a "Risk Curve in Probability of Frequency Format" or alternatively a "Risk Diagram." It consists of a family of curves, $\Phi_P(x)$, with the parameter being the cumulative probability. To use this diagram we would, for example, enter with a specific x value and choose, say the curve P = 0.90. The ordinate of this curve, $\Phi_{.90}(x)$, is then the 90th percentile frequency of x. That is to say, we are 90% confident that the frequency with which damage level x or greater occurs is not larger than $\Phi_{.90}(x)$.

Figure 0.5-2 is the pictorial form of our Level 2 definition of risk. It is of interest to express this definition also in terms of the set of triplets idea.

0.5.3 SET OF TRIPLETS INCLUDING UNCERTAINTY

In listing our set of triplets for a proposed course of action, suppose we now acknowledge that, to tell the truth about it, we do not know the frequency with which scenario category s_i occurs. We would then express our state of knowledge about this frequency with a probability curve:

$P_i(\phi_i)$ = probability density function for the frequency ϕ_i, of the ith scenario.

Thus we now have a set of triplets in the form:

$$R = \{\langle s_i, p_i(\phi_i), x_i \rangle\} \qquad (0.5\text{-}1)$$

which set of triplets we could say is the risk including uncertainty in frequency.

From the set (0.5-1) we can construct the risk family Figure 0.5-2 by cumulating frequencies from the bottom in a manner entirely parallel to that used in Section 0.3.2.

Similarly, if there is uncertainty in the damage also, we would have the set of triplets

$$R = \{\langle s_i, p_i(\phi_i), q_i(x_i) \rangle\} \qquad (0.5\text{-}2)$$

or more generally,

$$R = \{\langle s_i, p_i(\phi_i, x_i) \rangle\} \qquad (0.5\text{-}3)$$

using a joint distribution on ϕ_i, x_i. In cases (0.5-2) or (0.5-3) we can also construct the family of risk curves. It is conceptually and computationally much cleaner, however, to stick with the form (0.5-1) if possible. One way of doing this is to make the damage level part of the definition of the scenario. There is then no uncertainty in the x_i.

definition of the scenario. There is then no uncertainty in the x_i. All the uncertainty is then in the functions $p_i(\phi_i)$.*

0.5.4 COMMENTS ON THE LEVEL TWO DEFINITION

Figure 0.5-2, or equivalently Equations (0.5-1), (0.5-2), or (0.5-3), constitutes our expanded, Level 2 definition of risk. Observe that the Level 1 and Level 2 definitions of risk correspond to Method One and Method Two of Section 0.4.3. Observe also that Level 2 includes the Level 1 definition in the sense that the expected frequency, $\Phi(x)$, at any x is the probability P(x) at that x. Thus we have lost nothing in going from Level 1 to Level 2 and have gained the ability to explicitly include uncertainty in the calculation of risk. This ability will be particularly useful where the fundamental input data on failure rates is uncertain, as it always is.

Finally, we remark that our Level 2 definition of risk is "subjective" in the same sense that probability itself is (i.e., it expresses an internal state of knowledge rather than any property of the "real" world).

On the other hand, as Jaynes points out in the last sentence of his definition (Section 0.4.1), any two rational observers given the same totality of information must calculate the same Figure 0.5-2, and thus agree on the quantification of risk.

In this sense, we may say the Level 2 definition of risk is "absolute" and "objective." It depends upon the evidence at hand, but other than that is independent of the personality of the user. Any two rational beings given the identical evidence must assess the risk identically.

0.5.5 "CUTTING" THE FAMILY OF RISK CURVES

In the Level Two definitions, Figure 0.5-2, let us draw a vertical line which cuts the family of curves at a specific damage level, x_1. The intersection of this line with the family leads to the curve of P vs $\Phi(x_1)$, a cumulative probability curve as shown in Figure 0.5-3. This curve may be differentiated to yield a bell curve or probability density function, $p\,\Phi(x_1)$, which expresses our state of knowledge about the frequency with which events of level x_1 or greater occur.

*The price paid for this neat bit of legerdemain is that the distributions p_i may now not be independent from one i to another. That is, our state of knowledge of ϕ_i may now be dependent upon our knowledge of ϕ_j. Of course, this could also be true without the legerdemain. It is something to be aware of in either case.

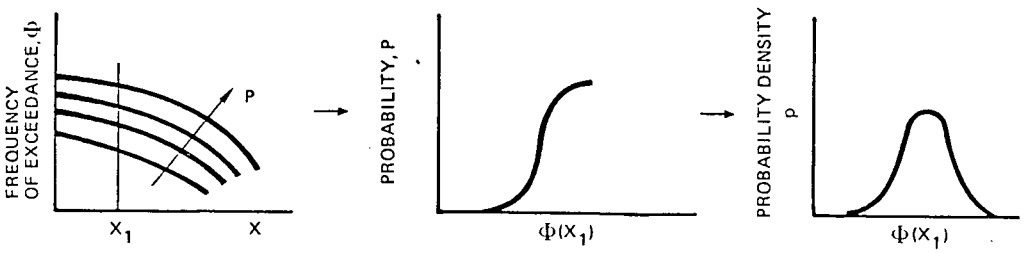


Figure 0.5-3. Development of "Cut Curves" From Level Two Family

Curves of this type, called "cut curves", will be useful in Section 8 for identifying the key physical contributors to the family of risk curves. Section 8 will also make use of the Level One definition of risk in discussing the contribution to risk from various "internal" events. This is used, for purposes of clarity, as an intermediate stage in building to the full Level Two presentation.

PROBABILISTIC RISK ASSESSMENT

METHODOLOGY

Part 2

MODELING AND ANALYSIS

0.6 IDENTIFYING AND STRUCTURING THE SCENARIO LIST

In light of our definition of risk as a set of triplets, it is clear that the first step in risk analysis must be to make a list of possible scenarios. As a matter of principle, we wish to make this list as long as possible, i.e., to think of and separately identify as many scenarios as we can. In the case of a nuclear plant, as we shall see, the list can run literally into the billions. It is necessary, therefore, to develop methods for identifying scenarios and for organizing and structuring the list so that it can be comprehended. These methods are the subject of this section.

We begin in Section 0.6.1 by following a deductive line of thought that leads to identification of possible initiating events. The next two sections discuss the event tree method for organizing the myriad of possible scenarios which can emanate from any given initiating event. Section 0.6.2 discusses the "plant" event tree which follows the scenarios up to the point where either the reactor is stabilized, or plant damage has occurred. At this point, as suggested in Figure 0.6-1, a coalescence of scenarios or "pinchpoint" occurs in that, given that a certain state, y_j, of plant damage has occurred, the remainder, or downstream portion of scenarios is the same irrespective of how that damage state was arrived at.

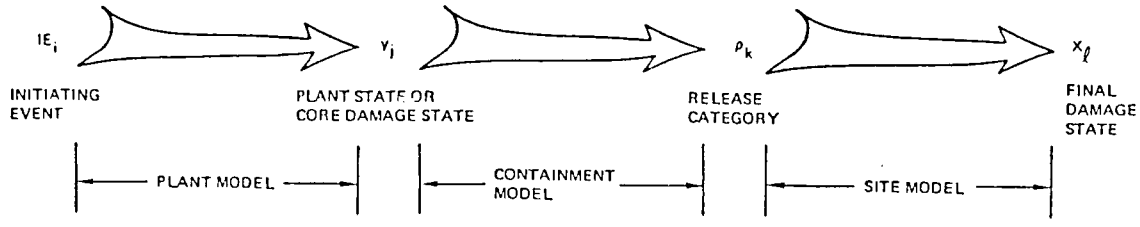


Figure 0.6-1. Structuring of Scenarios--Relationship of Pinch Points

The next portion of the scenarios is modeled by a "containment event tree" which follows the progress of the scenarios through the containment from the plant damage state to the occurrence or nonoccurrence of a release of radioactivity to the environment. Thus, the entry states to the containment event tree are the plant damage states, i.e., the exit states from the plant event tree.

The exit states from the containment are called "release categories" each of which specifies a certain quantity and mix of radioisotopes released. At this point another coalescence occurs in that the effects in the environment of a given category of release are the same irrespective of the particular scenario that led to that release category.

The environmental effects are then studied by a "site model" which takes the release category as its input event, follows the movement of the radioactivity, and computes the final damage state, x_ℓ, in terms of public health impacts and property damage.

0.6.1 IDENTIFYING SCENARIOS, INITIATING EVENTS, AND THE MASTER LOGIC DIAGRAM

We turn to the task of identifying scenarios that might lead to a release of radioactivity to the evironment. In particular, we wish to identify events which have the potential to initiate such scenarios. We refer to such events as "Initiating Events" or "IEs." Each IE is the root of a branching family of scenarios, some of which may lead to release.

As an aid to identifying these IEs and these scenarios, we use the device of a "Master Logic Diagram" shown in Figure 0.6-2. This diagram traces the thought process that follows from the question "How can a significant release to environment occur?" Level I in the diagram represents such a release. Level II then says that if such a release occurs, it must originate either in a damaged core or in a noncore source of radiation such as the spent fuel storage pool or the gaseous, liquid, and solid waste facilities. Past experience and anlysis (Reference 0-34) have clearly shown that releases from the core are by far the major source of risk at a nuclear power plant and, therefore, the remainder of the Master Logic Diagram emphasizes the core branch.

Level III then expresses the fact that a release from the core to the environment can occur only if both the core is damaged and simultaneously the containment function is failed.

The next question is "How can core damage occur?" Level IV answers that this can occur only if either there is an excess production of power in the core, or if there is not excess power, that there is a loss of cooling capability.

Following the excess power branch, the only way this could occur is if there were an increase in reactivity. We are thus led, at Level VII, to Category 13 of initiating events, "Core Power Increase." The next question is "What IEs can lead to core power increase?" This question and others like it can be answered out of experience and understanding of the reactor phenomena, and out of a study of compilations of Licensing Event Reports (LERs), Final Safety Analysis Reports (FSARs), and other pertinent literature. In the present case the IEs which have been identified for the plant are discussed in Section 1.1 in detail.

Following the loss of cooling branch we see at Level V that loss of cooling can occur only if there is a failure of the primary coolant boundary, and therefore a loss of coolant, or if there is a loss of heat removal capability which, in turn, will lead to opening of pressure relief valves and thus also to a loss of coolant.

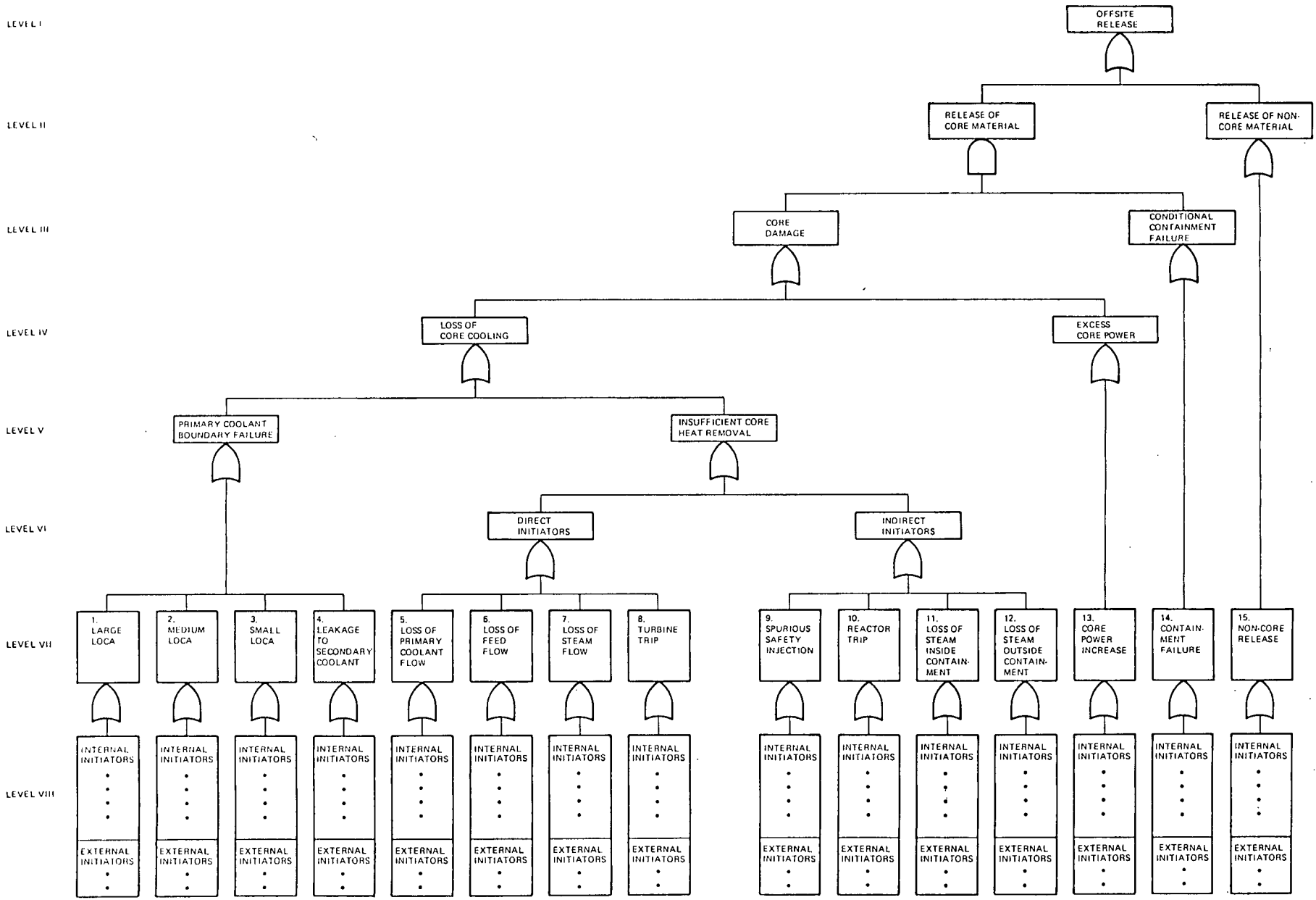


Figure 0.6-2. Master Logic Tree

Failure of the primary coolant boundary can occur in various ways which, in Figure 0.6-2, are grouped into four categories: large, medium, small LOCA, and leakage to secondary coolant. Loss of heat removal can result from IEs which are divided, at Level VI, into two types: "direct" and "indirect." Under each of these are four further categories of IEs as shown at Level VII. The individual IEs contributing to the Level VII categories are listed in Section 1, Table 1.1.1-1. They are divided there into "internal" plant events and "external" events. In this latter category are fires, earthquakes, floods, etc.

0.6.2 STRUCTURING THE LIST--THE PLANT EVENT TREE

Having identified the initiating events, we now direct our attention to the question, "what happens next?" That is, how does the initiating event propagate through the plant? The most useful tool for systematically thinking through these questions is the event tree diagram, introduced in WASH-1400, and used in the present study in Section 1.3.

Figure 0.6-3 is a symbolic representation of an event tree diagram. Arrayed across the top are various systems in the plant, e.g., reactor protection system, auxiliary cooling system, etc. At the left we enter the tree with the initiating event and then ask, "Does system A work or not?" Thus the tree branches at this point, the upper branch representing "system A works" and the lower "system A fails." At system B there is another branching, and so on.

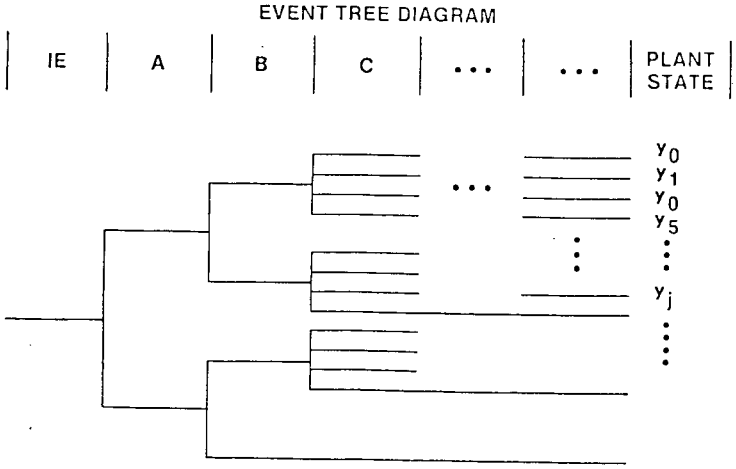


Figure 0.6-3. Structuring the Scenario List--The Plant Event Tree

In this way the tree diagram is developed. Each path through the tree thus represents a "scenario"--an envisioned sequence of events beginning with the specified initiating event and leading to a "plant damage state" represented by the symbol y_j.

These plant damage states are defined in terms of the conditions in the reactor vessel, the type and degree of coherence of core melt, and the status of the fan coolers and containment spray systems. These states are chosen and defined with sufficient specificity that once such a state has occurred the subsequent events in the containment are the same irrespective of the path by which that state was reached. As a result of this definition, a coalescence of scenarios occurs at this point which structures the scenarios list and greatly simplifies the computational labor involved in the analysis.

Also noted in Figure 0.6-3 is the fact that a given system need not be restricted to the two states "works" or "fails." In some cases it is appropriate to use a multistate model of the system, thus representing various states of partial failure. Electric power, for example, is a system which is treated this way as we shall see. How many states should be used for a given system is a question of modeling detail or "degree of aggregation" just as is the question of the number of systems identified in the event tree in the first place.

The situation here is identical to that present whenever a symbolic model is made of a real world entity, whether it be a mathematical model, an engineering model, a computer model, or indeed a verbal model. The point is, the model is not the entity. It is only a symbolic representation of the entity. Modelers sometimes tend to forget this. "The map is not the territory," and "the menu is not the meal."

The relevance of this point for our current work is that an event tree is a model of plant behavior. It can be made more accurate, more realistic by distinguishing more systems at the top and more states for each system. The price for this is greater complexity in the analytical work itself. The modelers' art and challenge is one of balancing the conflicting desires of realism and simplicity. The balance chosen in this study leans much farther in the direction of detailed event trees than does WASH-1400 or other studies of this type. The opinion of the present study team is that more detailed event trees make for greater scrutability and confidence in the model. They have the additional advantage that the fault tree modeling of the individual systems is made much simpler as we shall see later.

Another point needs to be made at this time; namely that, while we seek to make the model as realistic as we can, to the extent that some inaccuracy is necessarily inherent in every model we attempt to take this inaccuracy on the conservative side. Thus we design the model so that where in doubt it gives overstatements of the risk; that is

- Overstatements of the frequencies of scenarios.
- Overstatements of the damage.
- Overstatements of the uncertainties.

For example, if we model a system with a two state model, "succeeds" or "fails," even though the real system may have intermediate, partial success states, then we interpret the "success" state as being total

success. That is, any partial failure of the system is modeled as total failure. The frequency used for the failed state includes the frequency of any partial failures. Thus, when we simplify our model in this way, we do it so that the model gives results on the conservative side. This, of course, is just typical engineering practice everywhere.

0.6.3 THE CONTAINMENT EVENT TREE

The previous section described the development of event trees beginning with an initiating event, such as turbine trip or LOCA, for example, and continuing to the point where either the disturbance is controlled and the plant stabilized, or there is core damage or melt. If there is core damage or melt, we then enter a second tree called the containment tree, which represents the subsequent progress of the scenarios. This tree described fully in Section 2, is represented symbolically here in Figure 0.6-4. The entry states to this tree are the exit states from the plant tree, i.e., the plant damage states. In the containment tree such questions are asked as: Does the reactor vessel melt through? Is the core debris coolable? Does an overpressure develop? etc. The set of questions asked encompasses all the phenomena relevant to any of the scenarios stemming from the various plant damage states. It addresses the phenomenological questions in an approximate chronological order.

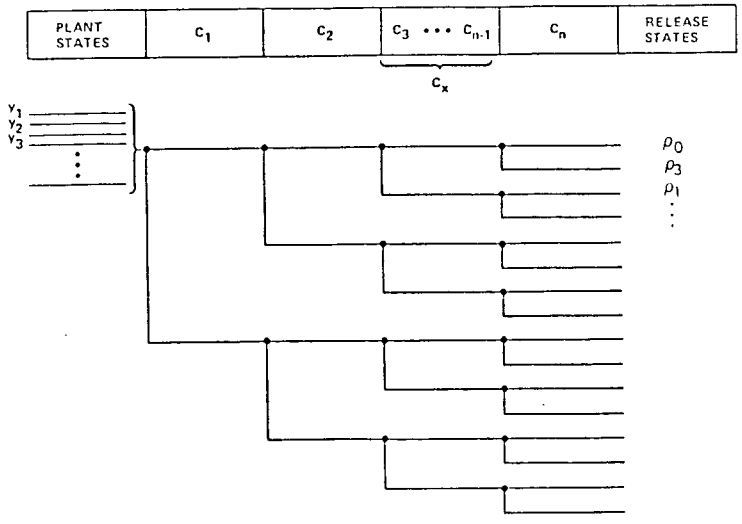


Figure 0.6-4. Containment Event Tree

The containment tree thus models the path of the scenarios from core damage to release or nonrelease of radioactivity to the environment outside the plant. Now once a given quantity and mix of radioisotopes has been released, what happens thereafter, the atmospheric dispersion is a matter of wind speed, direction, etc. It does not matter how the isotopes came to be released.

Therefore at this point we again impose a coalescence on the scenarios by defining a discrete set of "release categories" represented by the symbols P_k. Each P_k stands for a particular quantity and mix of isotopes, released with a particular time dependence, at a particular elevation or by means of a particular pathway through the containment structure.

The P_k therefore constitute the output states of the containment event tree.

A release category is assigned to each path of the tree, as shown in Figure 0.6-4, out of study of the physical process represented by that path. These release categories are essentially the WASH-1400 release categories. Modifications were made for some categories to account for new assessments of phenomenology.

0.6.4 THE SITE MODEL

Given that a particular release has occurred, the site model then determines the consequent damage to people. The site model, as described in Section 6, is embodied in an elaborate computer program called CRACIT and includes representation of the topography, meteorology, and demography specific to the site. For each particular release, this model traces the movement of the radioisotopes, their fallout to the ground, and their interaction with the population present. The resulting damage calculated depends not only on the specific release category but also on the weather conditions at the time (wind direction, speed, precipitation, etc.), upon the population pattern, upon the evacuation actions implemented, etc.

Conceptually these latter variables may be thought of as the top events of an event tree, so that an event tree diagram could actually be drawn for the site model. There is no particular value in doing this, however, for the site model, other than conceptually. It is important, though, to recognize that each combination of these variables defines a segment of a scenario just as each path through the plant and containment trees defines segments of scenarios. By linking these segments together we obtain total scenarios reaching from initiating events to final states of damage. The CRACIT code runs many thousands of different weather/evacuation combinations for each release category, including changes in weather and wind direction during the course of the incident. Thus, the number of different scenarios, already large up to the point of release, is multiplied by many thousands at this point, leading to a huge number of total scenarios.

The output of the site model is the damage that occurs in the neighborhood of the plant as a result of the release. Several different measures of damage are used: early deaths, early injuries, thyroid cancers, other cancers, and whole body dose. Let the variable x stand for a typical one of these measures, and let x_ℓ represent a discrete value of x. Then the site model can be thought of as a mapping, as shown in Figure 0.6-1, which takes us from release categories P_k to damage states x_ℓ.

0.6.5 RECAPITULATION OF THE IDENTIFYING AND STRUCTURING PROCESS

In Section 0.6.1 we began the identifying/structuring process by identifying a series of initiating events, which we shall label I_1, I_2... I_j,.... Each such event initiated a chain of events which branched into many, many paths each representing a different scenario. This myriad of paths was structured by a coalescense of scenarios or "pinching" at the point where core damage has occurred. We therefore defined a set of core damage states, y_1, y_2 ... y_j ..., and used a plant event tree diagram to display the set of paths between the IEs and the core damage states. We then used a containment tree diagram to show the branching that takes place between a core damage state and the occurrence of radiation release.

At this point we imposed another coalescense by defining a set of releases categories, ρ_1, ρ_2... ρ_k.... The progress of each release in the atmosphere was then simulated by a computer model leading to damage states x_ℓ.

If we now rank the release categories in order of increasing severity and think of the variable ρ as a continuous variable passing through the values ρ_k, then the quantity ρ itself can be used as an interim index of damage. Then we can express the results of the in-plant analysis, in the form of a risk diagram with ρ as the abscissa as shown in Figure 0.6-5. The results of the ex-plant part of the study may be expressed in a series of risk diagrams, one for each release category. We refer to these as "Conditional Risk Diagrams"; they show the risk of damage, x, to the population conditional on the occurrence of release category ρ_k.

When the in-plant and ex-plant parts of the study are done, the in-plant risk diagrams are combined with the ex-plant conditional risk diagrams to produce the final risk diagram as shown in Figure 0.6-5.

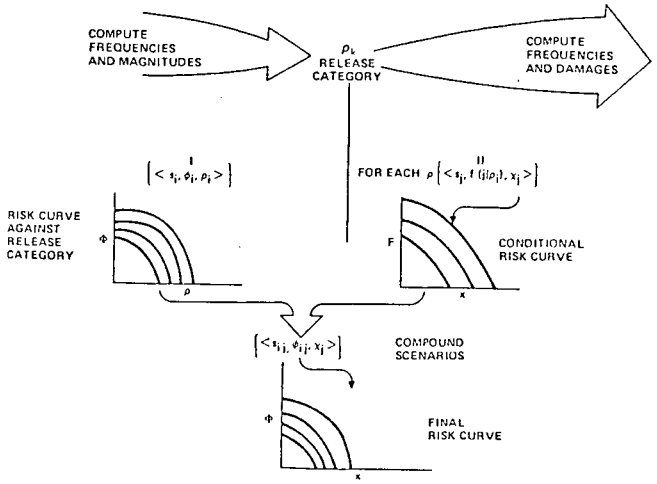


Figure 0.6-5. Combination of In-Plant and Ex-Plant Risk Diagrams

0.7 QUANTIFYING EVENT TREES

Each path through an event tree is characterized by the particular entry state and by the failed systems in the path. Thus, for example in the simplified plant event tree diagram of Figure 0.7-1 consider the scenario:

$$S = I\ A\ \bar{B}\ C\ \bar{D} \qquad (0.7\text{-}1)$$

meaning, the scenario consisting of initiating event or entry state I followed by success of systems A, C, and failure of B and D. This scenario is represented by the darkened line in the diagram (the lower branch at each node represents failure of the system).

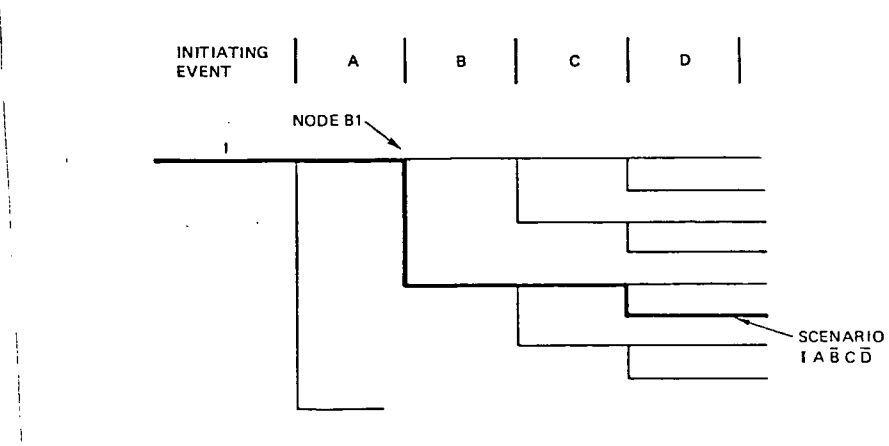


Figure 0.7-1. Simplified Plant Event Tree Diagram Showing Split Fraction at Nodes B1 and F9

The frequency of this scenario may be written:

$$\phi(S) = \phi(I)f(A|I)f(\bar{B}|IA)f(C|IA\bar{B})f(\bar{D}|IA\bar{B}C) \qquad (0.7\text{-}2)$$

where

$\phi(S)$ = the frequency of scenario S

$\phi(I)$ = the frequency of initiating event I

$f(A|I)$ = the fraction of times system A succeeds given that I has happened

$f(\bar{B}|IA)$ = the fraction of times system B fails given that I has happened and A has succeeded

$f(C|IA\bar{B})$ = the fraction of times C succeeds given that I has happened, that A has succeeded, and B has failed

$f(\bar{D}|IA\bar{B}C)$ = the fraction of times D fails given I, A, $\bar{B}$, and C.

The quantities f(A|I), etc., are called the "split fractions" at the nodes of the tree.

What this means exactly, for example, is that in our thought experiment out of all sequences which reach node B1, the fraction $f(\bar{B}|IA)$ takes the lower branch at this point.

The split fraction concept of course is not limited to nodes with two branches. Therefore, to generalize the notation, let $f_{n\ell}$ denote the split fraction for branch ℓ at node n.*

With the split fractions established at each branch point, we may then calculate the frequency of each scenario path as the frequency of the initiating event, times the appropriate split fraction at each branch on the path, i.e.,

$$\phi(S) = \phi(I)\, f_{1\ell_1}\, f_{2\ell_2} \cdots f_{n\ell_n} \cdots \qquad (0.7\text{-}3)$$

where ℓ_n is the branch chosen by the path at node n.

Now note in Equation (0.7-3) that if we divide by $\phi(I)$ we obtain:

$$\frac{\phi(S)}{\phi(I)} = f_{1\ell_1}\, f_{2\ell_2} \cdots f_{n\ell_n} \cdots = f(S). \qquad (0.7\text{-}4)$$

Here, the term on the right-hand side, the product of split fractions along a given path, thus has the interpretation of "conditional frequency" or the split fraction of that path. That is, out of all the times initiating event I occurs in our thought experiment, f(S) is the fraction of times in which scenario S results.

In this way we then may compute the split fraction for each path in the tree. These numbers thus characterize the tree itself, without reference to the frequency of the incoming entry state.

Now let us focus attention on a particular exit state, say y_j, and let S_{ih} denote a particular scenario going from entry state i to exit state y_j. By summing overall such scenarios we obtain

*Note here that we are adopting the Method Two mode of description of uncertainty. That is, we imagine a thought experiment in which we enter this node or branch point n many, many times. The $f_{n\ell}$ are the split fractions observed in this experiment. Since we have not, in fact, done this experiment, we shall have to express our uncertainty about its outcome by giving probability distributions against the $f_{n\ell}$.

If we had adopted a Method One point of view corresponding to a Level 1 definition of risk, we would replace the symbols $f_{n\ell}$ by $p_{n\ell}$ and would now call them "split probabilities" or "branch probabilities." $p_{n\ell}$ would now be the conditional probability, given that we enter branch point n of emerging in branch ℓ. There would now be, of course, in the Method One way of talking, no uncertainty in the $p_{n\ell}$ since there is no uncertainty about probabilities. The $p_{n\ell}$ already expresses our state of uncertainty.

$$m_{ij} = \sum_{h} f(S_{ih}) \qquad (0.7\text{-}5)$$

which has the interpretation

m_{ij} = split fraction, or conditional frequency, from entry state i to exit state y_j.

That is to say, out of all the times entry state i occurs, m_{ij} is the fraction of times that exit state j occurs.

To recapitulate then, what is meant by the phrase "quantifying an event tree" is the following process:

1. Specify the split fractions $f_{n\ell}$ at each branch point in the tree diagram.

2. Specify the exit state, y_j associated with each path.

3. Compute the split fractions for each path.

4. Sum the split fractions of all scenarios between each entry state and each exit state to obtain the numbers m_{ij}.

This process is carried out in detail for the plant event trees in Section 1 and for the containment tree in Section 2.

0.8 ASSEMBLING THE SCENARIO INFORMATION INTO FINAL RISK CURVES--THE MATRIX VERSION OF EVENT TREES

In this section we shall outline the process by which the final risk curves are assembled, in Section 8, from the following bodies of information:

- Plant event tree
- Containment event tree
- Site (ex-plant) consequence analysis
- Initiating event frequencies.

0.8.1 THE MATRIX FORMALISM

A key idea (Reference 0-8), which makes the assembly process easy to understand, is the recognition that an event tree may be regarded as equivalent to a transition matrix in the sense that the tree defines the likelihood of moving from various input conditions, or states, to various output states.

For example, in the plant event tree the entry states represent the various possible initiating events. The exit states represent various combinations of in-vessel conditions, mainly pressure at time of core melt, containment conditions (spray, fan coolers, leakage paths), and time of core melt (measured from initiation of the incident). We have already shown that from the tree we may calculate the numbers:

m_{ij} = the conditional frequency of leaving the plant event tree in exit state j given that initiating event i has occurred. (0.8-1)

The set of these m_{ij} may now be regarded as constituting a matrix:

$$M = \begin{bmatrix} m_{11} & m_{12} & m_{13} & \cdots \\ m_{21} & m_{22} & & \\ \vdots & & & \end{bmatrix} \qquad (0.8\text{-}2)$$

which we call, naturally, the "plant matrix".

In a similar way from the containment event tree we can obtain a "containment matrix" C.

$$C = \left[c_{jk}\right] \qquad (0.8\text{-}3)$$

where:

c_{jk} = the conditional frequency of emerging from the containment tree in exit state k given that we enter it in entry state j. (0.8-4)

Now the exit states from the containment tree are the various release categories, ρ_k. The entry states for the containment tree are by definition the exit states of the plant tree. Therefore, the matrix product MC gives the frequencies of transition from initiating event i to release category ρ_k. To see this, define the product matrix A as

$$A = MC \tag{0.8-5}$$

$$A = [a_{ik}]$$

where, by the definition of matrix multiplication,

$$a_{ik} = \sum_j m_{ij} c_{jk}. \tag{0.8-6}$$

Thus, in light of (0.8-1) and (0.8-4), a_{ik} has the meaning

a_{ik} = the conditional frequency of a release in category ρ_k occurring given that initiating event i has occurred. (0.8-7)

We see, therefore, that the matrix MC relates to the portion of the risk analysis lying to the left of the pinch point in Figure 0.6-5. In a similar way the right side of the pinch point, the site model, may be regarded as resulting in a "site matrix," S,

$$S = [s_{k\ell}] \tag{0.8-8}$$

where

$s_{k\ell}$ = conditional frequency of damage level x_ℓ, or greater, occurring given that a release in category ρ_k has occurred. (0.8-9)

The variable x here stands for any of the damage indices, e.g., early deaths, thyroid cancers, etc. We presume in (0.8-8) that this variable has been discretized and in (0.8-9) use x_ℓ to denote one of the discrete values.

Now multiplying MC by S, we see that the product matrix

$$B = MCS$$

has the meaning

$$B = [b_{i\ell}]$$

$b_{i\ell}$ = conditional frequency of damage x_ℓ, or greater, occurring given that initiating event i has happened.

0.8.2 THE INITIATING EVENT VECTOR AND THE MASTER ASSEMBLY EQUATION

We now finally introduce the row matrix

$$\phi^I = \left[\phi_1^I, \phi_2^I, \ldots \phi_i^I \ldots\right] \tag{0.8-10}$$

where

$$\phi_i^I = \text{the frequency of occurrence of initiating event } i \text{ (measured in occurrences per reactor year).} \tag{0.8-11}$$

ϕ^I is called the "initiating event vector." The product of this row matrix ϕ^I by the rectangular matrix M yields a new row matrix ϕ^y.

$$\phi^y = \left[\phi_1^y, \phi_2^y, \ldots \phi_j^y\right] \tag{0.8-12}$$

related to ϕ^I by

$$\phi^y = \phi^I M. \tag{0.8-13}$$

In expanded form, this relationship is

$$\phi_j^y = \sum_i \phi_i^I m_{ij} . \tag{0.8-14}$$

Thus, in light of (0.8-1) and (0.8-11),

$$\phi_j^y = \text{the frequency of plant damage state } y_j \text{ (in occurences per reactor year).} \tag{0.8-15}$$

ϕ^y is therefore called the "plant state vector."

If this vector, in turn, is now multiplied by the containment matrix, we obtain the row matrix ϕ^ρ

$$\phi^\rho = \phi^y C \tag{0.8-16}$$

$$\phi^\rho = \left[\phi_1^\rho, \phi_2^\rho, \ldots \phi_k^\rho \ldots\right] \tag{0.8-17}$$

where, by (0.8-15) and (0.8-4),

$$\phi_k^\rho = \text{the frequency of release category } \rho_k \text{ (in occurrences per reactor year).} \tag{0.8-18}$$

ϕ^ρ is therefore referred to as the "release vector."

Finally, multiplying ϕ^{ρ} by S we obtain

$$\Phi^{x} = \phi^{\rho} S \qquad (0.8\text{-}19)$$

where the row vector Φ^{x}

$$\Phi^{x} = \left[\Phi_1^{x},\ \Phi_2^{x},\ \ldots\ \Phi_{\ell}^{x}\ \ldots\right] \qquad (0.8\text{-}20)$$

has in its ℓth position the quantity

$$\Phi_{\ell}^{x} = \text{the frequency, occurrences per reactor year, of occurrence of damage level } x_{\ell}, \text{ or greater.} \qquad (0.8\text{-}21)$$

Φ^{x} is, in effect, a risk curve in frequency format, equivalent to Figure 0.5-1.

Substituting (0.8-13) and (0.8-16) into (0.8-19) we obtain the relationship between Φ^{x} and ϕ^{I} as follows:

$$\Phi^{x} = \phi^{I} MCS. \qquad (0.8\text{-}22)$$

This equation shows how the results of the plant, containment, and site models are assembled with the initiating event frequencies to yield the risk curve. It is therefore called the "Master Assembly Equation." It has various useful interpretations. Thus, the product MCS is the transition matrix from initiating events to damage levels.

Similarly, from (0.8-13) and (0.8-22),

$$\Phi^{x} = \phi^{y} CS. \qquad (0.8\text{-}23)$$

Thus, the product CS is the transition matrix from plant states to damage levels. The jth row of this matrix may thus be thought of as the risk associated with being in the jth plant state. In the same way, the kth row of S is, in effect, the risk associated with release category ρ_k, the ith row of MCS is the risk measure associated with the ith initiating event.

It is also worth noting that the row matrix

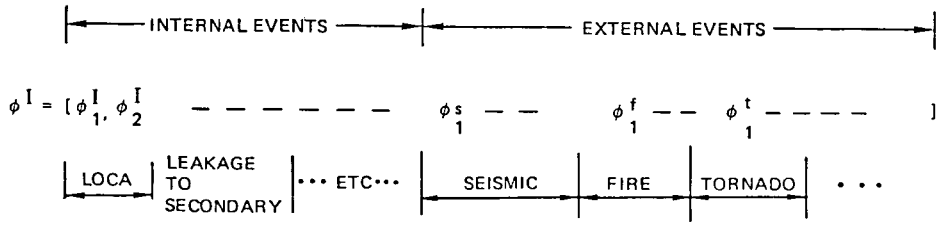
$$\phi^{\rho} = \phi^{I}\ MC, \qquad (0.8\text{-}24)$$

so that the ith row product MC may be thought of as the risk from initiating event i if the release category itself is taken as the measure of damage.

0.8.3 INCLUSION OF EXTERNAL EVENTS WITHIN THE MATRIX ASSEMBLY FORMALISM

Within the formalism of the master assembly equation, (0.8-22) we may readily include the treatment of external events. This is done by including in the meaning of the term "initiating events" both "internal" initiating events, such as turbine trip, etc., and "external" initiating events such as earthquakes of various size, fires at various locations, etc.

With this idea we can view the initiating event vector to be compartmented into subvectors as in the following sketch.



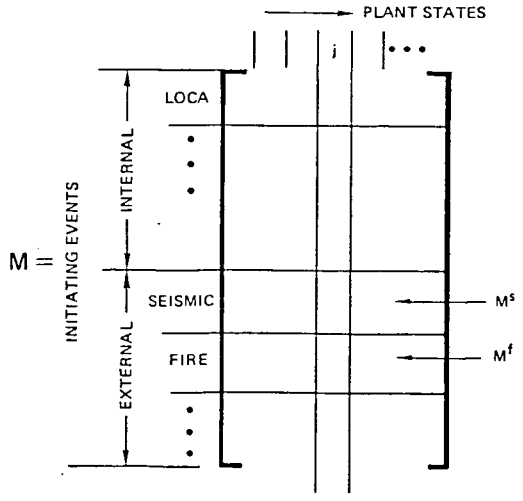
where

ϕ^I_1 = the frequency of internal initiating event 1.

ϕ^s_1 = the frequency of seismic initiating events (acceleration a_1) at the site.

ϕ^f_1 = the frequency of occurrence of fire type 1, and so on.

In a corresponding way the plant matrix M can be veiwed as partitioned



so that the submatrix M^s contains the transition fractions from initiating accelerations a_i, to plant states y_j. Similarly, M^f expresses the transition from various fires to the various plant states, and so on.

0.8.4 INCLUDING UNCERTAINTY WITHIN THE MATRIX FORMALISM (USING THE LEVEL TWO DEFINITION OF RISK)

0.8.4.1 Recapitulation

In the previous sections we operated within the context of our thought experiment in which all the interesting quantities had actually been measured. We recapitulate the definition of these quantities as follows:

ϕ_i^I = The frequency of initiating event i. That is the number of times event i occurred in the experiment, divided by the total number of reactor years logged.

m_{ij} = Out of all the times initiating event i occurred in this experiment, m_{ij} is the fraction of those times that the subsequent events in the plant resulted in plant exit state j.

c_{jk} = Out of all the times that the plant was in plant exit state j (equal to containment entry state j), c_{jk} is the fraction of those times that the subsequent events resulted in a release falling within release category ρ_k.

$s_{k\ell}$ = Out of all the times that a release in category ρ_k occurred, $s_{k\ell}$ is the fraction of those times in which damage level x_ℓ, or greater, resulted.

Φ_ℓ^x = The number of times damage level x_ℓ, or greater, occurred in the experiment, divided by the reactor years logged.

From these definitions, a relationship exists between these five quantities which was expressed in the matrix equation,

$$\Phi^x = \phi^I \text{ MCS.}$$

We now acknowledge the fact that we have not done the experiment and therefore do not know Φ^x. We wish, in fact, to calculate Φ^x using this relationship (0.8-22). In doing this, however, we must further acknowledge our uncertainties about the quantities ϕ^I, M, C, and S on the right-hand side. We must now, therefore, ask the question, "What uncertainty in Φ^x is implied by, consistent with, these uncertainties on the right side?" In other words, we need to "propagate" the uncertainties through the master assembly equation. The method for doing this is described in the next section.

0.8.4.2 Combining Uncertainties--The DPD Process

The propagation of uncertainties through the master assembly equation (0.8-22) is easily done using the concept of Discrete Probability Distribution arithmetic (See Section 0.12.3 and Reference 0-9). Applying these concepts we express the uncertainty in the matrix M by writing M as a discrete probability distribution (DPD) as follows,

$$M = \{<p_\beta, M_\beta>\} \qquad (0.8\text{-}25)$$

where the M_β are "versions" or "values" of the M matrix and p_β are the corresponding probabilities.

Similarly the matrices C and S are expressed as DPDs

$$C = \{<p_\gamma, C_\gamma>\} \qquad (0.8\text{-}26)$$

$$S = \{<p_\delta, S_\delta>\} \qquad (0.8\text{-}27)$$

and so also is the row ϕ^I

$$\phi^I = \{<p_\alpha, \phi^I_\alpha>\} \ . \qquad (0.8\text{-}28)$$

With this done, the right side of Equation (0.8-22) is viewed simply as a multiplication of DPDs, which is carried out by standard DPD procedure to yield a final DPD for Φ^X

$$\Phi^X_\epsilon = \{<p_\epsilon, \Phi^X_\epsilon>\} \ . \qquad (0.8\text{-}29)$$

This DPD, (0.8-29), is the final answer we were looking for. It is, in effect, a risk curve in probability of frequency format against damage variable x, and may be plotted graphically in that form.

0.8.4.3 Comments on Numerical Aspects

In Equation (0.8-29) ϵ denotes a combination of the indices α, β, γ, and δ

$$\epsilon = <\alpha, \beta, \gamma, \delta> \ . \qquad (0.8\text{-}30)$$

and

$$p_\epsilon = p_\alpha \, p_\beta \, p_\gamma \, p_\delta \qquad (0.8\text{-}31)$$

$$\Phi^X = \phi^I_\alpha \ M_\beta \, C_\gamma \, S_\delta \ . \qquad (0.8\text{-}32)$$

Equations (0.8-30), (0.8-31), and (0.8-32) are expressions of the standard DPD multiplication procedure applied at the level of matrix multiplication. Formally, the procedure is the same as for

multiplication of scalar quantities. As in the case of scalar multiplication, moreover, it is necessary to apply condensation procedures (Reference 0-9) in order to keep the quantity of computations within manageable limits. This necessity is even greater in the matrix case, of course, since the amount of computation in a matrix multiplication is much greater than in a scalar multiplication, according to the sizes of the matrices.

0.9 DETERMINING SPLIT FRACTIONS--FOR THE PLANT TREE

Let us turn now to the question of determining the split fractions for a typical branch point in the plant event tree. The basic process used here, described more fully in Section 1, is to perform a system analysis of the system to which that branch point relates. That is, we break the system down to its components and determine the relationship between the performance of the components and the performance of the system. From this relationship, from the likelihood of various component failures, and various combinations of component failures, the split fractions for the system are calculated.

A typical system analysis (see Section 1.5.2) contains the elements indicated in Figure 0.9-1.

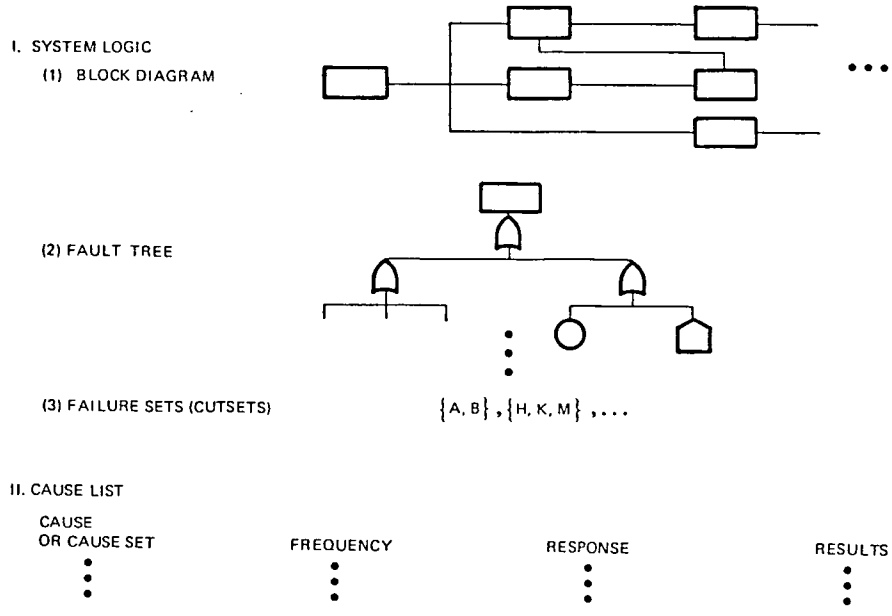


Figure 0.9-1. Form of a System Analysis

After a verbal description of the system and its purpose, there is a graphical portrayal of the structure of the system in terms of its components. This is typically done in the form of a block diagram as suggested in Figure 0.9-1. The next step, determining the system/component performance relationship, is typically done using a fault tree. Such a tree embodies the results of systematically pursuing the question: How can the system fail? If all its components succeed, then the system must succeed. The fault tree is a device, therefore, to help us identify those combinations of components which if they failed would cause system failure.

These combinations of components are called "failure sets" or, customarily, "cutsets." They are most concisely expressed by boiling them down to what are called "minimal" failure sets or "minimal" cutsets. A failure set is minimal if and only if it is true that, were we to restore to success any one of the components in the set, then the system would no longer fail.

The set of minimal failure sets for a system therefore embodies and expresses the logical relationship between the system and its components. Anything that could cause failure of the system must do so by acting through, that is to say "causing," failure of one or more failure sets.

So now the issue devolves upon what could possibly cause failure of all the components in a failure set. To facilitate and record the results of this inquiry, a device called a "cause table" is used (see for example, Tables 1.6.2.3.1-6 to 1.6.2.3.1-9). The precise form of these tables varies from case to case. Conceptually, in its maximum development, the cause table has the form sketched in Figure 0.9-2.

CAUSE TABLE FOR SYSTEM _____

ANALYST _____ EVENT TREE ___ BRANCH ___ POINT

DATE _____

(1) CANDIDATE CAUSE	(2) OCCURRENCE FRACTION ψ	(3) OPERATOR RESPONSE	(4) RESPONSE OCCURRENCE θ	(5) COMBINED OCCURRENCE $t = \psi\theta$	(6) COMPONENT FAILED	(7) SYSTEM STATE	(8) OTHER SYSTEMS	(9) INITIATING EVENTS
CRFS								
T&M + CRFS								
HUMAN ERRORS								
DESIGN ERRORS								
ENVIRONMENTAL FACTORS								
HE + CRFS								
HE + T&M								
.								
.								
.								
OTHER								

Figure 0.9-2. Cause Table

0.9.1 THE CAUSE TABLE

Down the left side of this table are listed various possible candidate "causes" for system failure. For example, there is the possibility of coincident random failure (CRF) of all the components in a cutset. There is the possibility that certain components may be out of action for purposes of test and maintenance (T&M) and, just at this time, certain other components fail coincidentally so as to result in failure

of a cutset. Likewise there are various possible human errors (HE) that could be made. These could occur coincidentally with test and maintenance outage or with random failure. There could be system failure as a result of design errors, unexpected environmental factors, and so on, all of this of course occurring just at the time the system is called upon to function because of the occurrence of an initiating event.

In the second column of the table we list the frequency or fraction of occurrence of each candidate cause. That is to say, with respect to our thought experiment, consider all the scenarios that arrived at a particular node in the event tree and therefore asked for the system in question to work. Out of all these instances, what fraction of times did the candidate cause occur? That fraction is the number that is entered in the second column. (Of course, if there is uncertainty about that number, that uncertainty must be quantified and entered also.)

Now if the candidate cause occurs, we need always ask how the human beings in the plant will react. It is possible almost always by correct action to recover system performance in spite of the occurrence of the candidate cause. On the other hand, there may be no action or incorrect action taken. In the third and fourth columns, then, are listed the various possible operator reactions along with the occurrence fraction of each.

Column five of the table lists the occurrence fractions for the combination of candidate cause and operator response. Columns six to nine record the consequences of each such combination. Column six shows which components fail as a result, and column seven shows whether or not the system fails.

Columns eight and nine ask whether this particular candidate cause/reaction combination could cause any other system to fail or whether it could cause another initiating event. These two columns thus are used as signals to highlight, call attention to, possible "common causes"--causes which could fail more than one system or fail a system plus produce an initiating event. Once identified, such potential common causes receive special attention in the analysis.

For example, the cause "earthquake" could fail several systems as well as generate an initiating event. The whole class of earthquake induced scenarios is therefore given special treatment. Similarly, a fire in a particular location could be the common cause of failure of several systems. These too are thus treated specially.

Leaving aside then the common causes failing more than one system, which are separately treated, we obtain the split fractions for each system state by summing appropriately in column five of the cause table.

0.9.2 "CAUSES" RELATED TO SCENARIOS

Figure 0.9-3 shows a representative scenario or path through the plant event tree. In this path two systems B and F have failed (indicated by the bars over B and F). All other systems work. Therefore denote this

scenario I $\bar{B}\,\bar{F}$. Under system B imagine that we have listed all the possible causes of system B failure. Let b_j denote one such cause. Similarly, list the causes of failure of F and let f_k denote a typical such cause. Then the combination I, b_j, f_k constitutes a particular way in which the path I $\bar{B}\,\bar{F}$ can be realized. In fact, varying j and k, we see that each such combination I realizes I $\bar{B}\,\bar{F}$. We can say in fact that each combination I b_j f_k is itself a scenario--a subscenario of I $\bar{B}\,\bar{F}$.

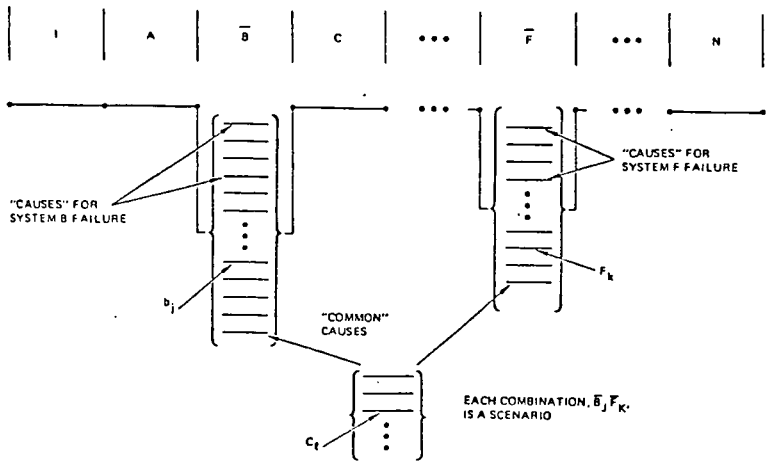


Figure 0.9-3. Structuring the Scenario List Scenarios Including System Failure Modes

Thus every path through the event tree really represents a whole multitude of scenarios comprised of the various combinations of causes that result in the system failures of that path. Thus we see how the event tree/fault tree/ cause table methodology is in effect a structured listing of a huge number of individual scenarios.

We also indicate in Figure 0.9-3 those causes which could fail both B and F. These causes are "common" to B and F. Each such common cause together with the initiating event constitutes a subscenario of I $\bar{B}\,\bar{F}$.

0.9.3 RELATION OF SYSTEM ANALYSES AND EVENT TREES

Figure 0.9-4 shows the relationship of the structuring ideas we have been discussing. It shows how they fit together. Thus at the top (the plant level) is indicated the event tree diagram. This level shows what combinations of system failures, together with what initiating events, result in any given plant state.

At the next level down, the system level, there are fault trees expressing the relationship of the system to its components. This level shows what combinations of component failures result in failure of the systems. Below that is the cause level, showing what causes could result in component failure and what combinations of causes could result in those combinations of component failure which cause system failures. Particular interest centers, at this level, on single causes which by themselves could fail more than one component or more than one system.

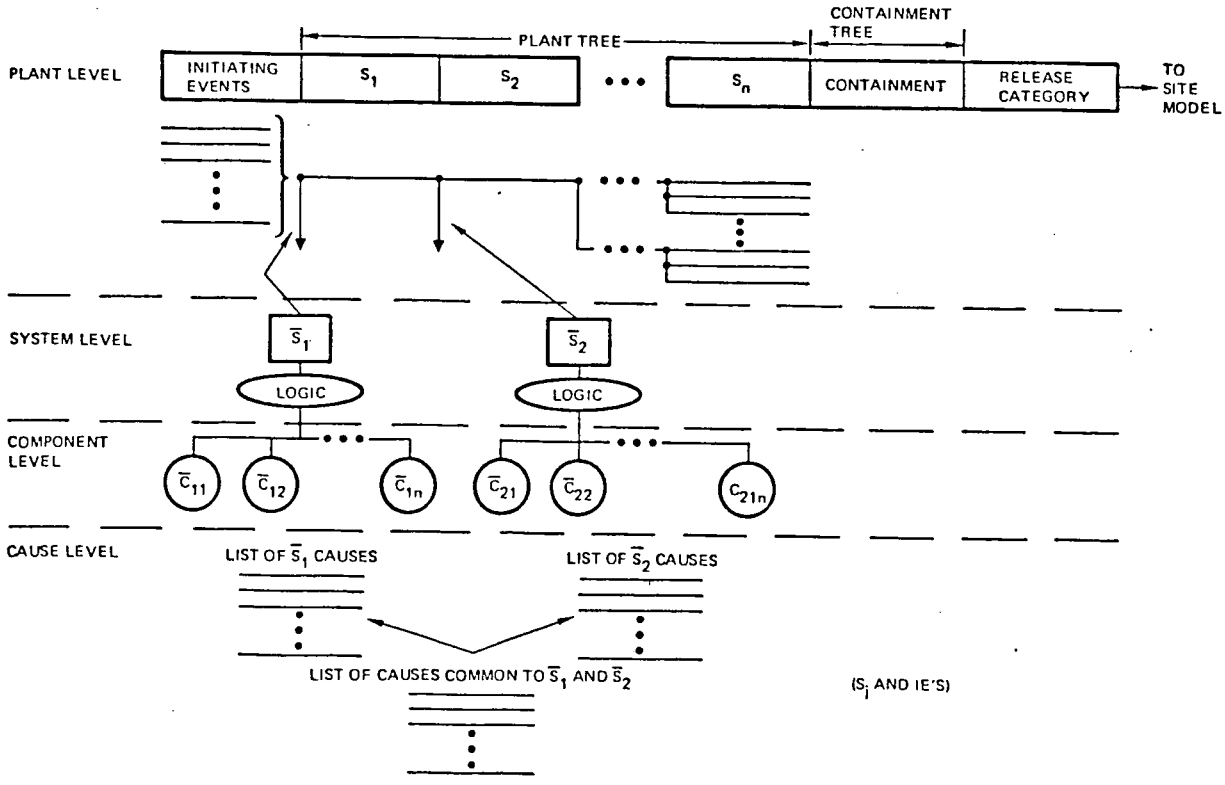


Figure 0.9-4. Structuring Scenarios

0.10 DETERMINING SPLIT FRACTIONS FOR THE CONTAINMENT EVENT TREE

On a plant event tree, each branch deals with the success or failure of a certain system or operator action to accomplish its mission. The split fraction expresses the frequency of successes and failures per demand which is determined by a detailed analysis of the system or operator action in terms of its components and identifying the possible causes of failure to accomplish the mission.

A containment event tree, while pictorially similar to a plant event tree, is philosophically very different. Its branch points deal with physical processes which are governed by the law of physics. For example:

- Does the hydrogen-oxygen mixture in the containment ignite?
- Does the core melt coherently?
- Is the debris coolable within the pressure vessel?
- Is the containment pressure higher than the failure pressure?

In the containment event tree, the split fractions are assigned by judgment which is supported by analyses and experimental data relevant to the physical process under consideration. Three distinctive, different types of analyses are performed.

First, the individual physical phenomena are investigated to: a) determine the conditions necessary for a particular process such as a steam explosion to occur--this investigation utilizes first principle analyses, a review of experiments, and a review of prior related analyses; and b) determine whether, for a given plant state and containment sequence, the conditions required for the process to occur are met. This first kind of analysis forms the basis for stating a judgment on the likelihood for a given phenomena or process on the containment event tree to occur.

Secondly, on the basis of the first type of analysis, a set of key integral analyses are defined and a transient analysis of the entire containment sequence is performed to determine whether the time-phased conditions in the containment are likely to cause conditions for containment failure.

Given these analyses, the basis is now established for judging the containment event tree split fractions.

0.11 ANALYSIS OF RELEASE CONSEQUENCES--THE SITE MODEL

Each of the major accident scenarios identified in this study has a release of radioactive material associated with its occurrence. Each scenario was assigned to one of several "release categories" based on the core and containment responses to the accident conditions. The objective of the consequence portion of this analysis is to estimate the potential for health effect impact on the surrounding population due to exposure to the releases assigned to each release category.

Assuming that a release occurs at any random time, there are a number of factors that would cause the consequences to vary. The most important factors affecting this variation are related to the weather conditions that exist during the release. Wind speed, wind direction, atmospheric turbulence, and rainfall dictate the area contaminated and the amount of radioactive material in the plume. Once the distribution of airborne and deposited radioactive material is defined in the environment as a function of time, an estimate of health effects can proceed. A large number of individual weather conditions scenarios are studied to determine the probability that selected health effects would occur. The calculation procedure shown in the block diagram of Figure 0.11-1 is repeated for each meteorological sequence.

In order to compute the health impact to individuals or population groups, information related to their locations and travel paths (during evacuation) with respect to the radioactive plume location must be characterized. The degree of radiation exposure is not only dependent on time of exposure to the plume or deposited material on the ground, but also on the shielding provided by dwellings and/or other protective measures taken.

Once the frequency of occurrence of selected health effects has been established for each release category, these results are combined with the frequency of occurrence of the release category itself to determine the overall potential impact on health (expressed as risk curves) due to operation of the plant.

0.11.1 CONSEQUENCE ANALYSIS METHODOLOGY

The principal model used to assess health impact is an extensively modified version of the CRAC (Calculation of Reactor Accident Consequences) computer program develped by the U.S. Nuclear Regulatory Commission for the Reactor Safety Study (RSS).

The CRAC program, as it existed for the RSS, was intended for use in computing the overall risk from a number of nuclear plants in different regions throughout the United States. However, for this study consequences specific to each site are estimated; therefore, modifications to the CRAC program were necessary to treat specific site conditions adequately and realistically in the vicinity of the site. The two fundamental changes are the use of variable direction plume trajectories and the incorporation of a variable direction evacuation scheme. By

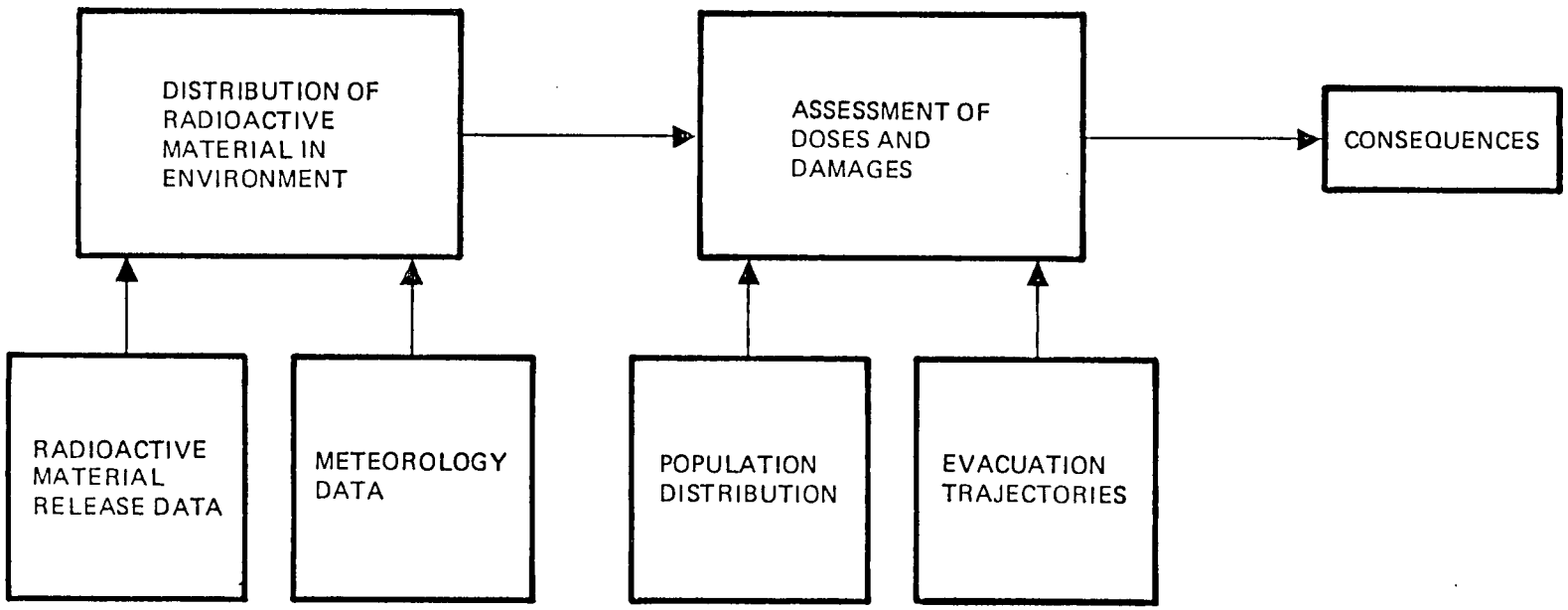


Figure 0.11-1. CRAC and CRACIT Consequence Assessment

comparison, CRAC models both plume travel and evacuation path in straight lines extending radially outward from the plant. The modified version of CRAC is referred to as CRACIT (Calculation of Reactor Accident Consequences Including Trajectories).

A series of mathematical and statistical models is coupled in CRACIT to assess risk. CRACIT requires as input an inventory of radioisotopes assumed to be released from the plant to the environment for each type of accident release category. Basic functions of the program include the calculation of meteorological dispersion of the released material as it travels down and the estimation of the health effects this material will have on the surrounding population. Each CRACIT run uses a continuous series of measured hourly meteorological conditions from a number of towers in the plant region.

To develop the distribution of frequency of occurrence versus health effects, the CRACIT model is run repeatedly with different start hours selected at random. Calculations follow the start hours in a sequential hour-by-hour manner until the plume has traveled far away from the site region. Assumed accident start dates are selected randomly in sets stratified so that start times are uniformly distributed over all months and each month has the same number of day and night samples.

The radiation dose calculations portion of the CRACIT consequence model was essentially unchanged from that used in the RSS. Estimated health effects, computed from the simulation of radiation exposure, are combined to form frequency distributions of consequence versus probability for a given hypothetical accident. Program libraries include detailed site information on meteorology and seasonal population specific to the site.

The following subsections provide further discussion of the important elements of a consequence calculation. Referring again to Figure 0.11-1, the distribution of radioactive material is the first calculation made, followed by the health effects calculation. Both are treated in more detail below.

0.11.1.1 Plume Trajectory and Distribution Modeling

The plume trajectory and distribution model can best be described as a time-dependent plume segment model. Picture the plume track as the projection of a round "puff" or cloud that moves with the wind (not necessarily in a straight line) and grows (and thus becomes diluted) as it moves downwind. As the cloud travels, particulate matter is deposited on the ground along its path leaving a swath of contaminated land. If it is raining, the deposition of particulates is accelerated.

At preselected distance intervals (referred to as spatial intervals), calculations of the plume concentration and amount of material deposited on the ground are made. The time of arrival and departure in each spatial interval is saved along with plume intensity information which is later used for dose calculations.

As the cloud moves downwind, its track is free to change due to several influences. If an hour has elapsed, a new set of meteorological data is available for the next sequential hour which may dictate a change in direction, speed, or rate of dispersal (or rain). Similarly, if the hour has not changed but the cloud has moved into a different meteorological region (see Section 6.2) which may have different wind patterns for the same hour, the cloud will respond accordingly. Finally, the cloud track may be affected by the wind flow field as it travels around or over terrain obstacles.

Now consider the persons in the site vicinity for which dose calculations are to be made. If they were to remain stationary, dose calculations could be made based on the plume distribution information calculated above for each pertinent location around the plant. However, in all likelihood these people will have evacuated during the release. These evacuation paths will likely not be along the plume path in a straight line. Therefore, to attempt to model the evacuees' dose, the model establishes evacuation trajectories and timing along the path for each population group. The dose calculation and the cloud (or its deposited material) coincide in space and time. Figure 0.11-2 illustrates the overlaying of the plume track and evacuee path on the "fine" calculation grid. This technique represents a significant departure from the straight-line calculations used in the RSS.

0.11.1.2 Atmospheric Dispersion Models

In an attempt to reduce uncertainties in the atmospheric dispersion part of the consequence analysis, many significant improvements were made to the original RSS dispersion model.

These modifications included development of the following:

- Use of wind speeds at/or near plume dispersion height to more accurately model time of arrival and greater dilution.
- A model using four wet deposition (washout) rates based on measured rain rate.
- Use of meteorological data from 14 locations in the site region including hourly estimates of lid height.
- Use of updated plume rise and lid penetration models.
- Use of terrain height correction model.
- Use of trapping and fumigation models under inversion lids.
- Use of a plume "lift-off" model.
- A turbulent internal boundary layer (TIBL) model to account for lake effects.

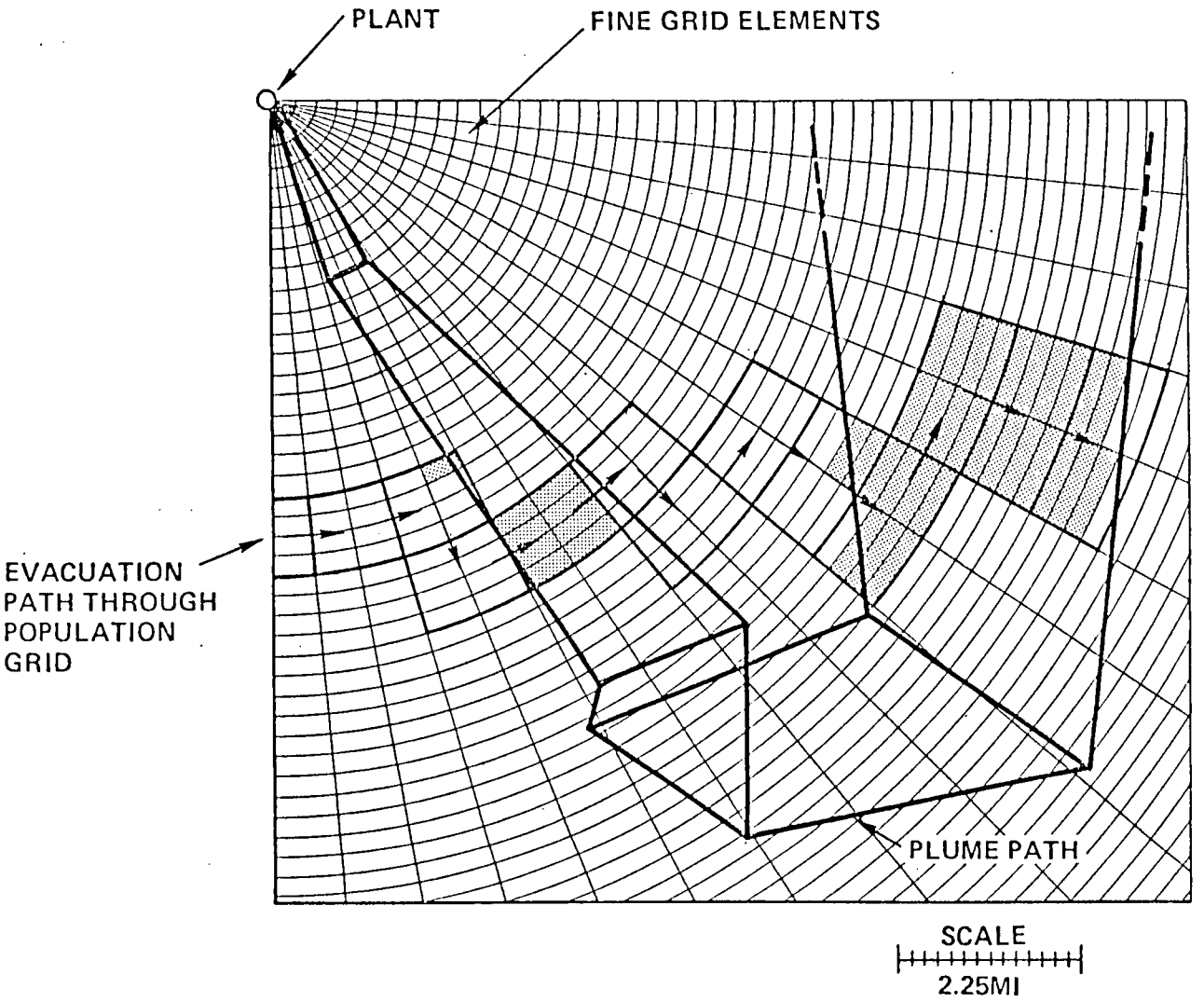


Figure 0.11-2. Illustration of Plume and Evacuation Paths on Fine Grid
(Dose Calculations Made in Shaded Fine Grid Areas)

- A three-dimensional model for determining the effects of terrain on wind flow and plume trajectory.

- The ability to modify plume characteristics based on expert judgment.

No matter how good the quality or quantity of site data, there are cases when the models described above for defining plume dispersion and trajectory are obviously not accurate. Examples of such cases are conditions over large bodies of water (where no measurements exist) and the influence of complex terrain on plume trajectory. To reduce the uncertainties, a method was devised for enabling a pre-edit of the parameters which define plume behavior prior to performing the dose calculations.

To accomplish this feature, a minicomputer was used with the aid of a graphics CRT. Each plume scenario was displayed on a site map along with printed values for pertinent parameters as illustrated in Section 6.2 for "before" and "after" edit cases. A team of diffusion meteorologists then inspected the plume characteristics for reasonableness and decided what, if any, changes were necessary. Edits were possible to each of six parameters: wind speed, wind direction, stability, lid height, rainfall or plume front width. After the editing process, the plume is redrawn for inspection and can be changed again if necessary. The editing process was repeated until the appropriate plume scenario was achieved. This procedure was done for each of two groups of release categories, one for elevated releases and one for ground level releases.

For the Zion site, the presence of the lake has significant influences on local dispersion. A study was conducted to classify each of 288 randomly selected meteorological sequences according to several categories of meteorological scenarios to determine how frequently lake effects conditions which could not be characterized by the plume dispersion modeling in the program would be expected. Of these, about 25 scenarios were selected for editing.

0.11.1.3 Computation of Doses and Health Effects

After plume trajectory segments have been laid out in space on a fine grid and data for each segment and concentrations and ground concentrations by isotope and arrival times have been stored, the remaining task is to quantify the impacts of the release in terms of public health. Public health impacts result from two exposure phases, early exposure and chronic exposure. Early exposure (exposure that occurs within a short time following release) may be distinctly different for two population groups--people who evacuate immediately and people who do not. Within both groups there are subgroups whose exposures are distinctly different because of differences in radiation conditions encountered at the different residence locations, and for evacuees at points along their different evacuation paths. A minor additional complication lies in the fact that the numbers of people in each group vary with start time (season or time of day).

Chronic exposure results from the return to normal use of the land contaminated during the passage of the radioactive plume (cloud) and accumulates over a period of years following the release. The program has been made flexible to accommodate studies of the effects of different mitigation strategies.

So, for each occupied fine grid element in the cloud path, exposures are initially set at full cloud passage for exposure to airborne radioactivity. It is assumed that the resident beyond the 10-mile evacuation zone would be relocated within 24 hours, and the ground dose is calculated to be that equivalent to 24 hours of exposure.

Once the radiation doses have been computed as described above, the effect on the health of the exposed individual must be estimated. The health effects model in CRAC was used unchanged in the CRACIT calculations. Detailed information concerning the clinical and experimental data on which the calculations are based is included in Appendixes F, G, H, and I of Appendix VI of WASH-1400 on the early and continuing somatic effects, late somatic effects, thyroid effects, and genetic effects, respectively.

The health effects that could be associated with a reactor accident are divided into categories as follows. Early and continuing somatic effects include the early mortalities and morbidities that are usually observed after large, acute doses of radiation and can occur within days to weeks after exposure; they also include illnesses and deaths that can become manifest within a year or so.

The late somatic effects include latent cancer fatalities and morbidities as well as benign thyroid nodules. In radiation effects experience, such effects are typically observed 2 to 30 years after irradiation. Finally, there are genetic effects which do not manifest themselves in the irradiated individuals, but rather in their descendants. In contrast to the early somatic effects, both latent cancer and genetic diseases are assumed to be random phenomena whose probability of occurrence is some function of the dose magnitude. For this reason, both late somatic and genetic effects are calculated on the basis of population dose (cases per million man-rem) rather than individual doses. Late somatic and genetic effects may result from very low doses but with a very low incidence. Consequently, these effects may occur at long distances (e.g., 200 miles) from the reactor; but because of the decrease in doses at great distances, the risk of late somatic and genetic effects for a given individual would be low.

As stated in the preceding sections, the underlying objective of this study is to make as realistic an assessment of risk as is possible and to indicate the uncertainties in the estimate. Until recently, the 1972 BEIR report (Reference 6-9, Section 6) was considered to represent the best authoritative data source for quantifying latent cancer and genetic effects and was used as input data to CRAC. There was experimental evidence which suggested that the 1972 BEIR report overestimated

cancer risks. The RSS used these experimental data to justify reductions to the BEIR linear risk coefficients for low doses and dose rates. The latest report by the BEIR committee (Reference 6-10, Section 6) supports the use of reduction factors used in the RSS for certain cancers. Accordingly, the RSS reduction factors have been used in this study, except for thyroid and breast cancers, in keeping with BEIR recommendations.

The health effects estimated in this analysis are similar to those used in the WASH-1400 analysis. Those included in this analysis are:

1. Acute fatalities
2. Injuries (excluding cancer)
3. Thyroid cancer cases
4. Cancer fatalities (other than those from thyroid cancer)
5. Whole body man-rem.

Other effects included in WASH-1400 can be estimated from these five. For example, genetic effects may be derived from whole body man-rem, using a conversion of 0.0002 genetic effects per man-rem to obtain values on a basis consistent with WASH-1400. Similarly, benign thyroid nodule incidence may be considered to be about 50% higher than thyroid cancer incidence. Thyroid cancers were estimated separately because of the low fatality rate for thyroid cancer--less than 10%.

0.11.2 SITE INFORMATION

To conduct a site-specific consequence analysis, a considerable amount of site information is needed, including:

- Population distribution data
- Evacuation path and timing data
- Meteorological data from the site and adjacent regions
- Content and magnitude of radioactive material released
- Land use information.

Following is a discussion concerning the data used for the Zion site in this evaluation.

0.11.2.1 Population Data

The town of Zion immediately west of the plant represents the highest population density near the site. There is a high summertime population along the beaches located south of the site. Population growth and distribution predictions were based on the population distribution within a 50-mile radius of the site region from the FSAR for the year 1985. The population was subdivided on a more detailed 32 sector grid than that used in the FSAR. Seasonally higher populations were assumed to apply from June through August.

Population figures for large distances were derived by estimating relative densities from a population density map of the United States compiled from the 1970 census. Seasonal changes are not considered at these large distances.

0.11.2.2 Evacuation Data

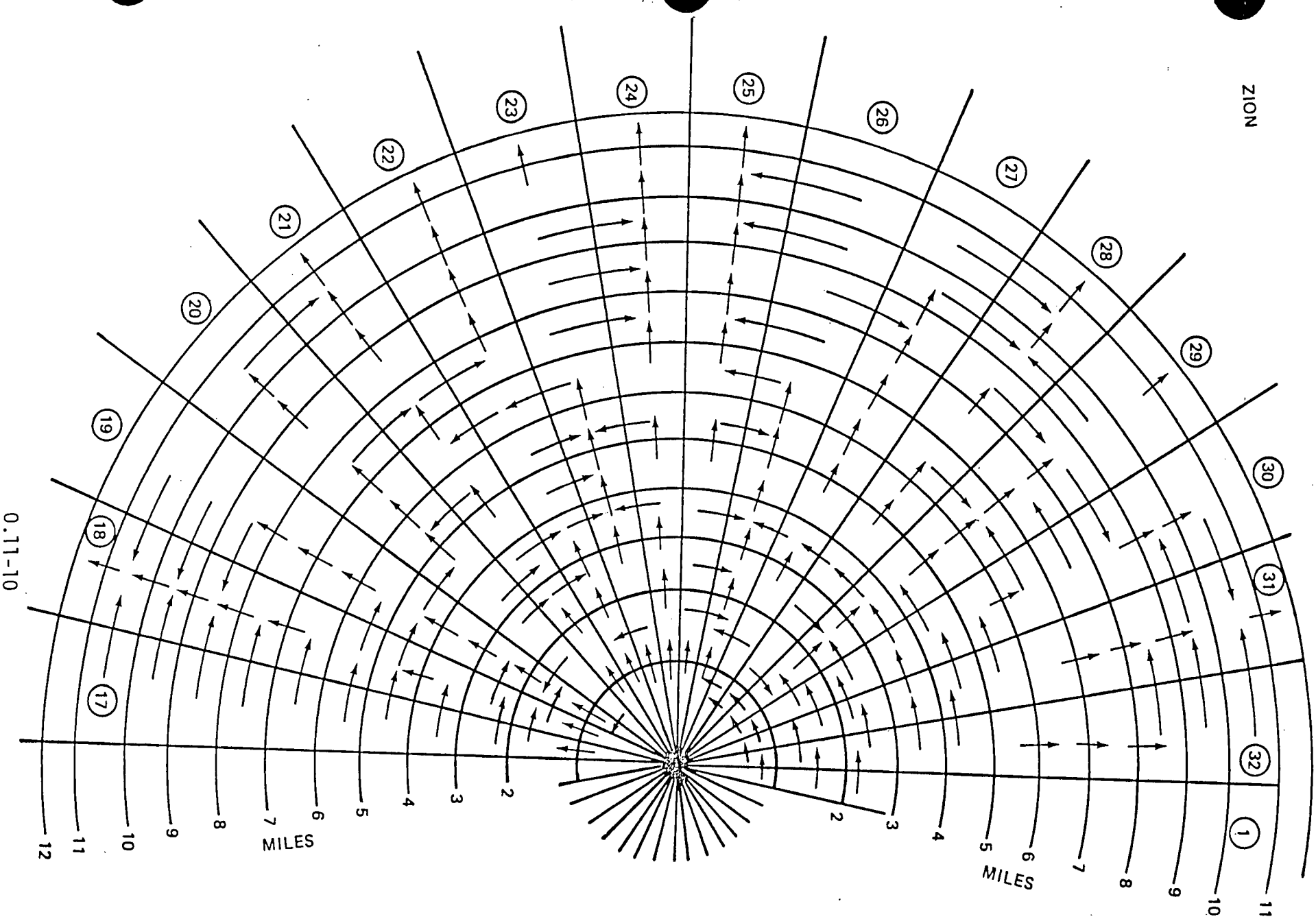
The evacuation estimates were made using the evacuation plan along with the detailed road and population maps of the site area out to 11 miles. The first task involved locating major evacuation routes and determining which of the 32 direction sectors was traversed along each route. Then the direction of evacuation in each radial segment was estimated based on the preferred evacuation route. A series of evacuation direction vectors was established as shown in Figure 0.11-3. After determining the direction of each vector, the distance from the center of each segment to the center of the next along the vector was estimated. Knowing the distance and speed, the time in the segment is easily determined and used in the dose calculations.

After determining the paths, the speeds along the vector in each segment were estimated based on a Preliminary Evacuation Time Study prepared by Stone & Webster for Commonwealth Edison Company dated January 1980. For initial timing runs, a uniform evacuation speed along each route was assumed. Then using a program called TRACK to simulate evacuation times, the uniform speed was adjusted to achieve an average evacuation time comparable to those in the Evacuation Time Study to account for the evacuation delays such as for the beach and amusement park populations. Speeds along the vectors in the higher population areas were adjusted downward. For Zion an average speed of 3 mph resulted in evacuation times averaging about 5 hours.

0.11.2.3 Meteorological Data

The primary source of weather data is the 250-foot tower instrumented at three levels in conformance with NRC Regulatory Guide 1.23. A 1-year period of these data, starting in January 1976 and ending in December 1976, was used. Meteorological parameters were stored as sequential hourly values on magnetic tape for use in the CRACIT program runs.

Due to local meteorological effects at the site, any plume that is projected to leave the site area is assumed to be influenced according to weather data measured in an adjacent region. Therefore, separate tapes were prepared in the I9 format for data from 14 sites in the region such that data were available concurrently for each hour from all sites. A map of the Zion site region appears on Figure 0.11-4 showing the location where meteorological data were measured. Regional boundaries defining the area of assumed influence of each data set are also marked on the figure.



0.11-10

Figure 0.11-3. Evacuation Vectors - Zion Site

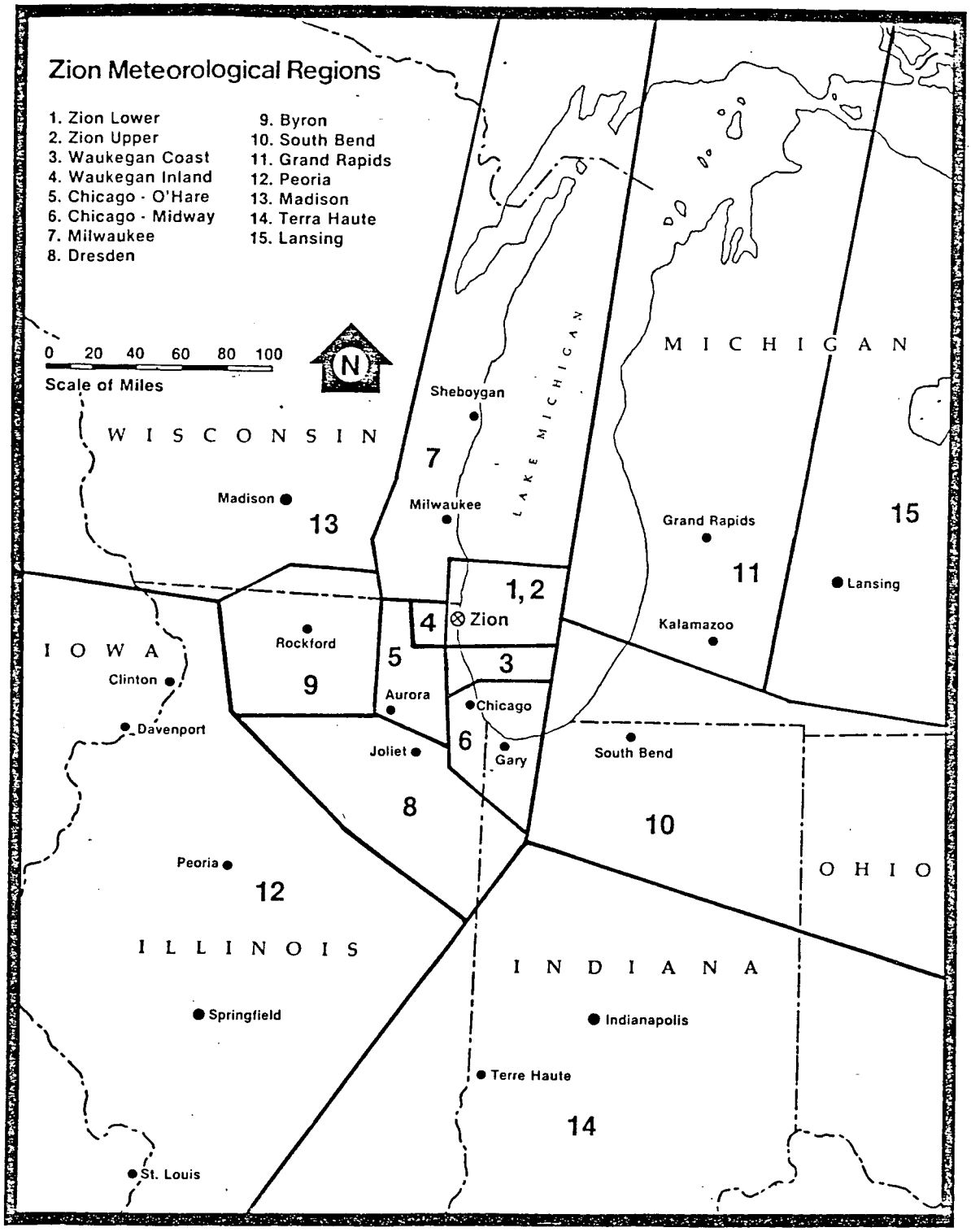


Figure 0.11-4. Zion Meteorological Regions

0.11.2.4 Release Categories

Separate CRACIT runs are made for each release category, with the probability of each release category set to 1.0. The magnitude of radioactive releases are input as fractions of the initial core radioactivity that might leak from the containment structure. In addition to release magnitude, the parameters that characterize the various hypothetical accident sequences are time of release, duration of release, warning time for evacuation, height of release and energy content of the released plume.

The time of release is the time interval between the start (core shutdown) of the hypothetical accident and the release of radioactive material to the atmosphere. It is used to calculate the initial decay of radioactivity. The release duration is the total time during which radioactive material is released and is used to adjust the horizontal dispersion due to wind meander during releases longer than one-half hour. The warning time for evacuation is defined as the interval between awareness of impending core melt and the release of radioactive material to the atmosphere. The height of release and the energy content of the released plume (which causes buoyancy) are used to compute plume height. Release categories are established which are composites of numerous accident sequences with similar characteristics. The assumed release quantities of each isotope are very similar to those in the RSS, even though recent information may support use of reduced release fractions for iodines and particulates.

0.11.2.5 Land Use Information

During the RSS, land use information was collected for each state. This same information was used in this study and was omitted for areas over water.

0.11.3 UNCERTAINTIES

The consequence analysis models used in this study contain a great number of variables. Examples include meteorological conditions, factors affecting plume dispersion, evacuation assumptions, shielding assumptions, dose calculation factors and health effect factors. For most of these parameters, a single value was assigned to each. The intent was to assign a "best estimate" value; however, there is a tendency to assign conservative values as the state of one's knowledge is less certain. Thus, consequences are based on single value or "point estimate" for each parameter with knowledge that some amount of uncertainty exists. A rigorous analysis technique would dictate the propagation of assigned uncertainties surrounding each individual variable through the CRACIT calculation; however, this would be an extremely difficult, if not impossible task. Therefore, uncertainties are expressed in a more aggregate way as described below.

First, let us mention that, in the current study, the considerable effort made to reduce uncertainties by improving the physical models which describe the dispersion (or transport) of released material in the

atmosphere is an important consideration. Secondly, the extensive amount of meteorological data from 14 sites improved the ability to predict plume location. Associated with the improved modeling and data activities was the injection of expert judgment concerning transport of the released material based on the expert's familiarity with local conditions. Another example of model improvements lies in the evacuation model which utilized a realistic evacuation time and path simulation.

Thus, the release consequence results in this study have incorporated state-of-the-art improvements in modeling; however, still no quantifiable value has been placed on the uncertainties in the results. One way to obtain quantifiable results is to run sensitivity studies. Several were run, including sensitivity to the number of meteorological sequences, evacuation speeds and finally, source term. Examination of sensitivities to evacuation scenarios led to the conclusion that, given a realistic warning time, the number of early fatalities was not particularly sensitive to evacuation speeds. The study of sample size showed that variations could be stablilized if at least 288 samples were taken; therefore, uncertainties were reduced by taking this number of samples for the release categories that dominate the risk curves. Finally, it was concluded that variation of source term quantities could be used in an aggregate way to simulate uncertainties in the data, in the dispersion model, in the dose model, and to some extent in the evacuation model. There are nonlinear sensitivities in the health effects model to such variations.

Calculations were therefore made for three assumed source term magnitudes in addition to the original "point estimate." The point estimate assumptions in the model tend to overestimate rather than underestimate doses, i.e., they are conservatively high assumptions. Results of this series of runs showed dramatic changes in health effects over the spectrum of source term changes as illustrated for a typical case in Figure 6.3-6 of Section 6.

0.12 PROBABILITY DISTRIBUTIONS--BASIC CONCEPTS

Probability distributions are, of course, fundamental to any discussion of risk and are used extensively throughout this study. For convenience, therefore, this section collects and reviews some of the basic ideas and standard language relating to such distributions.

0.12.1 DISTRIBUTION FUNCTIONS

Given an uncertain variable X and a number x, the notation $X \leq x$ represents the hypothesis that X has a value less than or equal to x. The (cumulative) probability distribution function, P(x), of the variable X is now defined as

$$P(x) \equiv \text{probability } (X < x). \qquad (0.12\text{-}1)$$

This definition applies to both discrete, i.e., variables that take on a countable number of values, and continuous variables. Frequently, we wish to have more detailed information than that provided by Equation (0.12-1). In particular, for a discrete variable we may wish to know the probability that X = x and, for a continuous variable, the probability that X falls between x and x + dx. Thus, we define the probability function for a discrete variable

$$p(x_j) = \text{probability } (X = x_j)$$

and the probability density function for a continuous variable

$$p(x) \equiv \frac{dP(x)}{dx} .$$

Clearly, the density function satisfies

$$\int_{-\infty}^{\infty} p(x)dx = P(\infty) - P(-\infty) = 1-0 = 1 \qquad (0.12\text{-}2)$$

and

$$P(x) = \int_{-\infty}^{x} p(s)ds \qquad (0.12\text{-}3)$$

(the integrals in Equations (0.12-2) and (0.12-3) are replaced by sums when X is discrete).

As an example of the above, let T be the time at which a particular piece of equipment first fails and let

$$P(t) = \text{probability that } T \leq t.$$

Then,

$$R(t) = 1-P(t),$$

known as the "reliability," is the probability that the equipment has not failed by time t.

The probability density function

$$p(t) = \frac{dP(t)}{dt} \qquad (0.12\text{-}4)$$

is the probability of failure, per unit time, at t. The "failure rate," or "hazard function," $\lambda(t)$ is defined as

$$\lambda(t) = \frac{p(t)}{R(t)} = \frac{1}{1-P(t)} \frac{dP(t)}{dt} .$$

The interpretation of the failure rate is that $\lambda(t)dt$ is the conditional probability that the equipment will fail in dt about t, given that it has not failed up until time t.

We now examine briefly several distributions of discrete and continuous types which are frequently encountered in risk analysis work.

0.12.2 DISCRETE DISTRIBUTIONS

0.12.2.1 Binomial

The binomial distribution is applicable when an experiment can have only two outcomes (e.g., success--failure such as in the case of a diesel generator starting or not). Let us say the frequency of failure is f and of success 1-f, and that an experiment is repeated n times. The following function p(r), then, gives the probability of exactly r failures in n trials, i.e.,

$$p(r) = \frac{n!}{r!\ (n-r)!} f^r(1-f)^{n-r} = \binom{n}{r} f^r(1-f)^{n-r}. \qquad (0.12\text{-}5)$$

0.12.2.2 Poisson

Items of equipment that operate continuously, pumps for example, are usually modeled as having a failure rate, λ, that is constant in time. In this case, the probability of having exactly k failures in t operating hours is given by:

$$p(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad (0.12\text{-}6)$$

viewed as a function of k this expression is known as the Poisson distribution.

0.12.3 CONTINUOUS DISTRIBUTIONS

0.12.3.1 Normal (Gaussian)

The Gaussian distribution, or normal curve of error, is fundamental in probability work.

The density function of this distribution is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad , -\infty < x < \infty. \tag{0.12-7}$$

To get the standardized (tabulated) normal distribution, define the new variable

$$z \equiv \frac{x-\mu}{\sigma} \tag{0.12-8}$$

in which case, Equation (0.12-7) becomes

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{z^2}{2}\right]. \tag{0.12-9}$$

Tables for the standardized distribution can be found in many textbooks (References 0-10 and 0-11).

0.12.3.2 Lognormal

The lognormal distribution is used extensively in risk and reliability work and is relevant in more physical processes where the underlying variable is restricted to positive values (i.e., 0 to ∞). In the present study, as in past safety studies, the lognormal is used to represent our state of knowledge of component failure rates and also to represent the variability, or frequency distributions, of populations of components.

The density function for the lognormal is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\ell nx-\mu)^2}{2\sigma^2}\right] \quad , 0 < x < \infty. \tag{0.12-10}$$

Comparison of Equations (0.12-7) and (0.12-10) reveals that if x is lognormally distributed, then ℓnx is normally distributed.

0.12.3.3 Exponential

Referring back to the Poisson (or constant failure rate) process, set k = 0. Then, the probability of zero failures up to time t is

$$R(t) = e^{-\lambda t}. \tag{0.12-11}$$

The density function is

$$p(t) = \lambda R(t) = \lambda e^{-\lambda t} \qquad (0.12\text{-}12)$$

which, we notice, has units of probability per unit time.

The cumulative distribution is

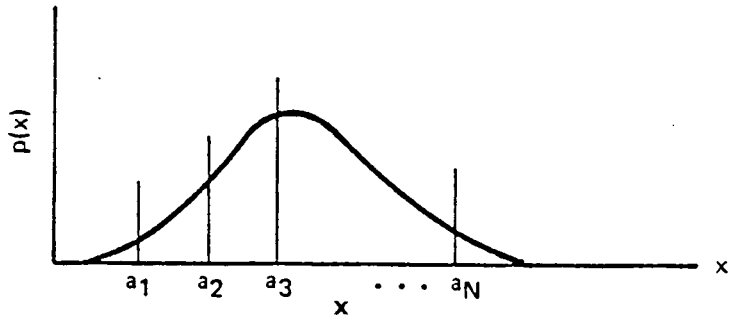
$$P(t) = 1 - e^{-\lambda t} \simeq \lambda t \qquad (0.12\text{-}13)$$

where the approximation holds for $\lambda t < 0.10$, as is usually the case in risk assessments. The implications of this approximation when uncertainties are propagated will be discussed in later sections.

0.12.4 DISCRETE APPROXIMATIONS TO CONTINUOUS DISTRIBUTIONS

Continuous distributions are, of course, the tools that we use to express our state of knowledge about continuous variables. For purposes of numerical calculation, however, it is convenient to approximate these continuous models by discrete distributions. This discretization is, of course, similar to the procedures used in evaluating integrals by numerical quadrature.

Consider the following distribution, p(x), of the continuous variable x.



If we wish now to get a discrete approximation to p(x), we can do this simply by carving x up into intervals as shown in the figure. The idea is to assign the probability that x will fall in an interval (a_{i-1}, a_i) to a single point x_i inside that interval. This probability, say p_i, is simply

$$p_i = \int_{a_{i-1}}^{a_i} p_X(x)dx. \qquad (0.12\text{-}14)$$

We can determine the points x_i in various ways. For example, x_i can be the mean value of the points in each interval. Thus, with the understanding

$$a_0 = -\infty, \qquad a_{N+1} = +\infty, \qquad (0.12\text{-}15)$$

we determine

$$x_i = \frac{1}{p_i} \int_{a_{i-1}}^{a_i} x p_x(x) dx. \qquad (0.12\text{-}16)$$

A second method is to simply take x_i as the midpoint of the interval, i.e.,

$$x_i = \frac{a_i + a_{i-1}}{2} \qquad (0.12\text{-}17)$$

or

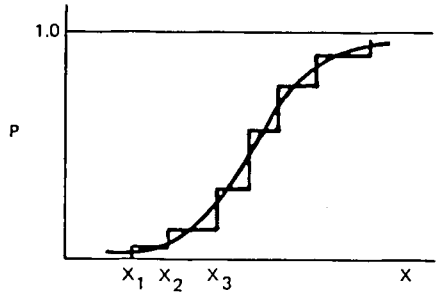
$$x_i = \sqrt{a_i a_{i-1}}. \qquad (0.12\text{-}18)$$

In this case we cannot use Equation (0.12-15). However, it will be satisfactory to choose a_0 and a_{N+1} appropriately, so that the probability that x falls outside the interval (a_0, a_{N+1}) will be negligibly small. With these finite values of a_0 and a_{N+1}, Equation (0.12-17) or (0.12-18) can be applied to all intervals.

The points x_i may also be determined using any other reasonable method that facilitates the calculations (since this is the major reason for the discretization). For example, if the lognormal distribution, Equation (0.12-10), is to be discretized, it will be convenient to take advantage of its relation to the normal distribution and the fact that the normal is tabulated. Thus, we work with the logarithm of x and we discretize the normal distribution and then switch back to the lognormal by taking exponentials.

The accuracy of the discretization increases as the number of intervals increases (i.e., for N large). The intervals do not have to be of equal length.

The above discussion has shown how to develop a discrete distribution from a continuous one. The reverse of this process, obtaining a continuous distribution from a discrete one, is simply a matter of "fitting" or "smoothing." A convenient way to do this is to plot the discrete distribution, in cumulative form, as a step function and then smooth in a sigmoid shape as shown below. This smoothed shape can then be differentiated graphically to obtain a density function.



0.12.5 MEASURES OF CENTRAL TENDENCY AND DISPERSION

We have seen that the proper way to express our state of knowledge about a random variable is to use a probability distribution. While a distribution gives in detail all that we know about the variable, occasionally it is convenient (mainly for communication purposes) to summarize the distribution by reporting one or more characteristic values which reflect its central tendency and its dispersion.

The most widely used measure of central tendency is the expected value (or mean), which is defined as

$$\alpha \equiv E[X] \equiv \begin{cases} \int_{-\infty}^{\infty} xp(x)dx \\ \text{or} \\ \sum_{i} x_i p_i \end{cases} \qquad (0.12\text{-}19)$$

according to whether x is continuous or discrete.

If the density function is interpreted as a mass distribution, then the expected value corresponds to the center of gravity. Besides the mean, there are two other measures of central tendency, the mode and median.

The mode (or most likely value) is defined for a discrete variable as the value where p_i is greatest, and for a continuous variable as the value at which the density p(x) is maximum.

The median is defined as that point x_{50} for which

$$P(x_{50}) = 0.50. \qquad (0.12\text{-}20)$$

Thus,

$$\int_{-\infty}^{x_{50}} p(x)dx = 0.50 \qquad (0.12\text{-}21)$$

or for a discrete variable

$$\sum_{x_i \leq x_{50}} p_i = 0.50. \qquad (0.12\text{-}22)$$

The percentile x_γ is defined as

$$P(x_\gamma) = \frac{\gamma}{100}. \qquad (0.12\text{-}23)$$

From Equation (0.12-21) we see that the median is the 50th percentile. Two percentiles, that are often used to indicate how broad the distribution is, are the 5th and 95th percentiles, which are determined by Equation (0.12-23) with γ = 5 and 95, respectively.

A measure of dispersion is the variance (and its square root, the standard deviation). It is defined to be the second moment about the mean; that is,

$$\beta^2 = E(X-\alpha)^2 = \begin{cases} \int_{-\infty}^{\infty} (x-\alpha)^2 p(x)dx \\ \text{or} \\ \sum_i (x_i-\alpha)^2 p_i \end{cases} . \qquad (0.12\text{-}24)$$

The variance is related to the mean and the second moment about zero by

$$\beta^2 = E[X^2] - \alpha^2 \qquad (0.12\text{-}25)$$

hence also

$$E[X^2] = \alpha^2 + \beta^2. \qquad (0.12\text{-}26)$$

This equation states that the mean of the square of a random variable is equal to the square of the mean of the variable plus its variance. This observation will be useful later in the quantification of fault trees.

0.12.6 THE LOGNORMAL DISTRIBUTION

Very often in risk analysis, the primary variables are assumed to be lognormally distributed. It is of interest, then, to investigate some useful properties of that distribution in the present context.

The lognormal density was given in Equation (0.12-10) and is repeated here

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ell nx-\mu)^2}{2\sigma^2}\right] . \qquad (0.12\text{-}27)$$

Note that this form contains two parameters, μ and σ. We sometimes write

$$x \sim \Lambda(\mu,\sigma)$$

to mean that x is lognormally distributed with parameters μ and σ.

Several characteristic values of the distribution are as follows:

Mean: $$\alpha = \exp\left(\mu + \frac{\sigma^2}{2}\right) \qquad (0.12\text{-}28)$$

Variance: $$\beta^2 = e^{2\mu+\sigma^2}\left[e^{\sigma^2}-1\right] = \alpha^2\left[e^{\sigma^2}-1\right] \qquad (0.12\text{-}29)$$

Mode: $$X_m = \exp(\mu-\sigma^2).$$

Inverting Equations (0.12-28) and (0.12-29), we get the parameters μ and σ as functions of the mean and variance, i.e.,

$$\sigma^2 = \ell n \left[\frac{\beta^2}{\alpha^2} + 1 \right] \qquad (0.12\text{-}30)$$

$$\mu = \ell n_{\alpha} - \frac{\sigma 2}{2}. \qquad (0.12\text{-}31)$$

Useful percentiles of the lognormal are:

5th Percentile: $x_{05} = \exp(\mu - 1.645\sigma)$ (0.12-32)

50th Percentile: (median) $x_{50} = e^{\mu} = \sqrt{x_{05} x_{95}}$ (0.12-33)

95th Percentile: $x_{95} = \exp(\mu + 1.645\sigma)$ (0.12-34)

Percentile: $x_{\gamma} = \exp(\mu + k_{\gamma}\sigma)$ (0.12-35)

where k_{γ} is the appropriate coefficient found in tables of the standard normal distribution. The error factor (EF) is defined as

$$EF = \sqrt{\frac{x_{95}}{x_{05}}} = e^{1.645\sigma}. \qquad (0.12\text{-}36)$$

It follows immediately that

$$x_{95} = x_{50}^{EF} \qquad (0.12\text{-}37)$$

and

$$x_{05} = \frac{x_{50}}{EF}. \qquad (0.12\text{-}38)$$

Since ℓnx is normally distributed, the most convenient way to look at a lognormal distribution is to write

$$x = me^{\sigma z}. \qquad (0.12\text{-}39)$$

When z in Equation (0.12-39) is a standard normal variate, then x is lognormally distributed. m is the median value of x and σ is called the lognormal standard deviation, or the "multiplicative" standard deviation. The reason for the latter term is seen in Equation (0.12-39) from the fact that if we increase z by the additive amount, 1.0, then x increases by the multiplicative factor e^{σ}. Thus, the multiplier e^{σ} plays the same role in a lognormal curve as the additive quantity σ plays in a normal curve. That is, when x changes by the factor e^{σ}, the cumulative probability changes by one standard deviation's worth.

Two important properties of the distribution are (see Reference 0-33, page 11)

- Property 1. If $X = \Lambda(\mu_x, \sigma_x)$ and $Y = C_1 X^{C_2}$, where C_1 and C_2 are constants, then

$$Y = \Lambda(\mu_y, \sigma_y) \tag{0.12-40}$$

where

$$\mu_y = C_2\mu_x + \ell n C_1 \tag{0.12-41}$$

$$\sigma_y = C_2\sigma_x. \tag{0.12-42}$$

This property simply states that if a lognormal variable is multiplied by a constant and raised to a power, the resulting variable will also be lognormal with parameters given by Equations (0.12-41) and (0.12-42)

We can easily prove this property by using Equation (0.12-39) to get

$$Y = C_1 m_x^{C_2} e^{C_2\sigma_x z}. \tag{0.12-43}$$

Equation (0.12-43) states that Y is also lognormal with median $C_1 m_x^{C_2}$ and lognormal standard deviation $C_2\sigma_x$. From Equation (0.12-33) we then get

$$C_1 m_x^{C_2} = e^{\mu_y} \tag{0.12-44}$$

and Equation (0.12-41) follows immediately.

- Property 2. If

$$X = \Lambda(\mu_x, \sigma_x)$$

and

$$Y = \Lambda(\mu_y, \sigma_y)$$

are independent lognormally distributed variables, and

$$V = X \cdot Y \tag{0.12-45}$$

then

$$V = \Lambda\left(\mu_x + \mu_y, \sqrt{\sigma_x^2 + \sigma_y^2}\right). \tag{0.12-46}$$

Using Equation (0.12-39) we get

$$V = m_x m_y \exp(\sigma_x z_x + \sigma_y z_y). \tag{0.12-47}$$

Since the sum of two normal curves is a normal curve, we have

$$\sigma_x z_x + \sigma_y z_y = \sigma_v z_v \qquad (0.12\text{-}48)$$

where

$$\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2} \quad . \qquad (0.12\text{-}49)$$

Thus, from Equation (0.12-47) we see that V is a lognormal variate with median $m_x m_y$ and lognormal standard deviation (0.12-49). Therefore, Equation (0.12-46) is proved.

These properties of the lognormal distribution will be used later in the quantification of system unavailabilities.

0.13 PROPAGATION OF UNCERTAINTIES, THE METHOD OF MOMENTS, AND THE METHOD OF DISCRETE PROBABILITY DISTRIBUTIONS

The use of logic diagrams (event and fault trees) in quantitative safety analyses leads to expressions for the unavailability Q_T of an event of interest in terms of the parameters X_i of more elemental events, i.e.,

$$Q_T = f(X_1, \ldots, X_n). \tag{0.13-1}$$

In general, the numerical values of the parameters X_i are not known with high precision. Therefore, we express our uncertainty about these values in terms of probability distribution functions.

Having determined the function f of Equation (0.13-1), the next step is to combine the distributions of the X_i's to get the distribution for Q_T. This step is known as the "propagation of uncertainties."

For an arbitrary function $f(X)_i$, a simple analytical form for the distribution of Q_T, of course, does not exist. However, simple numerical methods do exist. Among these the most widely used is the Monte Carlo Method (Reference 0-12). Two other methods, much more extensively used in this study, are the Method of Moments (Reference 0-13) and the Method of Discrete Probability Distributions (DPDs) (Reference 0-9). As background for discussion of these two methods, we first review the analytic expressions for the binary case, that is the case of combining two continuous and independent probabilistic variables.

0.13.1 COMBINING PROBABILITY DISTRIBUTIONS, ANALYTIC OR CONTINUOUS VARIABLE CASE

Let x and y be independent variables having the probability density functions $p_X(x)$, $p_Y(y)$. If $z = x + y$, then the density function for z is expressed by the convolution integral

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(z-x) dx. \tag{0.13-2}$$

Similarly if

$$z = x\,y \tag{0.13-3}$$

then

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y\left(\frac{z}{x}\right) \frac{1}{x}\, dx \tag{0.13-4}$$

(with any ambiguity at $x = 0$ handled by limit operations from both sides in the obvious way).

More generally, let

$$z = f(x, y) \tag{0.13-5}$$

where, for any specific values of z and x, y has a specific value denoted by

$$y = f^{-1}(z, x); \tag{0.13-6}$$

that is

$$z \equiv f\left(x, f^{-1}(z, x)\right) . \tag{0.13-7}$$

Then

$$p_z(z) = \int_{-\infty}^{\infty} p_x(x)\, p_y\left(f^{-1}(z,x)\right) \frac{\partial}{\partial z} f^{-1}(z, x)dx \tag{0.13-8}$$

which may be thought of as a more general form of convolution. Again, there are obvious further generalizations possible but this is sufficient for our purposes.

In real life applied work we rarely have the luxury of dealing with analytic forms and even in those rare cases may be unable to perform the the integrations (0.13-8) analytically. We are therefore led to seek approximate procedures. Two such procedures, the Method of Moments and the Method of Discrete Probability Distributions, are described in the next two sections. The moments method was used in this study in the system analyses primarily in connection with lognormal distributions. The discrete distribution method was used primarily in the Bayesian analysis of data, in the seismic analysis, and in the overall master assembly process of Section 0.8.4.

0.13.2 THE METHOD OF MOMENTS

The Method of Moments is an approximation method in which we abandon trying to get the system distribution from the component distributions and settle instead for calculating the moments of the system distribution from the moments of the component distributions. Thus, for example, if

$$z = x_1 + x_2, \tag{0.13-9}$$

then the first moment, or mean, of z is the sum of the first moments of x_1 and x_2

$$z = x_1 + x_2. \tag{0.13-10}$$

The variance, or second moment about the mean, is the sum of the component variances

$$v[z] = v[x_1] + v[x_2] . \tag{0.13-11}$$

For the case of a product

$$z = xy \ , \tag{0.13-12}$$

the first moment is the product of first moments

$$\bar{z} = \overline{xy} \tag{0.13-13}$$

and the variance is

$$v[z] = v[x]v[y] + x^2 v[y] + y^2 v[x] \ . \tag{0.13-14}$$

These results, (0.13-10), (0.13-11), (0.13-13), and (0.13-14), hold for arbitrary probability distributions $p_X(x)$, $p_Y(y)$, not just normal or lognormal distributions. To see this we need simply work, in the case of addition, with (0.13-2) as follows:

$$\bar{z} = E(z) = \int_{-\infty}^{\infty} z p_Z(z) dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z p_X(x) p_Y(z-x) dx dz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x + y) p_X(x) p_Y(y) dx dy$$

$$= E(x) + E(y). \tag{0.13-15}$$

Thus, the mean of a sum is the sum of the means for all distributions--not just normal distributions.

Similarly,

$$v(z) = \int_{-\infty}^{\infty} (z - \bar{z})^2 \, p_Z(z) dz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(x - \bar{x}) + (y - \bar{y})\right]^2 \, p_X(x) p_Y(y) dx dy$$

$$= v(x) + v(y). \tag{0.13-16}$$

Thus, the variance of a sum is the sum of variances--and this also is true for arbitrary distributions.

In the case of multiplication $z = xy$, the convolution is

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(x) p_Y\left(\frac{z}{x}\right) \frac{1}{x} \, dx \tag{0.13-17}$$

from which

$$z = \int_{-\infty}^{\infty} z p_z(z) dz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p_x(x) p_y(y) dx dy$$

$$= xy \ . \qquad (0.13\text{-}18)$$

Thus, the mean of a product is the product of means for arbitrary (not just lognormal) distributions.

Finally,

$$v(z) = \int_{-\infty}^{\infty} (z-\bar{z})^2 p_z(z) dz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy - \overline{xy})^2 p_x(x) p_y(y) dy$$

$$= (\overline{x^2})(\overline{y^2}) - (\bar{x})^2(\bar{y})^2$$

$$= [(\overline{x^2}) - (\bar{x})^2][(\overline{y^2}) - (\bar{y})^2] + (\bar{x})^2[(\overline{y^2}) - (\bar{y})^2]$$

$$+ (\bar{y})^2[(\overline{x^2}) - (\bar{x})^2] .$$

Thus,

$$v[z] = v[x]v[y] + (E[x])^2 v[y] + (E[y])^2 v[x] \qquad (0.13\text{-}19)$$

and again this holds for all distributions.

We have thus obtained the propagation rules for the case where the system relationship f is a sum or product of two variables. For a general system relationship, the method of moments may be derived from a Taylor Series argument as follows.

Let the system relationship be

$$Q = f(X_1, X_2, \ldots X_n). \qquad (0.13\text{-}20)$$

Now expand the function f about the mean values of its arguments in a multivariable Taylor series,

$$Q = f(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial X_i} (X_i - \bar{X}_i) + \frac{1}{2}\left(\sum_{i=1}^{n} \frac{\partial^2 f}{\partial X_i^2} (X_i - \bar{X}_i)^2 \right.$$

$$\left. + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial^2 f}{\partial X_i \partial X_j} (X_i - \bar{X}_i)(X_j - \bar{X}_j) \right) + \ldots , \qquad (0.13\text{-}21)$$

where $\bar{X}_i$ is the mean of X_i and all the partial derivatives are evaluated at the mean values of the X_i variables.

To find the mean of Q we simply take the expectation of both sides of (0.13-21) and get

$$\bar{Q} = f(\bar{X}_1, \ldots, \bar{X}_n) + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial X_i^2} \mu_2(X_i)$$

$$+ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial^2 f}{\partial X_i \partial X_j} E\left[(X_i - \bar{X}_i)(X_j - \bar{X}_j)\right] + \ldots , \qquad (0.13\text{-}22)$$

where

$$\mu_2(X_i) \equiv E\left[(X_i - X_j)^2\right] \equiv \text{ variance of } X_i$$

and

$$E\left[(X_i - X_i)\ (\bar{X}_j - \bar{X}_j)\right] \equiv \text{ covariance of } X_i \text{ and } X_j.$$

This covariance is zero when the variables are uncorrelated.

The variance of Q can be obtained from

$$\mu_2(Q) = E(Q^2) - \bar{Q}^2. \qquad (0.13\text{-}23)$$

Here $E(Q^2)$ can be obtained by squaring both sides of (0.13-20) expanding in Taylor series, and taking the expected value. Thus, we get

$$\mu_2(Q) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial X_i}\right)^2 \mu_2(X_i)$$

$$+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\partial f}{\partial X_j} \frac{\partial f}{\partial X_i} E\left[(X_i - \bar{X}_i)(X_j - \bar{X}_j)\right] + \ldots . \qquad (0.13\text{-}24)$$

We next show that these general formulae specialize to the results given previously for the case where f is a simple sum or product.

0.13.2.1 Sum of Probability Distributions

$$f(X_1 \ldots X_n) = \sum_{i=1}^{n} X_i.$$

- Mean. In this case, (0.13-22) yields

$$\bar{Q} = \sum_{i=1}^{n} \bar{X}_i \qquad (0.13\text{-}25)$$

i.e., the mean of the sum is the sum of the means.

- Variance. Equation (0.13-24) gives

$$\mu_2(Q) = \sum_{i=1}^{n} \mu_2(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E\left[(X_i - \bar{X}_i)(Y_j - \bar{Y}_j)\right] . \qquad (0.13\text{-}26)$$

For uncorrelated variables the last term of (0.13-26) is zero and we get

$$\mu_2(Q) = \sum_{i=1}^{n} \mu_2(X_i), \qquad (0.13.27)$$

i.e., for uncorrelated variables, the variance of the sum equals the sum of the variances.

0.13.2.2 Product of Distributions

In this case

$$f(X_1 \ldots X_n) = \prod_{i=1}^{n} X_i$$

- Mean. Equation (0.13-22) now becomes

$$\bar{Q} = \prod_{i=1}^{n} \bar{X}_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\bar{Q}}{\bar{X}_i \bar{X}_j} E\left[(X_i - \bar{X}_i)(X_j - \bar{X}_j)\right] \qquad (0.13\text{-}28)$$

which, for uncorrelated variables, gives

$$\bar{Q} = \prod_{i=1}^{n} \bar{X}_i . \qquad (0.13\text{-}29)$$

- Variance. From Equation (0.13-24) we get

$$\mu_2(Q) = \sum_{i=1}^{n} \left(\frac{\bar{Q}}{\bar{X}_i}\right)^2 \mu_2(X_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E\left[(X_i - \bar{X}_i)^2 (X_j - \bar{X}_j)^2\right]$$

$$+ 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\bar{Q}^2}{\bar{X}_i \bar{X}_j} E\left[(X_i - \bar{X}_i)(X_j - \bar{X}_j)\right] \qquad (0.13\text{-}30)$$

which, for uncorrelated variables, gives

$$\mu_2(Q) = \sum_{i=1}^{n} \left(\frac{\bar{Q}}{X_i}\right)^2 \mu_2(X_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mu_2(X_i)\mu_2(X_j) \qquad (0.13\text{-}31)$$

For the case of a product of two variables, Equation (0.13-31) specializes to our previous result, (0.13-14).

0.13.3 THE METHOD OF DISCRETE PROBABILITY DISTRIBUTIONS (REFERENCE 0-9)

The method of discrete probability distributions is used extensively in this study and thus far not widely used elsewhere. We therefore give a fairly complete exposition as follows.

0.13.3.1 Discrete Probability Distributions

Let x be an ordinary scalar variable and let $x_1, x_2, \ldots, x_n$ denote particular discrete values of x. Let $p_1, p_2, \ldots, p_n$ be associated probability values such that:

$$\sum_{i=1}^{n} p_i = 1. \qquad (0.13\text{-}32)$$

Then the set of doublets

$$\langle p_1, x_1\rangle, \langle p_2, x_2\rangle, \ldots \langle p_n, x_n\rangle = \{\langle p_i, x_i\rangle\} \qquad (0.13\text{-}33)$$

may be called a Discrete Probability Distribution (DPD) and may be thought of as a discrete approximation to a continuous probability density function p(x). (See Section 0.12.4.)

More fundamentally we need not introduce p(x) and may instead regard the DPD directly as an expression of our state of knowledge with respect to variable x. This is the point of view we shall take from here on.

We shall sometimes also refer to a set of doublets (0.13-33) as a probability "histogram." This usage, however, is explicitly not intended to suggest that the x_i should be regularly spaced or anything of that sort. On the contrary, coming from the point of view of the previous paragraph, we allow ourselves total freedom to select the x_i and p_i any way at all, save only that the set $\{\langle p_i, x_i\rangle\}$ adequately represents our state of knowledge and that it be suited to the numerical procedures we have in mind.

Thus, for example, suppose we were particularly interested in low values of x, i.e., in the low side tail of the distribution. We would then place several of the x_i within this low side tail, even though the corresponding p_i were small. In this way, we would ensure that the low values of x would be appropriately represented in our subsequent calculations.

0.13.3.2 Probabilistic Addition

Suppose the variables x, y, z, are related by:

$$z = x + y \tag{0.13-34}$$

and suppose our state of knowledge with respect to x and y is expressed by the DPDs

$$x = \{\langle p_i, x_i\rangle\}, \; y = \{\langle q_j, y_j\rangle\}. \tag{0.13-35}$$

Moreover, suppose these states of knowledge are independent in the sense that, if we found out the true value of y, this would not affect our DPD for x, and vice versa.

For this situation, we may now define the operation of addition of two DPDs as follows:

$$\{\langle p_i, x_i\rangle\} + \{\langle q_j, y_j\rangle\} = \{\langle p_i q_j, x_i + y_j\rangle\} \tag{0.13-36}$$

We now regard the set of doublets on the right as a DPD representing our state of knowledge of the variable z.

$$z = \{\langle r_{ij}, z_{ij}\rangle\}$$

where

$$r_{ij} = p_i q_j, \; z_{ij} = x_i + y_j.$$

Equation (0.13-36) then is our algorithm for probabilistic addition. It may be regarded as a discrete analog to the convolution operation (0.13-2). As an example of this algorithm, if

$$x = \{\langle .1, -1\rangle \langle .5, +1\rangle \langle .4, +2\rangle\}$$

$$y = \{\langle .2, 5\rangle \langle .8, 10\rangle\}$$

then

$$z = \{\langle .02, 4\rangle \langle .1, 6\rangle \langle .08, 7\rangle \langle .08, 9\rangle \langle .4, 11\rangle \langle .32, 12\rangle\}.$$

0.13.3.3 Probabilistic Multiplication

Similarly, we define multiplication of DPDs as:

$$\{\langle p_i, x_i\rangle\} \; \{\langle q_j, y_j\rangle\} = \{\langle p_i \, q_j, x_i \, y_j\rangle\}. \tag{0.13-37}$$

Thus if

$$z = xy$$

then

$$z = \{\langle r_{ij}, z_{ij}\rangle\} \tag{0.13-38}$$

with

$$r_{ij} = p_i q_j, \; z_{ij} = x_i y_j. \tag{0.13-39}$$

As an example, let us take the x and y of the previous paragraph. Then, z is the set of doublets:

$$z = \{\langle .02, -5\rangle \langle .1, 5\rangle \langle .08, 10\rangle \langle .08, -10\rangle \langle .4, 10\rangle \langle .32, 20\rangle\}$$

Equation (0.13-39) is the discrete analog of (0.13-4). In a similar way we can write the discrete version of (0.13-8). We summarize all this in the following section.

0.13.3.4 General Rule of Probability Arithmetic for Binary Operations

If

$$z = f(x, y) \tag{0.13-40}$$

where x, y are independent DPDs

$$x = \{\langle p_i, x_i\rangle\}, \; y = \{\langle q_j, y_j\rangle\}, \tag{0.13-41}$$

then z is the DPD

$$z = \{\langle r_{ij}, z_{ij}\rangle\} \tag{0.13-42}$$

where

$$r_{ij} = p_i q_j, \; z_{ij} = f(x_i, y_j). \tag{0.13-43}$$

We note in passing that

$$\sum_{ij} r_{ij} = 1.0$$

so that (0.13-42) is a bona fide DPD. We also note that obtaining the DPD for z is a simple matter of two nested "do loops" on a computing machine. We finally note the straightforward generalization to the case of more than two arguments in f. That is, if

$$z = f(x^1, x^2, \ldots x^M) \tag{0.13-44}$$

where each x^m is a DPD

$$x^m = \{\langle p^m_{i_m}, x^m_{i_m}\rangle\}, \tag{0.13-45}$$

then z is the DPD

$$z = \{\langle r_{i_1 \cdots i_M}, z_{i_1 \cdots z_{i_M}} \rangle\} \tag{0.13-46}$$

where

$$r_{i_1 \cdots i_M} = \prod_{m=1}^{M} p^m_{i_m} \tag{0.13-47}$$

$$z_{i_1 \cdots i_M} = f\left(x^1_{i_1}, x^2_{i_2}, \ldots x^M_{i_M}\right). \tag{0.13-48}$$

If the number of the variable, here M, is large, and f complicated, then the DPD approach, (0.13-48), becomes computationally burdensome. At this point, the Monte Carlo approach becomes more feasible. For the types of f and M that arise in this study, however, the DPD approach serves quite nicely.

0.13.4 PROBABILISTIC FUNCTIONS

Notice that what we have done so far is to develop an arithmetic for DPDs. The algebraic and functional notation associated with this arithmetic is identical to the usual notation for ordinary scalar variables. We simply reinterpret the symbols x, y, z, etc., as now standing for DPDs and the symbols +, ·, ÷, etc., as standing for appropriate operations between DPDs.

Thus all equations remain exactly the same. We see them now, however, through different eyes, interpreting them as equations and operations between sets of doublets. This interpretation is consistent with, in fact a manifestation of, the basic philosophical point of view which says that in computing real world phenomena we do not know any numerical values with complete certainty. Therefore, all our numbers should be replaced by probability distributions and all symbols representing numbers should thus be regarded as symbols representing probability distributions. All operations between such symbols thus now represent operations between probability distributions.

Now let us carry this a little further. Consider the statement

$$z = f(x). \tag{0.13-49}$$

If x is the DPD $\{\langle p_i, x_i \rangle\}$ we interpret (0.13-49) as saying that z is the DPD

$$z = \{\langle p_i, f(x_i) \rangle\}. \tag{0.13-50}$$

This is shown pictorially in Figure 0.13-1.

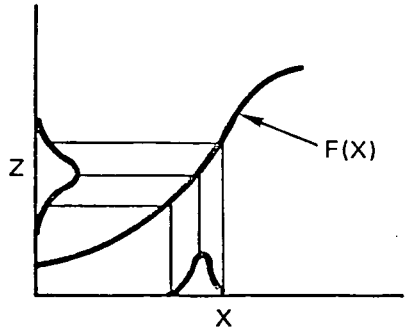


Figure 0.13-1. Probabilistic Input for Deterministic Function f

We interpret (0.13-49) as saying that we enter the curve f(x) with a probability distribution for x and emerge with a distribution for z. This interpretation holds, of course, for both discrete and continuous distributions.

Thus, probabilistic input gives probabilistic output. Now what happens if the function f itself is probabilistic?

By saying f is probabilistic we mean we have uncertainty as to the form of f. To proceed then, we must have a way of quantitatively expressing this uncertainty. This is simply done from the discrete point of view by setting out a series of different functional forms, f_j, and assigning probabilities to each. Thus we have a set of doublets:

$$f = \{\langle q_j, f_j \rangle\}, \quad \sum_j q_j = 1.0 \qquad (0.13\text{-}51)$$

which expresses our state of knowledge with respect to the form of f, and which we can regard as a histogram, or DPD, erected over the function space from which f comes.

Now returning to (0.13-49), if both x and f are probabilistic, we understand z to be the DPD.

$$z = \{\langle p_i q_j, \quad f_j(x_i) \rangle\} . \qquad (0.13\text{-}52)$$

Graphically, the DPD for f can be plotted as a "band", i.e., as a family of curves with probability as parameter (Figure 0.13-2).

We now thus think of ourselves as entering a probabilistic function with a probabilistic argument and emerging with a probabilistic output.

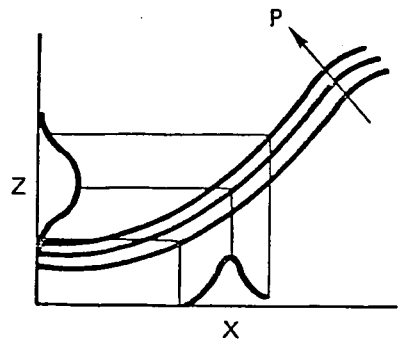


Figure 0.13-2. Probabilistic Input to Probabilistic Function f

0.13.5 FURTHER EXTENSIONS, VECTOR AND MATRIX VALUED VARIABLES

The pattern is now clear and it is a simple matter to extend to the case, say, where x is a vector valued variable. For example, if x is a column vector

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^N \end{bmatrix}$$

the set of doublets

$$x = \{\langle p_j, x_j \rangle\}$$

is simply regarded as a histogram on the space of N element column vectors, each x_j being a point in that space. Similarly if x and y are matrices, operators, or whatever, and if +, ÷, and f are understood to represent appropriate operations between such entities, then the probability arithmetic of Section 0.13.3 remains totally valid. Again the equations remain the same--we simply reinterpret the symbols.

This extension of the DPD approach to matrices is particularly useful in this study in light of the matrix format adopted in it for the assembly of final results (see Sections 8 and 0.8, especially 0.8.4.2).

0.13.6 COMMENT

We have now done enough to drive home the point that, by adopting the "set of doublets" point of view, we can "scale up" all our ordinary mathematics so that it has a probabilistic dimension. The scale-up simply involves a reinterpretation of the symbols we have written down in our ordinary analysis. The appearance of the writing itself, all the

equations, remains entirely the same. We now see the writing, however, differently. All quantities in the equations now become sets of doublets, and all the usual arithmetic, calculus, or matrix operations are now analogous operations defined between sets of doublets.

So everything goes through very neatly. There are two pitfalls, however, to be wary of in actually carrying out the numerical work. These we discuss next.

0.13.7 NONDISTRIBUTIVITY OF PROBABILISTIC OPERATIONS

In the ordinary arithmetic of numbers if

$$Q = x[y + z] \tag{0.13-53}$$

then also

$$Q = xy + xz \tag{0.13-54}$$

This property is referred to as the distributivity of the multiplication operation over addition. If x, y, and z are probability distributions, however, then (0.13-53) is not the same as (0.13-54).

If Equation (0.13-53) says:

1. First, add the distributions for y and z using, for example, (0.13-36).

2. Then, multiply the resulting distribution by the distribution for x using (0.13-37).

Equation (0.13-54) says:

1. Multiply the distributions for x and y using (0.13-37).
2. Multiply the distributions for x and z using (0.13-37).
3. Add the two resulting distributions using (0.13-36).

In this case the second procedure, (0.13-54) would give an incorrect result since the distributions (xy) and (xz) are not probabilistically independent. The distribution coming from procedure (0.13-54) would be narrower than it should be.

Thus probabilistic multiplication is not distributive, and in general one needs, in probabilistic arithmetic, to pay careful attention to the sequencing of operations. This must be understood in connection with the comment of Section 0.13.6.

0.13.8 DEPENDENCE AND INDEPENDENCE

The nondistributivity noted above is one example of a general feature that needs to be treated with great care in probabilistic calculations, namely dependency. An even simpler example is as follows:

Suppose $z = 2x$ (0.13-55)

and take for x the DPD:

$$x = \{<.2, 1><.6, 2><.2, 3>\} \quad (0.13\text{-}56)$$

then, according to our function rule (0.13-50),

$$z = \{<.2, 2><.6, 4><.2, 6>\}. \quad (0.13\text{-}57)$$

On the other hand, if we interpret (0.13-55) as

$$z = x + x \quad (0.13\text{-}58)$$

and use the addition rule (0.13-36), we get

$$z = \{<.04, 2><.24, 3><.44, 4><24, 5><.04, 6>\}, \quad (0.13\text{-}59)$$

a very different and much narrower DPD than (0.13-57). Thus, we see that, for probabilistic calculations, 2x is not the same as x + x. The difference is that when we did the operation x + x, we treated the two distributions as if they were independent. They are not. Thus, we obtained compensating uncertainties and a more narrow distribution than we should have.

This situation arises in risk calculations when we have similar components (see Section 0.16.5). For example, consider the system in Figure 0.13-3. We have

$$\lambda_s = \lambda_1 + \lambda_2. \quad (0.13\text{-}60)$$

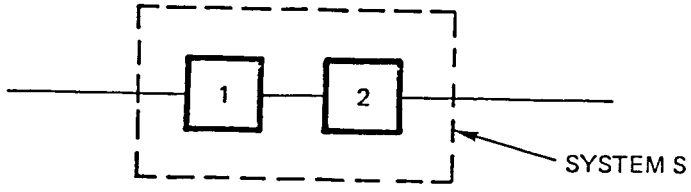


Figure 0.13-3. Series System

Now, suppose that components 1 and 2 are the same type of component, e.g., two valves of the same make and model. Our state of knowledge about λ_1 and λ_2 is therefore identical. That is to say, the DPD (or pdf) for λ_1 is the same as that for λ_2.

More importantly, these pdfs are not only identical, they are dependent in the sense that, if we were somehow to learn the true failure rate of valve 1, this would certainly affect our state of knowledge about λ_2.*

*Note, however, that this does not mean that we would know the value λ_2 exactly; for, although it is the same make and model, it is a physically distinct valve. Our DPD for λ_2, however, would be much narrower.

Thus, if the components are the same, it is important to write Equation (0.13-60) as

$$\lambda_s = 2\lambda_1 \text{ (not } \lambda_s = \lambda_1 + \lambda_1). \tag{0.13-61}$$

Similarly, in a parallel system, if the components are the same, it is important to write the system failure rate not as

$$\lambda_s = \lambda_1 \lambda_1, \tag{0.13-62}$$

but rather as

$$\lambda_s = \lambda_1^2.* \tag{0.13-63}$$

0.13.9 NUMERICAL CONSIDERATIONS, THE CONDENSATION OPERATION

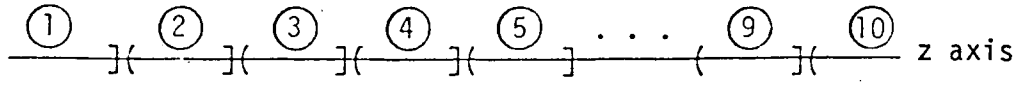
Consider again the probabilistic addition operation x + y = z, i.e.,

$$\{<p_i, x_i>\} + \{<q_j, y_j>\} = \{<p_i q_j, x_i + y_j>\}. \tag{0.13-64}$$

Observe that in passing through this operation the number of doublets in the z histogram is the product of the numbers in the x and y histograms. Thus if x and y each have 10 doublets, z would have 100.

At this rate if we had a series of operations to do, for example adding a string of histograms, we would very soon be into a serious storage and running time problem. The obvious answer to this problem is to introduce, after each probabilistic operation, a condensation or aggregation operation to reduce the number of doublets. This is done as follows.

In the preceding example, we now have a hundred-doublet histogram for the variable z. But for numerical work we can argue that we do not need a hundred doublets to represent our state of knowledge about z. Ten will do just fine. So we need only to aggregate the hundred down to ten. A useful way to do this is to define ten intervals on the z axis as indicated on the following diagram:



Note that we include the open-ended outer intervals among our ten. Now all we need to do is take all the points of the hundred-doublet histogram lying within the same interval and aggregate them into a single point.

*In doing this, note that we say, in effect, that our state of knowledge is totally dependent. In fact, it is not totally dependent since the components are physically distinct, as we pointed out in the last footnote. However, it is much better to err on the side of too much dependence rather than too little in our calculations, for too little will yield final DPDs that are too narrow, thus misleading us into an understatement of uncertainty.

Thus suppose

$$\langle p_5, z_5\rangle \langle p_6, z_6\rangle \cdots \langle p_{12}, z_{12}\rangle$$

all lie within a single one, the ith, of our ten intervals. We aggregate these doublets into a single new doublet $\langle p_i, z_i\rangle$ in the following way:

$$\hat{p}_i = \sum_5^{12} p_j \tag{0.13-65}$$

$$\hat{z}_i = \frac{1}{\hat{p}_i} \sum_5^{12} p_j z_j. \tag{0.13-66}$$

The condensation operation so defined preserves the total probability and the mean z within each interval.

Thus to add a series of histograms:

$$x^1 + x^2 + x^3 + \ldots x^n$$

we would simply add x^1 and x^2 probabilistically, put the resulting histogram through a condensation operation, then add the condensed histogram, representing $x^1 + x^2$, to x^3, condense again, and so on. At each stage we choose the intervals so as to retain the detail in the ranges of most interest.

This simple condensation trick eliminates many of the computational difficulties--certainly for all scalar operations. When z is a matrix valued or function valued variable, the condensation process is naturally more delicate--but similar strategies can be applied.

0.14 DATA ANALYSIS (DETERMINING THE FREQUENCY OF ELEMENTAL EVENTS)

Thus far, by the process of event tree/fault tree/cause table analysis, we have expressed the frequency of our scenarios in terms of the frequencies and occurrence fractions of various "elemental" events--initiating events and causes. It remains, therefore, to describe the methodology for determining these elemental frequencies and occurrence fractions.

Since the various elemental events are of different types, the methodology must be suited to the nature of each and to the information available about each.

0.14.1 TYPES OF INFORMATION AVAILABLE

In general, there are three types of information available for the frequency of elemental events:

- E_1: General engineering knowledge of the design and the manufacture of the equipment in question.
- E_2: The past experience in the specific plant being studied.
- E_3: The historical performance in other plants similar to the one in question.

The information of type E_1 and E_3 together constitutes the "generic" information and E_2 the "plant-specific" or "item-specific" information. Generic information is contained in various industry compendia such as WASH-1400, IEEE 500, and NUREG/CR-1363. Information of type E_2 can be culled from the plant records.

These three types of information are available, to various degrees of detail, for elemental events such as hardware failure rates, initiating events, human error rates, test and maintenance unavailabilities, abnormal environments (seismic, fires), etc. It remains now to discuss how to integrate and synthesize the three types of information. The central conceptual tool used for this purpose in this study is Bayes' theorem from the Theory of Probability. This theorem, being the fundamental law of logical inference, is the ideal tool (and, as a matter of principle, the only one) for quantitatively assessing the significance of the various items of information.

The methodology of this study makes extensive use of Bayes' theorem in several different applications. We illustrate first what we call the "one stage" application (References 0-14 and 0-15).

0.14.2 ONE-STAGE APPLICATION OF BAYES' THEOREM (DATA SPECIALIZATION)

To illustrate the way Bayes' theorem has been used in this study, suppose that we wish to develop a probability distribution, which expresses our state of knowledge about the failure rate λ of a specific valve to open on demand. We write Bayes' theorem in the form

$$p(\lambda|E_1E_2E_3) = p(\lambda|E_1E_3)\,\frac{p(E_2|\lambda E_1E_3)}{p(E_2|E_1E_3)} \qquad (0.14\text{-}1)$$

where

$p(\lambda|E_1E_2E_3)$ = the "posterior" probability distribution in light of evidence $E_1E_2E_3$.

$p(\lambda|E_1E_3)$ = the probability distribution in light of the generic information E_1 and E_3 only, i.e., the state of knowledge "prior" to the plant-specific evidence E_2.

$p(E_2|\lambda E_1E_3)$ = the "likelihood" that evidence E_2 would be observed if the value of the failure rate were λ, and given also the generic information.

$p(E_2|E_1E_3)$ = the probability that the specific evidence E_2 would be observed, given the generic information.

For purposes of numerical calculation we discretize the λ axis, establishing a set of possible values, λ_j. The denominator of Equation (0.14-1) can then be written as

$$p(E_2|E_1E_3) = \sum_j p(\lambda_j|E_1E_3)p(E_2|\lambda_jE_1E_3), \qquad (0.14\text{-}2)$$

and therefore (0.14-1) itself can be written as

$$p(\lambda_j|E_1E_2E_3) = \frac{p(\lambda_j|E_1E_3)\;p(E_2|\lambda_jE_1E_3)}{\sum_j p(\lambda_j|E_1E_3)\;p(E_2|\lambda_jE_1E_3)}\,. \qquad (0.14\text{-}3)$$

We see here that the denominator is the sum of the numerator terms so that the posterior probabilities automatically add to 1.0 as they should. It remains then only to describe how the numerator terms, the likelihood and the prior, are calculated. Consider first the likelihood. The evidence E_2 will be in the form:

E_2 = k failures out of n trials. (0.14-4)

The likelihood is therefore (from (0.12-5))

$$p(E_2|\lambda_jE_1E_3) = \frac{n!}{k!(n-k)!}\,\lambda_j^k(1-\lambda_j)^{n-k}. \qquad (0.14\text{-}5)$$

For a continuously operating system, the evidence is of the form:

E_2 = k failures in T hours.

In this case the likelihood is

$$p(E_2|\lambda_j E_1 E_3) = \frac{(\lambda_j T)^k}{k!} e^{-\lambda_j T} . \qquad (0.14\text{-}6)$$

Observe that this equation holds also for the case k=0. Thus, the evidence of zero failures is handled within the Bayesian framework just like any other evidence. Examples of this one stage Bayesian specialization procedure are contained in References 0-14 and 0-15.

0.14.3 DETERMINING THE PRIOR (OR GENERIC) DISTRIBUTIONS

The prior distributions $p(\lambda_j|E_1E_3)$ are obtained from published generic information, modified as described below.

For many components there does not exist a data source with a content and format that allows an unambiguous selection of the prior distribution. For example, in the published compendia, it is not always clear what failure modes are represented, what environments the data are applicable under, etc. It is evident, therefore, that judgment must be exercised in the derivation of the prior distribution.

Most sources give, for each failure rate, not a single number but a range or a distribution of some sort. For example, Table III 4-1 of the Reactor Safety Study gives, for the failure rate, λ, of pumps under normal environments, the range 3×10^{-6} to 3×10^{-4} failures per hour, which are described in the text as the 5th and 95th percentiles of a lognormal distribution. Indeed, most of the RSS data are given in the form of lognormal distributions. The question thus arises as to exactly how these distributions are to be interpreted.

We have chosen to interpret these distributions as "population variability curves," that is, as curves showing the variation of performance of individual components within the population (see Section 0.4.7). Thus, in the case of pumps, for example, we imagine that if we tested a large population, and plotted the frequency with which specific failure rates were measured, then that plot would turn out to be a lognormal curve with the 5th percentile, 3×10^{-6}, and 95th percentile, 3×10^{-4}.

These lognormal curves from the RSS, therefore, are regarded as the known results of experiments on populations. They are frequency distributions representing the variability stemming from different manufacturers, different models, different operating and maintenance conditions, as well as the random fluctuations occurring in presumably identical components.

Now, if we are interested in a specific component in a specific plant, then that component will have at its end of life a specific failure rate. At that time, then, we will know λ for that specific component. Until then we have a state of uncertainty about what λ is. We express that state of uncertainty with a probability distribution.

Now if all we know about the specific component, before we have any experience with it, is that it is one member of the population, then we would set our probability curve equal to the population variability curve. If all we know then is that we have a pump, we would say that our state of knowledge of the λ for that pump is expressed by a lognormal with the 5th and 95th percentiles equal to 3×10^{-6} and 3×10^{-4}.

Thus, our prior probability distributions are set numerically equal to the population variability curves.

After we have some operating experience with our specific component, we compute a posterior distribution which will be different from the population variability curve, and may in fact be thought of as expressing our posterior state of knowledge, in light of our new evidence, about where within the population our specific component falls.

This specialized distribution is the posterior distribution $p(\lambda \mid E_1 E_2 E_3)$ of Section 0.14.2.

0.14.4 TREATMENT OF THE GENERIC DISTRIBUTIONS FROM WASH-1400

As discussed in Section 0.14.3, the Reactor Safety Study provides lognormal distributions for failure rates. The 5th and 95th percentiles were the endpoints of the "assessed range" which was derived by the RSS from point estimates of each failure rate supplied by experts (organizations, individuals, etc.)

We have decided to broaden these curves by taking the endpoints of the assessed range as the 20th and 80th percentiles of a lognormal distribution. We thus express greater uncertainty since we allow a 40% chance to the possibility that the failure rate will be outside the assessed range. This we do for the following reasons.

In a previous site-specific study (Reference 0-14) we used the WASH-1400 curves (as given) as generic distributions. We then found that several posterior distributions, reflecting the evidence of the specific plant, lay in the tail region of the prior distributions on the high side. These results led us to the conclusion that the generic curves had to be broadened to reflect greater uncertainty.

References 0-17 and 0-18 provide further support to our decision. In Reference 0-17 the authors review experimental results that test the adequacy of probability assessments, and they conclude that "the overwhelming evidence from research on uncertain quantities is that people's probability distributions tend to be too tight. The assessment of extreme fractiles is particularly prone to bias." Referring to the Reactor Safety Study they state "The research reviewed here suggests that distributions built from assessments of the 0.05 and 0.95 fractiles may be grossly biased."

Commenting on judgmental biases in risk perception, Reference O-18 states

> "A typical task in estimating uncertain quantities like failure rates is to set upper and lower bounds such that there is a 98% chance that the true value lies between them. Experiments with diverse groups of people making many different kinds of judgments have shown that, rather than 2% of true values falling outside the 98% confidence bounds, 20 to 50% do so (Reference O-17). Thus, people think that they can estimate such values with much greater precision than is actually the case."

Our practice of broadening the generic distributions obtained from the literature is consistent with the conclusions of these studies and with our own experience. We adopted this practice as a general rule, but did not apply it invariably. In some few cases (e.g., items 46 and 47 in Table 1.5.1-5), the WASH-1400 distribution was judged to be broad enough to express our state of knowledge and was left unchanged.

The numerical effect of our broadening procedure is illustrated by the following example.

For the failure mode "failure of air-operated valves to remain open," WASH-1400 gives the range 2.8×10^{-8} to 2.8×10^{-7} per hour. Using these as the 5th and 95th percentiles, we get a median of 8.9×10^{-8} h^{-1} and a mean equal to 1.1×10^{-7} h^{-1}.

Our broadened distribution will be specified by the equations

$$\exp(\mu - 0.84\sigma) = 2.8 \times 10^{-8}$$

$$\exp(\mu + 0.84\sigma) = 2.8 \times 10^{-7}.$$

Thus, we get $\mu = -16.24$ and $\sigma = 1.37$. Therefore, our broadened distribution has the characteristic values shown in the following table:

	5%	Median	Mean	95%	95/5 Ratio
WASH-1400	2.8×10^{-8}	8.9×10^{-8}	1.1×10^{-7}	2.8×10^{-7}	10
Broadened Distribution	9.3×10^{-9}	8.9×10^{-8}	2.3×10^{-7}	8.4×10^{-7}	90

We see here that the median remains the same, as it must; the 95 to 5 percentile ratio goes from a factor of 10 to almost 100; and the mean value increases by a factor of 2 reflecting the extension of the high side tail of the curve.

0.14.5 TREATMENT OF THE GENERIC DISTRIBUTIONS FROM IEEE STD-500

This standard (Reference 0-16) contains data for electronic, electrical, and sensing components. The reported values were mainly synthesized from the opinions of some 200 experts (a form of the Delphi procedure was used). Each expert reported a "low," "recommended," and "high" value of the failure rate under normal conditions and a "maximum" value which would be applicable under all conditions (including abnormal ones). The pooling of the estimates was done using geometric averaging techniques, e.g.,

$$\lambda_{max} = \left[\prod_{i=1}^{n} \lambda_{max,\, i} \right]^{1/n} . \qquad (0.14\text{-}7)$$

This method of averaging was judged to be a better representation of the expert estimates, which were often given in terms of negative powers of ten. In effect, the usual arithmetic averages of the exponents were used.

The standard does not recommend a distribution. The method of averaging, however, suggests that the authors have in mind a lognormal distribution. Our task now is to determine this distribution from the given information.

The recommended value is suggested to be used as a "best" estimate. The word "best" is, of course, subject to different interpretations. We decide to use it as the median value mainly for two reasons. Firstly, for skewed, lognormal type distributions, the median is a more representative measure of central tendency than the mean, which is very sensitive to the tails of the distributions. Thus, we suspect that the experts who submitted their "recommended" estimates actually had in mind median values. Experimental evidence (Reference 0-19) also indicates that assessors tend to bias their estimates of mean values toward the medians. The second reason is that this choice is conservative, since the mean value of our resulting distribution is then larger than the "recommended" value. The "maximum" value is taken to be the 80th percentile of the lognormal distribution.

0.14.6 GENERIC DISTRIBUTIONS: OTHER SOURCES

An additional source of generic data is a series of NRC reports on pumps, valves, and diesel generators (References 0-20 through 0-22). They contain actual experience, and they do not rely on expert opinion.

These reports, when applicable, provide the mean failure rate of the generic distribution. However, they do not explicitly give a measure of plant-to-plant variability. We therefore used the mean values from the NRC reports when available, together with an error factor obtained from WASH-1400. This error factor we determine by using the WASH-1400 range as the 60% range as explained in Section 0.14.4.

0.14.7 TWO-STAGE APPLICATION OF BAYES' THEOREM

In the method of the previous sections the information E_3 is incorporated into the process in basically a judgmental way, through the use of prior distributions taken from industry compendia such as WASH-1400 and IEEE 500. This method was used for most of the components dealt with in this study. For the initiating events it was felt desirable to use a more sophisticated approach.

In this more sophisticated (Reference 0-23) approach the prior distribution $p(\lambda|E_1E_3)$, instead of being taken from the compendia, is itself derived from more fundamental information, namely E_3. This is done by introducing a "first stage" application of Bayes' theorem in which E_1 is the prior information and E_3 is the "current" information. The posterior distribution, which is the output of this first stage, then becomes, in effect, the prior distribution for a second stage use of Bayes' theorem, in which E_2 is the current evidence as in Section 0.14.2.

In this way the information E_3 is treated explicitly and quantitatively, rather than judgmentally. It is milked of all its information content, and the perennial question of "where do you get the prior from?" is addressed by pushing the prior back to a more fundamental level. We outline this approach as follows.

Consider a particular initiating event, say I_1, and suppose that we wish to say what can be inferred about the frequency λ_1, of this event at a specific plant, say plant j. This plant has already operated for T_j years, during which the event I, has occurred k_j times. The evidence E_2 is thus

$$E_2 = \langle k_j, T_j \rangle. \qquad (0.14\text{-}8)$$

Now suppose that plant j is one of a population of similar plants, 1, 2, 3,... M, and suppose that we consider the experience of these plants with respect to event I_1 as relevant to our assessment of the frequency at plant j. The experience of the mth plant in the population may again be summarized as a doublet:

$$\langle k_m, T_m \rangle , \qquad (0.14\text{-}9)$$

i.e., k_m occurrences in T_m years. The evidence E_3 is thus the set of such doublets,

$$E_3 = \{\langle K_1, T_1 \rangle, \langle k_2\ T_2 \rangle \ldots \langle k_m\ T_m \rangle\}. \qquad (0.14\text{-}10)$$

Now each plant in the population has its own "true" occurrence rate λ_m. Since the plants are not identical the true rates, λ_m will likewise be not identical. Considering the population, therefore, there will be a frequency distribution of λ over the population.

We now imagine that these m plants are part of, in fact samples from, a much larger population. Each member of this larger population has its own "true" occurrence rate λ, which varies from member to member. There is thus a frequency distribution , $\phi(\lambda)$, of λ over the population. This distribution, if we knew it, would constitute a prior probability distribution for λ_j.

Of course we do not know this distribution $\phi(\lambda)$ but we do have some information about it; namely, that the M sample plants selected from the population have had the experience E_3. From this information, using Bayes' theorem, we are able to express a posterior state of knowledge about the nature of the function $\phi(\lambda)$. From this state of knowledge, we can compute the distribution $p(\lambda|E_1E_3)$ which is the prior for the second stage as in Section 0.14.2.

0.15 HUMAN ERROR RATES

0.15.1 BASIC HUMAN ERROR RATES

The principal source of information is the NRC Human Reliability Handbook (Reference 0-24). This work is a substantial extension of the human reliability analysis contained in the Reactor Safety Study. It provides numbers for human error rates in numerous situations and it discusses at length the various factors that may influence human performance.

Despite the impressive amount of work that the handbook contains, the user should not forget that the numbers given are essentially the judgment of its authors and that they are not based on actual data. This is acknowledged by the authors, who state that the numbers "represent our best judgment based on our experience in complex systems (including nuclear power plants) and on our background in experimental and engineering psychology" (page 1-6, Reference 0-24). It is not surprising, therefore, that we have to use our own judgment as to how the information that is supplied by the handbook is to be used in our analysis.

For a specific human error rate the handbook usually provides a best estimate and upper and lower bounds. The use of a lognormal distribution is suggested with the two given bounds to be used as its 95th and 5th percentiles. The handbook points out that these are merely suggestions and that the users may, in some situations, wish to assign a larger uncertainty band.

We agree that, in most cases, the lognormal distribution is a satisfactory distribution to use, because "the performance of skilled persons tends to bunch up towards the low human error probabilities" (page 16-6, Reference 0-24). Unless otherwise stated we also use the lognormal distribution.

We determine the lognormal distribution by using the best estimate as the median and the upper bound as the 90th percentile, rather as the 95th percentile that the handbook recommends. This is consistent with our approach of expressing greater uncertainty about the error rates than the generic sources of data recommend.

Having made these decisions, we obtain the parameters μ and σ of the lognormal distribution from the equations

$$\exp(\mu) = \text{BE (Best Estimate)} \qquad (0.15\text{-}1)$$

$$\exp(\mu + 1.28\sigma) = \text{UB (Upper Bound)} \qquad (0.15\text{-}2)$$

The solution is

$$\mu = \ell n(\text{BE}) \qquad (0.15\text{-}3)$$

$$\sigma = [\ell n\ (\text{UB}) - \mu]/1.28 \qquad (0.15\text{-}4)$$

Example. For the rate of omission in nonpassive tasks (e.g., maintenance, test, etc.) when written procedures with checkoff provisions are used correctly, Table 15-2 of the handbook gives

Best Estimate (BE): 3×10^{-3}

Lower Bound (LB): 10^{-3}

Upper Bound (UB): 10^{-2}

when the procedure consists of more than ten special instruction items. From Equations (0.15-3) and (0.15-4) we get

$\mu = -5.81$ and $\sigma = 0.94$

Therefore, we have

$$\text{Mean} \equiv \alpha = \exp\left(\mu + \frac{\sigma^2}{2}\right) = 4.67 \times 10^{-3}$$

$$\text{Variance} \equiv \beta^2 = \alpha^2 [\exp(\sigma^2) - 1] = 3.08 \times 10^{-5}$$

$$95\text{th Percentile} = \exp(\mu + 1.645\sigma) = 1.41 \times 10^{-2}$$

$$5\text{th Percentile} = \exp(\mu - 1.645\sigma) = 6.40 \times 10^{-4}$$

We observe that our numbers are not very much different from those of the handbook.

Following the above procedure we derive the numbers of Table 0.15-1 for the human error rates, γ, that are frequently used in this study.

0.15.2 DEPENDENCE

When two or more tasks are to be performed, the question of dependence between human errors must be addressed. The handbook defines five levels of dependence as follows (Chapter 7):

- Zero Dependence (ZD): "The quality of performance, including nonperformance, of one activity has no effect on the performance of subsequent activities."
- Low Dependence (LD): "It is a convenient assumption to make when the dependence between actions is clearly greater than zero but not much greater."
- Moderate Dependence (MD): "... a level of dependence between LD and HD."
- High Dependence (HD): "It is a convenient assumption to make when the dependence between two actions is not complete but is definitely towards the higher end of the dependence continuum."

TABLE 0.15-1

BASIC HUMAN ERROR RATES
(Per Demand)

Task	NRC Handbook*	This Study			
		μ	σ	Mean	Variance
ERRORS OF COMMISSION					
1. Change or tag wrong valve where the desired valve is one of two or more adjacent, similar appearing manual valves, and at least one other valve is in the same state as the desired valve, or the valves are MOVs of such type that valve status cannot be determined at the valve itself.	5×10^{-3} (2×10^{-3} - 2×10^{-2})	-5.30	1.08	9×10^{-3}	1.8×10^{-4}
2. Change or restore wrong MOV switch or circuit breaker in a group of similar appearing items (in case of restoration, at least two items are tagged).	3×10^{-3} (10^{-3} - 10^{-2})	-5.81	0.94	4.7×10^{-3}	3.1×10^{-5}
3. General error of commission in nonpassive tasks such as maintenance, test, or calibration when written procedures are used.	3×10^{-3} (10^{-3} - 10^{-2})	-5.81	0.94	4.7×10^{-3}	$3\ 1\times10^{-5}$
ERRORS OF OMISSION					
1. Nonpassive tasks (maintenance, test, calibration); using procedures with checkoff provisions.					
i. Short list ($\leq$10 special instruction items).	10^{-3} (5×10^{-4} - 5×10^{-3})	-6.91	1.26	$2\ 2\times10^{-3}$	1.9×10^{-5}
ii. Long list (>10 special instruction items).	3×10^{-3} (10^{-3} - 10^{-2})	-5.81	0.94	4.7×10^{-3}	$3\ 1\times10^{-5}$
2. Passive tasks such as walk-around inspections.					
i. Failure to recognize an incorrect status when checking each item as he looks at it.	10^{-2} (5×10^{-3} - 5×10^{-2})	-4.61	1.21	2.2×10^{-2}	1.9×10^{-3}
ii. Failure to recognize an incorrect status when checking off several items after looking at several.	10^{-1} (5×10^{-2} - 5×10^{-1})	-2 30	1.26	$2\ 2\times10^{-1}$	1.9×10^{-1}

*Best Estimate (Range)

- Complete Dependence (CD): "Complete dependence between the actions of two people is rare, but not as rare as ZD. CD between two actions performed by the same person is more common."

The error rate γ_N for the Nth action, given failure on the (N-1)th action, is given by the following equations:

ZD: $\gamma_N = \gamma$ (the unconditional rate of Table 0.15-1) (0.15-5)

LD: $\gamma_N = \frac{1+19\gamma}{20}$ (0.15-6)

MD: $\gamma_N = \frac{1+6\gamma}{7}$ (0.15-7)

HD: $\gamma_N = \frac{1+\gamma}{2}$ (0.15-8)

CD: $\gamma_N = 1$ (0.15-9)

From the above summary it is evident that judgment needs to be exercised in deciding what the level of dependence between two specific tasks is and how the above equations are to be applied.

As an example we take the tasks of successively restoring valves to their proper position after test or maintenance. For these routine-type actions where written procedures are used, we judge that the level of dependence between the restoration of the first two valves is moderate and for all remaining valves is complete.

From Table 0.15-1 we find that the basic error of omission by an operator (with written procedures, short list) has a median of 10^{-3}, which, of course, is the best estimate listed in the handbook. We judge that Equations (0.15-5) through (0.15-9) have been derived mainly for best estimates, i.e., the best estimate for the conditional error of omission (MD) is

$$\gamma_1^{MD} = \frac{1+6\times10^{-3}}{7} = 0.144$$

Uncertainty bounds are not given in the handbook. We note that this uncertainty is mainly due to the use of the equation for γ_1 itself and not so much due to the variability of γ_0. If we assume a low level of dependence, we will get Equation (0.15-6)

$$\gamma_1^{LD} = \frac{1+19\times10^{-3}}{20} = 0.05$$

while, for a high level of dependence

$$\gamma_1^{HD} = \frac{1+10^{-3}}{2} = 0.50.$$

Since we believe that moderate dependence is appropriate, we take as the median of the conditional error of omission the value 0.144.

We express our uncertainty about this value by assigning an error factor of 5. The upper and lower bounds are then 0.72 and 0.029, respectively, and they include the point estimates for high and low dependencies derived above.

The parameters μ_1 and σ_1 of the lognormal distribution of γ_1 are

$$\mu = \ell n\ 0.144 = -\ 1.94$$

$$\sigma = \frac{\ell n 5}{1.645} = 0.98.$$

The error rate for nonrestoration of two valves is given by

$$\gamma_{TV} = \gamma_1 \gamma_0$$

where γ_1 and γ_0 are lognormally distributed. Consequently, γ_{TV} is also a lognormal variable with

$$\mu = -1.94 - 6.91 = -8.85$$

$$\sigma = \sqrt{0.98^2 + 1.26^2} = 1.60.$$

Therefore, the mean and variance of γ_{TV} are

$$\alpha_{\gamma_{TV}} = 5.1 \times 10^{-4}$$

$$\beta^2_{\gamma_{TV}} = 3.1 \times 10^{-6}$$

and the percentiles

5th Percentile: 10^{-5}

Median: 1.4×10^{-4}

95th Percentile: 2.0×10^{-3}.

0.15.3 HIGH STRESS SITUATIONS

The numbers that have been presented apply to normal tasks. When a high stress situation exists, e.g., following a LOCA, the response of the operators depends strongly on the time available, the information that they have, etc. These special cases are analyzed where they arise.

0.16 SYSTEM ANALYSIS

As outlined in Section 0.9, the determination of the split fractions for each system in the plant event tree is done by a process called system analysis. The present section adds a bit more detail about this process and, in particular, points out certain pitfalls which occur along the path of numerical computation of these split fractions and which must be handled correctly in order to avoid getting misleading results.

Because the systems in the plant differ greatly from each other, the analyses of different systems have a certain amount of difference in format. Nevertheless, all the systems analyses contain the following essential structural elements.

0.16.1 SYSTEM DESCRIPTION

The logic and the hardware of the system are described with the use of P&ID and block diagrams. The system configuration is often simplified by removing components, e.g., instrumentation lines, that are not pertinent to the required function of the system. Interfacing systems and test and maintenance requirements are also listed.

0.16.2 LOGIC MODEL

The meaning of different states of the system, i.e., the definition of the branches at each branch point, is of course done as part of the event tree. For those states which represent failure or partial failure a fault tree model of the system is constructed. The failure or partial failure state of the system becomes the "top event" of the fault tree. The fault tree logic then expresses the relationship between the states of the system and the states of the components of the system.

In the top structure of the fault tree certain "boundary conditions" are expressed under which the analysis is carried further. These boundary conditions define, in effect, "subtrees" within the system fault tree. For example, in Figure 0.16-1, the top structure of the tree shows explicitly the parts of the system that are disabled during maintenance (or testing, if applicable). Special attention is paid to the maintenance and test subtrees because the configuration of the system changes, the position of valves may change, and the system may be most vulnerable under these conditions.

As another example, different boundary conditions may represent the availability of electric power at various buses. When the system analyzed is fed by, say, two buses, there are four boundary conditions for the TOP event, namely, "both buses available," "bus 1 available and bus 2 unavailable," "bus 1 unavailable and bus 2 available," and "both buses unavailable." A separate subfault tree is developed in this case for each of the four boundary conditions.

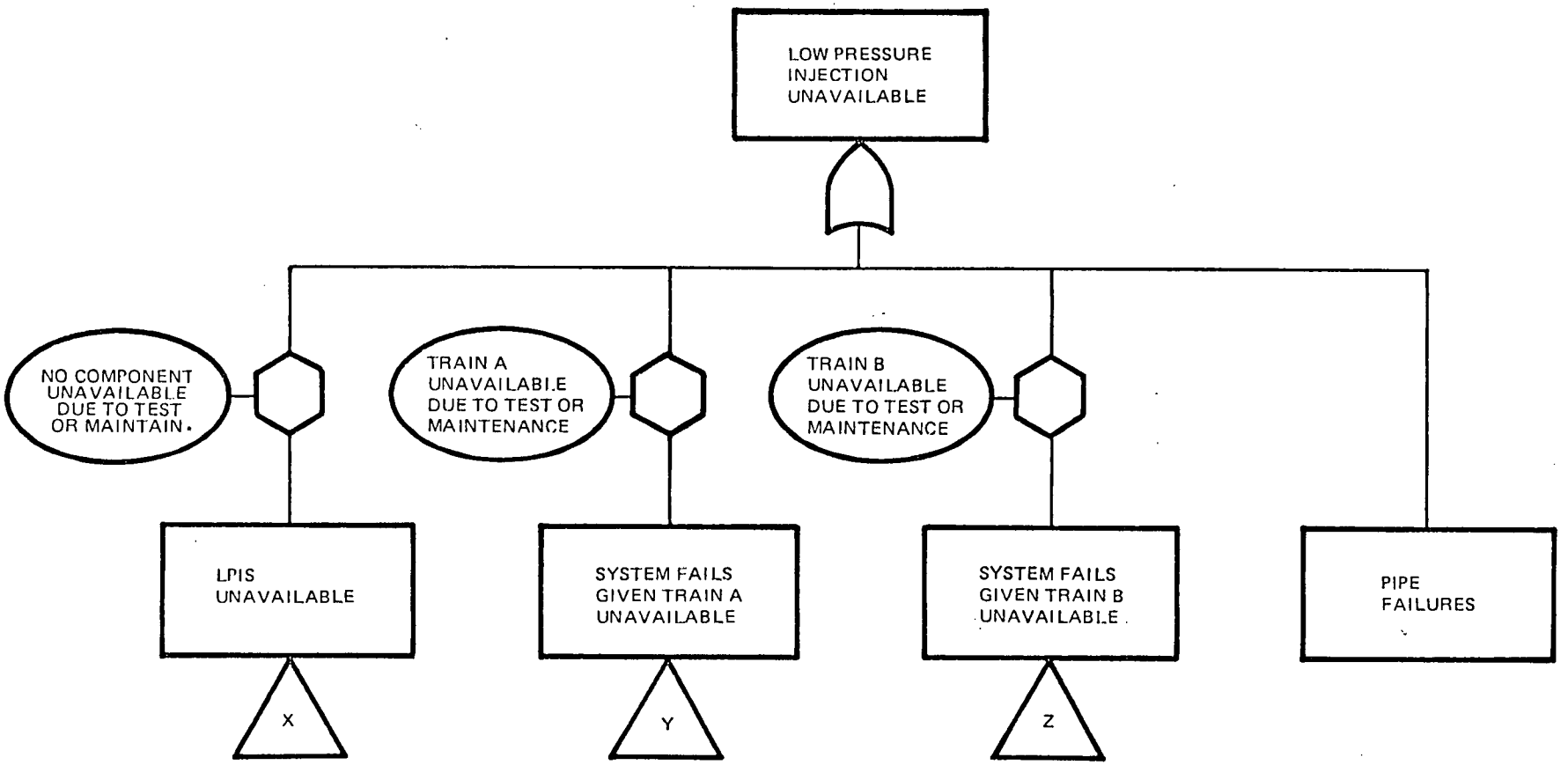


Figure 0.16-1. Top Structure of the Fault Tree

The system logic as illuminated by the fault tree analysis is finally expressed concisely in the form of the "minimal failure sets," or "minimal cutsets" of the system. These sets represent those minimum combinations of component failure which lead to system failure, i.e., to the TOP event. At this point the system logic part of the analysis is complete; that is, the minimal cutsets are logically equivalent to the TOP event.

0.16.3 CAUSES OF SYSTEM FAILURE

The fault trees and the minimal cutsets identify the failure modes of the system in terms of system components. We must now identify and quantify the failure causes of the system, i.e., causes that could make the components of a minimal cutset unavailable. This part of the analysis is where questions of completeness arise. By making the analysis systematic, we expect that the dominant failure causes will be identified. Furthermore, when all is said and done, we should think about the "other" category of causes, i.e., we should explicitly acknowledge our limitations and make some allowance for failure causes that we have not thought of.

To make the analysis systematic, we search for causes of single failures first, then double failures, etc. The latter refers to the order of minimal cutsets.

For each category of minimal cutsets we develop a table of causes and effects like the example shown in Table 0.16-1.

This table is developed for the second order minimal cutsets (double failures) of a system under the boundary conditions of both electric power buses being available. The table shows the causes of failure and their frequency, as well as their effects on components, on the system, on other systems, and on initiating events. This information alerts us to potential failure causes which can be common to more than one system.

The causes for which the analysis is more or less standard (but not always straightforward) are:

1. Hardware Failures. This is the standard treatment of "random independent" failures.

2. Test and Maintenance. This is where we use the branches Y and Z of the fault tree (Figure 0.16-1). Since technical specifications do not allow systems to be disabled during test and maintenance, additional failures must occur for the system to fail, i.e., at least one cause must exist in addition to the effects of test or maintenance for system failure.

3. Human Errors. The most common kinds are errors of omission and commission, e.g., forgetting to open or close valves after test or maintenance.

TABLE 0.16-1

SAMPLE CAUSE TABLE FOR DOUBLE FAILURES (BUSES AVAILABLE)

Cause	Mean Frequency	Effects			
		Components	System	Other Systems	Initiating Events
Coincident hardware failures	4.5×10^{-6}	mainly pumps	fails	no effect	no effect
Testing	10^{-10}	pumps	no effect	no effect	no effect
Maintenance and hardware failure	2.0×10^{-4}	pumps or M-V8700A, B	fails	no effect	no effect
Human error and hardware failure	8.2×10^{-9}	MOV8809A, B closed, failure on other side	fails	no effect	no effect
Other	4.6×10^{-5}	valves or pumps	fails	no effect	no effect
TOTAL	3×10^{-4}				

Dominant Contributor = Maintenance combined with hardware failure.

1099A

The frequency of these causes, human errors, hardware failures, etc., are determined from the available evidence and expert opinion as described in Sections 0.14 and 0.15. Since there is always uncertainty in these elemental frequencies, we express our state of knowledge about them, always, in the form of a probability distribution against frequency.

The question next arises as to how to combine these elemental probability of frequency curves into probability of frequency curves for failure of the whole system; these latter constitute, of course, the probability curves for the split fractions that we need at the event tree level.

0.16.4 SYSTEM QUANTIFICATION

This process of combining elemental frequencies into system failure frequencies, i.e., split fractions, is called system quantification. If the elemental frequencies were known with certainty, the process of combination is quite straightforward and based on the system logic as expressed in the fault tree or the minimum failure sets. When uncertainty in the elemental failure rates is taken into account, however, certain subtle pitfalls arise in this combination process, which, if they are not handled properly, can make customary combination procedures yield results substantially in error. These pitfalls have to do with dependencies or correlations between our state of knowledge of elemental failure rates and with certain convenient mathematical approximations which are customarily used, among them the use of lognormal shapes to represent probability distributions of elemental failure rates. The next several sections discuss these pitfalls in some detail.

0.16.5 PITFALLS RESULTING FROM CORRELATED VARIABLES

In the unavailability calculations, it is important not to overlook the correlation or dependence among variables for this would result in an underestimation of the uncertainties. The correlations we are speaking of here are not those due to dependence among events (common cause failures) but rather those stemming from the fact that for nominally identical components the states of knowledge that specify their failure rates are identical (Reference 0-25).

This correlation is easily understood when we recall our discussion on generic distributions and data specialization (Section 0.14). In deriving the specialized distribution for, say, the failure of motor-operated valves to open on demand we have assumed that there is a single value of the failure rate, q, which applies to all MOVs in our plant and, which is unknown to us. This state-of-knowledge uncertainty is expressed by the specialized distribution. The evidence from the plant is collected by aggregating all the failures of MOVs to open on demand and all the demands. This, of course, is consistent with the assumption of the existence of a single value for this particular failure rate.

Suppose, now, that a system fails when one of two MOVs in series fails to open. In conventional calculations we would evaluate this contribution to system unavailability by

$$Q = q_1 + q_2 \qquad (0.16\text{-}1)$$

where q_1 and q_2 are independent variables having the same distribution. Equation (0.16-1), however, ignores the correlation described above. We will be consistent with our model when we replace Equation (0.16-1) by

$$Q = 2q. \qquad (0.16\text{-}2)$$

Equation (0.16-2) expresses the fact that the failure rates of the two MOVs are identical (more precisely, that the probability distributions for the two failure rates express the same state of knowledge and are thus dependent).

Similarly, for two events in parallel, a conventional calculation would be

$$Q = q_1 q_2 \qquad (0.16\text{-}3)$$

while the correct equation is

$$Q = q^2. \qquad (0.16\text{-}4)$$

To further appreciate these differences, we calculate the mean and variance of Q (which we denote by

$$\alpha_Q \text{ and } \beta_Q^2)$$

in terms of the mean and variance of q (which we denote by

$$\alpha_q \text{ and } \beta_q^2).$$

- Series Configuration

  In Equation (0.16-1) we have the sum of two variables. According to Section 0.13 we get

  $$\alpha_Q = \alpha_{q_1} + \alpha_{q_2} = 2\alpha_q \qquad (0.16\text{-}5)$$

  and

  $$\beta_Q^2 = \beta_{q_1}^2 + \beta_{q_2}^2 = 2\beta_q^2 . \qquad (0.16\text{-}6)$$

  On the other hand, Equation (0.16-2) gives

  $$\alpha_Q = 2\alpha_q \qquad (0.16\text{-}7)$$

and

$$\beta_Q^2 = 2^2 \beta_q^2 = 4 \beta_q^2. \qquad (0.16\text{-}8)$$

Comparing Equations (0.16-5) and (0.16-7) we conclude that the mean value is unaffected when the correlation of the variables is taken into account. However, Equations (0.16-6) and (0.16-8) show that the uncertainty is underestimated if the correlation is ignored.

- Parallel Configuration

In the multiplicative case, Equation (0.16-3), we get

$$\alpha_Q = \alpha_{q_1} \alpha_{q_2} = \alpha_q^2. \qquad (0.16\text{-}9)$$

From Equations (0.16-4) and (0.12-26), we get

$$\alpha_Q = \alpha_q^2 + \beta_q^2. \qquad (0.16\text{-}10)$$

Comparing Equations (0.16-9) and (0.16-10) we conclude that even the mean value is underestimated when the correlation is not properly accounted for.

The effect of using (0.16-3) versus (0.16-4) on the "spread" or variance of Q can be further illuminated in the important case of lognormally distributed q as follows.

- Lognormal Case

In Section 0.12 we discussed a convenient way to think about the lognormal distribution (Equation (0.12-40)). Using Equation (0.12-43) with $C_1 = 1$ and $C_2 = 2$, we get for Q of Equation (0.16-4)

$$Q = q^2 = m^2 e^{2\sigma z}. \qquad (0.16\text{-}11)$$

Therefore, q^2 is also lognormal with median m^2 and lognormal standard deviation 2σ.

Also, if q_1 and q_2 are two lognormal variates

$$q_1 = m_1 e^{\sigma_1 z_1}, \quad q_2 = m_2 e^{\sigma_2 z_2}, \qquad (0.16\text{-}12)$$

then (see Section 0.12)

$$q_1 q_2 = m_1 m_2 e^{\sigma_1 z_1 + \sigma_2 z_2}. \qquad (0.16\text{-}13)$$

Since the sum of two normal curves is a normal curve we have

$$\sigma_1 z_1 + \sigma_2 z_2 = \sigma_3 z_3$$

where

$$\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2} \quad . \qquad (0.16\text{-}14)$$

Thus, from (0.16-3) we see that

$$Q = q_1 q_2 = m_1 m_2 e^{\sigma_3 z_3},$$

i.e., that Q is a lognormal variate with median $m_1 m_2$ and lognormal standard deviation (0.16-14).

Now if q_1 and q_2 have the same distribution

$$m_1 = m_2 = m$$

$$\sigma_1 = \sigma_2 = \sigma$$

then

$$Q = m^2 \, e^{\sqrt{2}\sigma z}. \qquad (0.16\text{-}15)$$

Comparing (0.16-15) with (0.16-11) we see that using Equation (0.16-3) rather than Equation (0.16-4) leads to a tighter distribution for Q. Thus, neglect of the dependence of the probability distributions for q_1 and q_2 leads to a severe understatement of the uncertainty in Q! The median of Q is correct, but the distribution about that median is too narrow, and leads to an erroneous and unconservative calculation of the mean value of Q.

0.16.6 ERRORS IN MEANS AND VARIANCES RESULTING FROM USE OF LOGNORMAL DISTRIBUTIONS

The lognormal is very convenient mathematically, but has the defect that the mathematical curve has a tail extending to infinity, whereas the true state of knowledge distribution truncates at finite values. (For example, a true failure rate on demand cannot be greater than 1.0). The presence of this tail can be regarded as a mathematical artifact which, in most cases, does not distort the results of the analysis. In other cases, however, it does. The purpose of this section, in fact, is to point out that this artifact can result in severe error in calculated means and variances even while the usual percentiles are unaffected (Reference 0-25).

To demonstrate the sensitivity of the mean value to the tails of distributions we can use the following simple example. Let the variable X take on two values as follows:

$$\Pr(X = 10^{-2}) = 0.9999 = 1 - 10^{-4}$$

$$\Pr(X = 2) = 10^{-4}.$$

The mean value of X is, then,

$$\alpha_X = 10^{-2} \times 0.9999 + 2 \times 10^{-4} = 1.020 \times 10^{-2}.$$

The major part of this value comes from the point 10^{-2}, which has the highest probability.

Suppose, now, that we are interested in the cube of X. Then

$$\Pr(X^3 = 10^{-6}) = 0.9999$$

$$\Pr(X^3 = 8) = 10^{-4}.$$

The mean value is now

$$\alpha_{X3} = 10^{-6} \times 0.9999 + 8 \times 10^{-4} = 8.009 \times 10^{-4},$$

a result that is completely dominated by the value X = 2. If this value happens to be unrealistic but tolerated due to some modeling approximation that assigns to it a very small probability, it will distort any calculations that include X^3. Situations like this do arise in risk analysis in the following situations.

- The Exponential Approximation

  It is common practice to treat the time-to-failure of items as an exponentially distributed random variable, i.e., the failure distribution is

  $$F(t) = 1 - e^{-\lambda t} \qquad (0.16\text{-}16)$$

  where λ is the failure rate. Due to the smallness of λ, F(t) is almost always approximated by

  $$F(t) \cong \lambda t, \text{ for } \lambda t < 0.10. \qquad (0.16\text{-}17)$$

  The requirement that λt be less than 0.10 is often forgotten since the most likely values of λ are indeed small. However, if a lognormal distribution is used for λ, then λt is unbounded on the right. The condition of Equation (0.16-17) then, is violated by the high percentiles resulting in an overestimate of the mean and variance.

- Frequency of Failure on Demand

  The frequency of failure on demand (unavailability), q, can take on values in the interval (0,1), while the lognormal distribution is defined over the interval $(0, \infty)$. Even though the probability of the (nonphysical) values $(1, \infty)$ is usually negligibly small, this

tail may distort the mean and variance just as in the preceding case. To prevent this distortion in the current study we truncate the lognormal distribution at q = 1 or less.

0.16.7 EXPRESSION OF SYSTEM FAILURE RATES IN TERMS OF COMPONENT FAILURE RATES. THE SUPERCOMPONENT IDEA. NECESSITY FOR PROPER TREATMENT OF DEPENDENCY

The results of the preceding section suggest that the unavailability contribution from hardware failures cannot be calculated from the minimal cutsets in a mechanical way, because of the correlations, which the standard computer codes ignore.

To demonstrate the effects of correlation, we use the example of Figure 0.16-2.

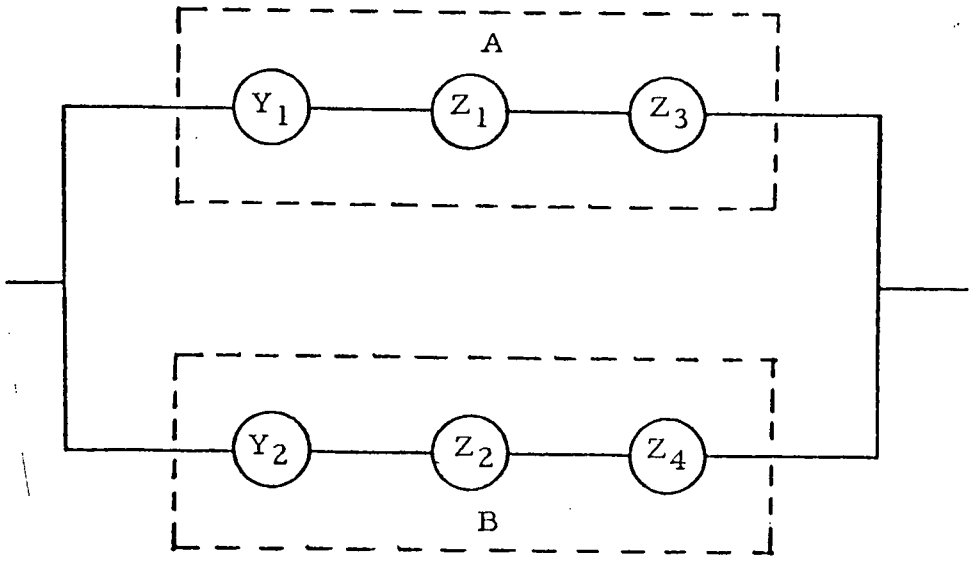


Figure 0.16-2. Sample Case of Supercomponents

The system consists of two parallel identical trains. In particular, the Y components are similar, as well as the Z components. Here Y and Z denote failure rates on demand. Each is lognormally distributed with the following characteristics:

Y: $\mu = -6.96$ $\sigma = 0.698$

mean value = $\alpha_Y = 1.21 \times 10^{-3}$

variance $\beta_Y^2 = 9.28 \times 10^{-7}$

Z: $\mu = -9.26$ $\sigma = 0.698$

$\alpha_Z = 1.21 \times 10^{-4}$; $\beta_2^2 = 9.28 \times 10^{-9}$.

There are nine minimal cutsets of order two, namely,

Y_1Y_2	Z_1Y_2	Z_3Y_2
Y_1Z_2	Z_1Z_2	Z_3Z_2
Y_1Z_4	Z_1Z_4	Z_3Z_4

A conventional calculation of the mean unavailability of the system, α_Q, ignores the correlation and is based on the expression

$$Q = (Y_1 + Z_1 + Z_3)(Y_2 + Z_2 + Z_4). \qquad (0.16\text{-}19)$$

Evaluating this expression we get

5th Percentile: 4.2×10^{-7}

Median: 1.4×10^{-6}

Mean: 2.1×10^{-6}

95th Percentile: 5.8×10^{-6}.

The correct expression, which replaces Equation (0.16-19), is

$$A = (Y + 2Z)^2 \qquad (0.16\text{-}20)$$

which gives

5th Percentile: 2.4×10^{-7}

Median: 1.5×10^{-6}

Mean: 3.1×10^{-6}

95th Percentile: 10^{-5}.

We see, therefore, that ignoring the correlation leads to substantial inaccuracy.

Equation (0.16-20) suggests that we define supercomponents A and B (Figure 0.16-1) as

$$A \equiv \{<Y_1, Z_1, Z_3>\}$$

$$B \equiv \{<Y_2, Z_2, Z_4>\}.$$

The two supercomponents are identical. The mean and variance of A are

$$\alpha_A = \alpha_Y + 2\,\alpha_Z = 1.452 \times 10^{-3}$$

$$\beta_A^2 = \beta_Y^2 + 4\,\beta_2^2 = 9.466 \times 10^{-7}.$$

The double failures of the system are due to failure of both supercomponents, i.e.,

$$Q = A^2. \qquad (0.16\text{-}21)$$

The mean of Q is then

$$\alpha_Q = \alpha_A^2 + \beta_A^2 = 3.1 \times 10^{-6}$$

which is the correct value.

0.16.8 TEST AND MAINTENANCE CONTRIBUTION TO SYSTEM UNAVAILABILITY

0.16.8.1 Test Contribution

Referring to Figure 0.16-2, suppose that every T hours (usually T = 720 h) supercomponents A and B are tested consecutively. There is a contribution to system unavailability only if the test disables the supercomponent. Let the mean duration of testing be τ. System failure occurs when A is being tested and B fails, or B is being tested and A fails. We have

Unavailability of A due to testing: frequency = $\frac{\tau}{T}$

B fails on demand: frequency $\equiv q_B = q_A$;

therefore,

$$Q_{test} = 2\,\frac{\tau}{T}\,q_A. \qquad (0.16\text{-}22)$$

The mean unavailability due to testing is then

$$\alpha_{Q_{test}} = 2\,\frac{\alpha_\tau}{T}\,q_A \qquad (0.16\text{-}23)$$

and the variance is

$$\beta^2_{Q_{test}} = \frac{4}{T^2}\left[\alpha^2_{q_A}\beta^2_\tau + \alpha^2_\tau\,\beta^2_{q_A} + \beta^2_\tau\,\beta^2_{q_A}\right]. \qquad (0.16\text{-}24)$$

Usually the uncertainty on τ is negligible and Equation (0.16-24) is replaced by

$$\beta^2_{Q_{test}} = 4 \left(\frac{\alpha_\tau}{T}\right)^2 \beta^2_{q_A}. \qquad (0.16\text{-}25)$$

0.16.8.2 Maintenance Contribution

Suppose now that supercomponent A is under maintenance with frequency f. The mean duration of maintenance is m. Then A is unavailable due to maintenance with frequency fm. The system fails when B fails on demand, therefore

$$Q_{maintenance} = 2\ fmq_A \qquad (0.16\text{-}26)$$

where the factor of 2 accounts for the symmetric event (B is under maintenance and A fails).

The mean unavailability contribution is then

$$\alpha_{Q_{maintenance}} = 2\ \alpha_f\ \alpha_m\ \alpha_{q_A}. \qquad (0.16\text{-}27)$$

The variance is calculated using Equation (0.13-19) twice. Note that we do include in the calculations the uncertainty in f, which WASH-1400 does not.

0.16.9 HUMAN ERROR CONTRIBUTIONS TO SYSTEM FAILURE RATE

A common kind of human error is forgetting to return the tested item to its proper configuration thus rendering that train unavailable (error of omission). Again referring to Figure 0.16-2, suppose that γ_0 is the human error rate of omission and γ_1 the conditional rate of repeating the error (see also Section 0.15). The system unavailability due to human error is then

$$Q_{HE} = \gamma_0\ \gamma_1 \qquad (0.16\text{-}28)$$

with mean and variance

$$\alpha_{Q_{HE}} = \alpha_{\gamma_0}\ \alpha_{\gamma_1} \qquad (0.16\text{-}29)$$

and

$$\beta^2_{Q_{HE}} = \alpha^2_{\gamma_0}\ \beta^2_{\gamma_1} + \alpha^2_{\gamma_1}\ \beta^2_{\gamma_0} + \beta^2_{\gamma_0}\ \beta^2_{\gamma_1}. \qquad (0.16\text{-}30)$$

These equations do not include the possibility of discovery (in the control room or otherwise) of the wrong position of the items, i.e., the disabled item will remain unavailable for the whole period T until the next test. If the mean time to discovery is τ_D, then Equation (0.16-28) becomes

$$Q_{HE} = \gamma_0 \gamma_1 \frac{\tau_D}{T}. \qquad (0.16\text{-}31)$$

PROBABILISTIC RISK ASSESSMENT

METHODOLOGY

Part 3

EXTERNAL EVENTS

0.17 SEISMIC ANALYSIS METHODOLOGY

Section 0.8.3 pointed out that ground accelerations of various magnitudes can be looked upon as just another type of initiating event. From that point of view the analysis of seismic events falls naturally within our overall matrix formulation and can be regarded as included in what has gone before. Because of the unique nature of this type of IE, however, the present section presents more details on the specific features of the seismic methodology.

0.17.1 METHODOLOGY

A seismic safety analysis consists of five main steps:

1. Seismicity: determination of how frequently ground motions of various sizes occur at the site.

2. Fragility: determination of the ability of various structures and equipment in the plant to survive earthquakes of various sizes.

3. Plant Logic: determination of the effects of various structural and equipment failures on the behavior of the plant.

4. Initial Assembly: combining the above three types of information into curves showing the likelihood of occurrence of various plant states as a result of an earthquake.

5. Final Assembly: further combination of these curves with the results of the containment model to obtain the likelihoods of various release categories, and these, in turn, with the site-specific release consequences model to obtain the final seismic risk curve.

0.17.2 SEISMICITY

We choose to characterize earthquakes in terms of a single parameter, a, the ground level acceleration. In terms of this parameter, the seismicity of a site may be characterized by a curve such as Figure 0.17-1 which shows for each acceleration, a, the frequency $\Phi(a)$ in times per year, that an earthquake of that acceleration or larger occurs at the site.

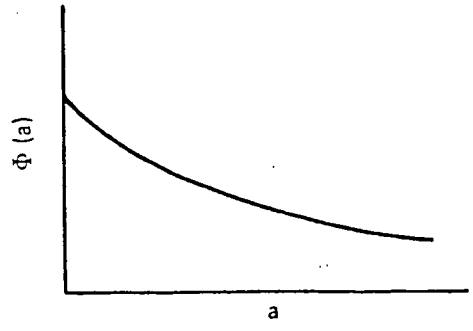


Figure 0.17-1. Seismicity Curve (Deterministic)

In using this mode of characterizing seismicity, we have adopted a certain traditional model of earthquake occurrence. In this model, quakes occur as a realization of an underlying random process. Thus, suppose at acceleration a_1, the frequency $\Phi(a_1)$ is 10^{-5}. We are saying that if we could see a record of the earth's history over millions of years (past and future), we would find that earthquakes of size a_1 or larger occurred at this site at random intervals with average frequency of once in one hundred thousand years.*

The curve $\Phi(a)$ for the site is supplied by seismologists from whatever recorded history and knowledge of crustal dynamics is available. Now, since records have been kept for only a short time, geologically speaking, this history and this knowledge are limited. Thus, the seismologists do not know the curve $\Phi(a)$ with great accuracy. To tell the truth then, about our state of seismic knowledge, and in keeping with the overall philosophy of this study, we need to quantitatively express our uncertainty about the curve $\Phi(a)$. The most convenient format for this expression is to put forth a family of curves, $\Phi_i(a)$, and assign a probability, p_i, to each. Thus we obtain Figure 0.17-2.

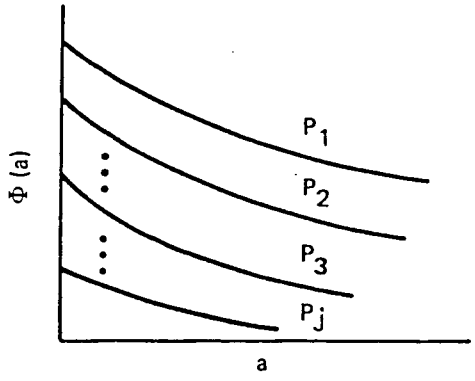


Figure 0.17-2. Family of Seismicity Curves

We may express this family of curves as a set of doublets:

$$\{<p_i, \Phi_i>\}$$

where Φ_i stands for the whole curve $\Phi_i(a)$, and

$$\sum_{i=1}^{J} p_i = 1.0. \qquad (0.17\text{-}1)$$

*A more ideal model would consider the detailed state of stress, the building up and release of tension in the earth's crust in the vicinity of the plant. Our current knowledge of crustal dynamics is not yet sufficient to support this level of modeling.

Observe that this set of doublets is nothing more than a discrete probability distribution (DPD) on the space of curves $\Phi(a)$. Note also that this is an instance of usage of the "Probability of Frequency" framework.

0.17.3 FRAGILITY

By fragility we mean the likelihood of failure as a function of effective ground acceleration for plant structures, equipment, and other components. Let us focus now on a specific structural component of the plant, say for example, a particular "shear wall." We wish to know if this wall will fail under earthquake size a. We have a definite meaning assigned to the word "fail," so it is now a question of structural analysis--analyzing the movements and stresses in the wall when subjected to earthquake size a. The problem is that there are many different earthquakes, all having the same peak acceleration a. Thus, when we say earthquake "a," we really define thereby a whole category of ground motions with different time histories, frequency content, etc. In the same way, when we say "shear wall" we actally have in mind a conceptual model of the specific physical wall in the plant. This conceptual wall is defined by certain specifications of thickness, height, materials, etc. This definition, however detailed, actually defines a whole category of walls, all meeting the specifications but differing from each other at levels beneath the level of definition.

Thus, when we ask if the wall will fail under quake "a," we are really envisioning a series of thought experiments in which we subject structures randomly selected from the category "shear wall," to quakes randomly selected from the category "a." We wish to know in what fraction, F, of these experiments did the wall fail. This fraction F, of course, will be a function of earthquake size a. Plotting this function, we obtain the fragility curve, Figure 0.17-3.

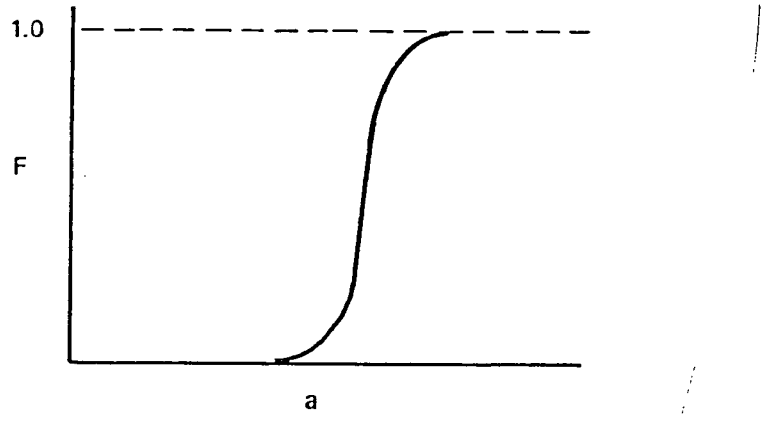


Figure 0.17-3. Fragility Curve for Typical Component

This curve characterizes the seismic performance of the wall. If we had run the above series of experiments, we would know this curve. Since we have not, we have uncertainty about what this curve is and again express our uncertainty in the form of a family of curves, Figure 0.17-4.

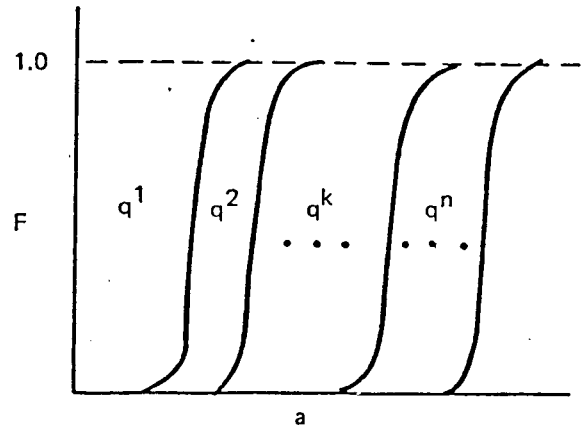


Figure 0.17-4. Family of Fragility Curves for a Typical Component C

To each such curve, $F_k(a)$, we assign a probability, q_k, and thus obtain the DPD:

$$\{<q_k, F_k>\} . \qquad (0.17\text{-}2)$$

Note that this is another instance of the Probability of Frequency format.

The family of fragility curves is developed by the structural dynamicists for each important structure and equipment item in the plant. This set of fragility families constitutes step 2 and is the structural dynamics input to the analysis. When this input is received, the items are sorted and listed in order of increasing strength. Thus, the most fragile elements--those most vulnerable to seismic action--come first on the list. This ordering is a key step which makes it possible to handle the plant logic step, step 3, in a computationally feasible form.

0.17.3.1 Plant Logic

If we have the fragility family for each structural and equipment component in the plant, we need next to aggregate these into fragility families for the plant as a whole. The meaning of this, more specifically, is as follows. Let y represent a certain plant damage state, one of the exit states from the plant matrix M. From the fault and event trees for the plant, we identify the set(s) of structural and equipment components whose failure in combination leads to state y. These combinations are stated in the form of Boolean expressions. For example, the expression

$$y = ① \wedge [④ \vee ⑧]$$

would say that state y occurs if component ① fails and, in addition, either component ④ or component ⑧ fails.

Next, by means of the logic of these combinations, the fragility families for these components can be aggregated into a single fragility family, which again can be written in the form of a DPD as follows:

$$\{< q_k^y, F_k^y >\} \quad . \tag{0.17-3}$$

This DPD, (0.17-3), represents the fragility family for the plant state y. The set of these DPDs for all plant states characterizes the seismic response of the plant.

0.17.3.2 Initial Assembly

It remains next to assemble the seismicity and fragility information into a statement of the frequency of core damage and plant states due to earthquake (step 4). To explain how this is done, consider first the case of no uncertainty. Thus, we have a single curve $\Phi(a)$ and a single fragility curve $F_y(a)$ for damage state y. In this case, we can say that the frequency with which state y occurs due to earthquake is:

$$\phi^y = - \int_0^\infty \left[\frac{d\Phi(a)}{da} \right] F^y(a) da. \tag{0.17-4}$$

To understand this integral, note that

$$- \left[\frac{d\Phi(a)}{da} \right] da \tag{0.17-5}$$

is the frequency with which quakes occur in the size range da about a, and $F_y(a)$ is the fraction of such quakes which result in plant state y.

0.17.3.3 Final Assembly (Step 5)

Once the numbers ϕ^y are obtained, giving the frequency of plant states as a result of seismic activity, the combination of these results with the containment and site matrices proceeds as for internal initial events, and leads to risk curves against release category and against the various damage indices. These curves express the seismic risk. They may then be combined with those from the other IEs to obtain the total risk.

To include uncertainty, we apply the basic idea of DPD arithmetic as follows.

Write

$$\phi_{jk}^y = \int_0^\infty \left[\frac{d\Phi_j(a)}{da} \right] F_k^y(a) da. \tag{0.17-6}$$

The probability that goes along with this value of frequency is the product of the probability of curve Φ_j and that of F_k:

$$p_{jk}^y = p_j \, q_k^y \quad . \tag{0.17-7}$$

Thus, we have the DPD for each ϕ^y

$$\{< p^y_{jk},\ \phi^y_{jk} >\} \ . \qquad (0.17\text{-}8)$$

These DPDs constitute a probabilistic statement of the row matrix for plant states

$$\phi^y = \phi^{S_M S}$$

resulting from the seismic events.

0.18 FIRE ANALYSIS METHODOLOGY

0.18.1 INTRODUCTION

The evaluation of the core melt frequency due to fires consists of six major steps:

1. Identification of critical areas where fires can cause an initiating event and, at the same time, fail redundant engineered safety functions.

2. Calculation of the probability distribution of the frequency of fires in these areas.

3. Calculation of fire growth and of effects of suppression activities.

4. Assessment of the effects of fires and inducing LOCAs and transients.

5. Assessment of the effects of fires on accident sequences originating from the initiating events identified in Step 4.

6. Calculation of the probability of frequency distribution of the plant states.

The occurrence of fires and their effects on plant safety are very complex issues which have not attracted the attention that other parts of risk assessment have in the literature. It is natural, therefore, that assumptions, usually conservative, have to be made for the analysis to be completed. The following remarks will place the results of this study into perspective:

1. The analysis is limited to rooms where the most damage can occur. A complete fire risk analysis would have to be much more detailed and would have to include fires that partially disable the engineered safety functions. Many more areas of the plant would have to be investigated. It is our judgment, however, that the results of this study would not change significantly.

2. The frequencies of fires are derived from evidence collected from all U.S. nuclear power generating stations. They are average frequencies and do not necessarily reflect the conditions that exist at the plant. For example, it is debatable whether some actual fire occurrences, e.g., the Browns Ferry fire, should be part of our data base. All these fires have been included in the data base.

3. A simple model is used for the propagation of fires in cable trays and the temperature rise in compartments due to the heat released by the fire.

4. Detailed analysis of the accident sequences is not done. Such an analysis would include explicitly the timing of events, the possibility of restoration of lost functions, the possibility of errors of commission by the operator, etc.

5. Whenever a fire is postulated in an area where it can affect instrumentation, the question of completeness of the analysis becomes very important. It is very difficult to know what information reaches the operators and how they respond. We have explicitly identified such situations and the frequency distributions that we have used reflect our concern about completeness.

0.18.2 IDENTIFICATION OF CRITICAL AREAS

A preliminary analysis is carried out to identify the locations that require a more detailed investigation. The first criterion for qualifying an area as a critical one is that a fire occurring in that location must be capable of causing an initiating event (a LOCA or a transient).

Given that a fire can cause an initiating event, we proceed to investigate whether the same fire can induce failures that will prevent:

1. Reaching and maintaining a condition of negative reactivity.

2. Removing decay heat.

3. Monitoring and controlling the primary system coolant inventory and pressure.

0.18.3 THE FREQUENCY OF FIRES

Reference 0-28 presents distributions for the average frequency of fires (per year) in the following areas: control room, cable spreading room, diesel room, containment, turbine building, and auxiliary building. These distributions are derived using Bayes' theorem.

The prior distributions are almost noninformative, i.e., no significant prior beliefs are injected into the analysis. The evidence is derived from actual fire incidents as reported to the American Nuclear Insurers (ANI) (a survey of ANI records is reported in Reference 0-29). The room-years (as of May 1, 1978) are calculated by surveying the FSARs of operating reactors.

These distributions are generic, i.e., they are the distributions of the frequency of fires averaged over all existing areas. For example, the distribution for the cable spreading room is a conservative distribution because it includes the Browns Ferry fire and does not give any credit to the changes in procedures and/or design that have resulted from that incident. We decided to be conservative and include the Browns Ferry fire even though we suspect that the frequency of fires in cable spreading rooms has been reduced. The derived distribution represents fires of any magnitude. The only criterion for their inclusion in the data base is that they have been reported to ANI, which has stricter reporting requirements than the NRC.

0.18.4 FIRE GROWTH

The fire growth is modeled using simple heat-transfer models (Reference 0-30). In some cases, we need to estimate the temperature rise within an electrical cabinet due to an external fire. The model computes the heat flux from the fire, the geometrical attenuation of the heat flux in passing from the fire to the cabinet, and the cabinet air temperature behavior given the external heat flux impinging on the cabinet surface. In this latter portion, both convective and radiative gains and losses are accounted for.

For the cable spreading rooms the situation is more complicated because we have to estimate the time it takes for cable trays of one and two ESF divisions to be within the flames (Figure 0.18-1).

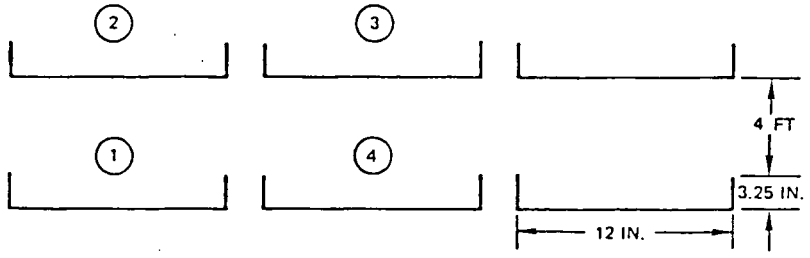


Figure 0.18-1. Cable Tray Configuration in a Cable Spreading Room

The physical parameters of interest, e.g., temperatures, heat fluxes, and burning rates, are deterministically calculated by the computer code COMPBRN (Reference 0-30).

The COMPBRN calculations proceed as follows. A small (12" x 12") established pilot fire is assumed at the top surface and in the center of tray 1, Figure 0.18-1. The mean times to propagate to trays 2 and 4 (τ_V*, and τ_H*, respectively) are computed for a number of trials. The purpose of this calculation is to establish a functional relationship between τ_V* and τ_H* and the input parameters that turn out to be important. In other words, response surfaces are generated (Reference 0-31). The advantage of this approach is that the uncertainty analysis can be done more economically using the response surface, rather than running the code many times.

These response surfaces are useful for the propagation of uncertainties due to our imperfect knowledge of the relevant parameters (e.g., burning rate of the fuel, heating value of the fuel, etc.). However, there is an additional source of uncertainty due to the basic model that COMPBRN implements. Combination of these two sources of uncertainty gives the uncertainty in the mean time τ_V* for two trays to be within the flames, Figure 0.18-2.

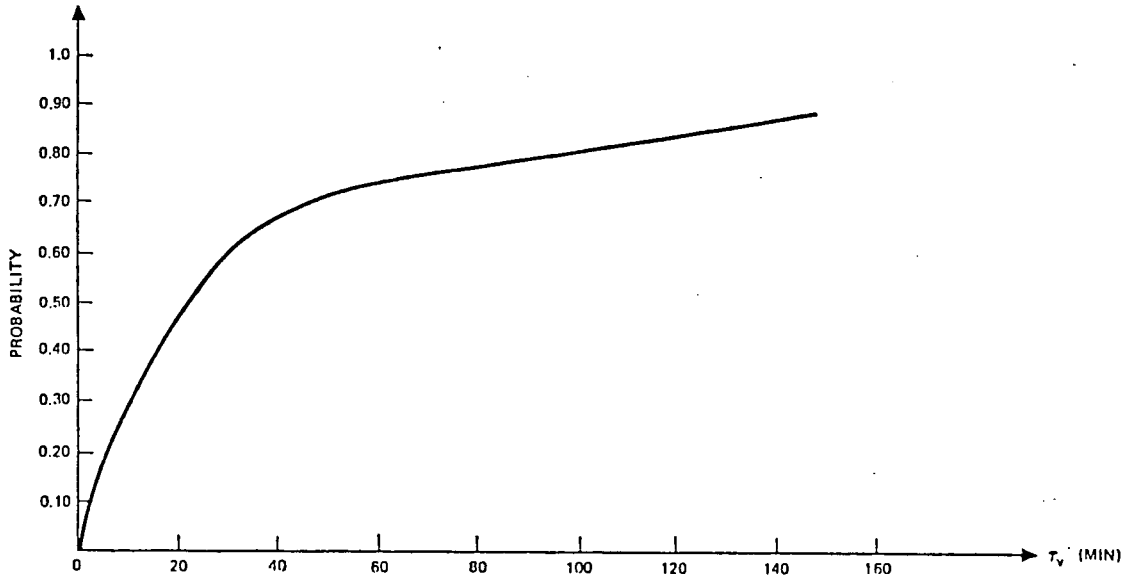


Figure 0.18-2. Uncertainty in τ_V* Due to Parameter Uncertainties (Cumulative Distribution Function)

0.18.5 FIRE SUPPRESSION

The distribution for τ_S, the mean time between fire ignition and suppression for cable-insulation fires, is assessed from information given in Reference 0-32. Fires occurring during plant construction have not been included.

The following discrete model for τ_S is used:

$$\Pr(\tau_S = 5 \text{ min}) = 0.40$$
$$\Pr(\tau_S = 15 \text{ min}) = 0.30$$
$$\Pr(\tau_S = 30 \text{ min}) = 0.20$$
$$\Pr(\tau_S = 60 \text{ min}) = 0.10$$

The justification for this distribution is as follows. The estimates reported in Reference 0-32 are the results of expert evaluation. There are 17 fires that are studied. The experts estimated that seven ($\sim 40\%$) of them were extinguished within 5 minutes. Four of the seventeen fires were estimated to have been extinguished between 5 and 25 minutes after initiation, and another group of four had extinguishment times between 30 and 35 minutes. Finally, there were two fires that were extinguished after 60 minutes (but before 85 minutes). Of course, these estimates are not statistical data, and we really do not know what the experts had in mind when they developed them. We judge that our discrete model is fairly consistent with the experts' estimates. Furthermore, it conforms with our belief that it is very likely that most of the fires in the cable spreading room will be extinguished fairly quickly by plant personnel.

We note that this suppression model is rather crude since it does not include any dependencies of the suppression time on the size of the fire. Furthermore, slowing down of the growth of the fire is not explicitly modeled. On the other hand, the data in Reference 0-32 are said to represent times to "control" the fires. This terminology remains unexplained and it may include slowing down effects.

0.18.6 COMBINED GROWTH AND SUPPRESSION

We can now derive the conditional frequency of two divisions being within the fire, given that a fire has started. This frequency is given by $\exp(-\tau_V^*/\tau_S)$, where τ_V^* has a probability distribution given in Figure 0.18-2 and τ_S is given in Section 0.18.5. Using DPD arithmetic, we get a histogram for the conditional frequency which, typically, has a range of several orders of magnitude reflecting the large uncertainties in our state of knowledge.

0.18.7 THE FREQUENCY OF FIRES INVOLVING TWO TRAYS

We can now derive the unconditional frequency of fires (per reactor year) that involve two adjacent trays, i.e., one division or two divisions. This is achieved by multiplying the histogram of Section 0.18.6 and the distribution of the fire frequency.

0.18.8 ACCIDENT SEQUENCES

The objective is to assess the potential of a fire in a critical location to cause an initiating event and defeat some (or all) of the engineered safety functions. For cable spreading rooms, a detailed analysis is carried out using the cable routings. In other cases, the concern is that readouts (flow meters, temperature, or pressure indicators) could drift or fail due to the temperature rise in the solid state signal processing circuitry.

In all the scenarios, the operator interpretation of the information that reaches the control room is important. The possibilities of errors of omission and commission are numerous. All of these are represented by the "other" category of scenarios.

0.18.9 UNCONDITIONAL FREQUENCIES

Having determined the conditional frequencies (i.e., given a fire) of core melt for each room, we make them unconditional by multiplying them with the frequency of two divisions being on fire (cable spreading rooms) or the frequency of fires in the particular location of interest. The frequencies for the rooms are added (using DPD arithmetic) to derive a histogram for the frequency of core damage.

0.18.10 PLANT STATES

The final step in the analysis is to determine the plant states. In addition to the dominant sequences identified in Section 0.18.8, an investigation of the potential effects of the fires in the critical

locations on the containment systems (e.g., fan coolers, spray systems) is performed. If the fire cannot affect these systems, their failure due to other causes is considered. The frequencies of failure of the containment functions are then combined with the unconditional frequencies of Section 0.18.9 to derive the final state-of-knowledge distributions concerning the risk from fires at the plant.

0.19 THE "OTHER" CATEGORY--COMPLETENESS AND LIMITATIONS OF RISK ANALYSIS, COMMON CAUSE EVENTS, AND SYSTEM INTERACTIONS

In Sections 0.3 and 0.5, it was pointed out that a risk analysis may be viewed, fundamentally, as a listing of scenarios or, more accurately, of categories of scenarios, each with its frequency and consequent damage. From this point of view, the question of completeness of the list is obviously a matter of much importance. Closely connected to the question of completeness are the much-talked-about subjects of "common cause," "systems interaction," and the "other" category introduced in Section 0.3.5. It is, therefore, of interest at this point to look back over the methodology outlined in this section, and forward to the application of the methodology in the report proper, in order to point out the places and ways in which completeness and related questions are dealt with.

0.19.1 COMPLETENESS

To begin with, note that certain categories of scenarios have been eliminated from the study as a matter of choice. Sabotage, war, terrorist attack, social chaos, etc., have, by conscious decision, been considered outside the scope of the present study.

With respect to identifying those scenarios which are within its scope, the study begins with the master logic diagram, Figure 0.6-2. This diagram and the reasoning process it represents are intended to be complete or exhaustive at each step. Thus, for example, since all significant radioactivity is held within the reactor core and, thus, within the plant containment system, the diagram states that there can be no significant release of radioactivity to the environment unless both the core melts and the containment system fails. The core cannot melt unless either the core power is above normal or the core cooling is lost. Cooling cannot be lost unless certain other things happen, and so on. If the master logic diagram is viewed as a pyramid, then each level of the pyramid may be thought of as a set of scenarios. The diagram has been designed so that, at each level, the set of scenarios is complete.

Proceeding down the pyramid then, the diagram comes to a set of categories of initiating events. These categories, by the way they are defined, are considered to be a complete set. That is, unless at least one of these initiating events occurs, the core cannot melt.

Given, then, that an initiating event does occur, the possible subsequent scenarios are delineated in one of the event trees for the reactor plant. Arrayed across the top of this tree as the "top events" are the various plant safety systems or safety functions that would come into play given such an initiating event. The possible scenarios, again actually categories of scenarios, are represented by the various paths through the event tree, the branching at each safety function representing the success or failure of that function.

The set of functions shown in the event tree is a complete set in the sense that, if all those functions succeed, the core cannot melt. The core can melt only if an appropriate combination of safety functions fail.

Thus, the set of paths in the plant event trees is also complete in the sense that any scenario which could possibly lead to core melt must be and is represented by, or included in, one of the paths in the plant event trees.

In the same way, an event tree-type analysis has been performed in this study for the containment system. Given a core melt, this analysis traces the course of the subsequent scenario through the containment. These scenarios are keyed to the occurrence or nonoccurrence of significant physical phenomena that relate to possible release mechanisms of the containment structure, i.e., design leakage, overpressure failure, missile penetration, basemat penetration, or containment bypass. This analysis is also complete in the sense that any possible scenario leading to release of radioactivity to the environment is represented by, or included in, one of the paths identified in the containment analysis.

Thus, at the level of the event trees the list of scenarios or scenario categories leading to radioactive release is complete. No "other" category need be added at this level.

Given, then, that the set of event tree level scenarios is complete, what remains is to assign the frequencies and damages to each of these scenarios. In the course of doing this, the question of completeness and "other" comes up again at more detailed levels. Let us review how this happens.

0.19.2 INITIATING EVENT FREQUENCIES

The first step in calculating the frequency of radioactive release scenarios is to determine the frequencies of the initiating events. As a basis for this determination, we now have considerable statistical evidence in the experience of the Zion plants themselves and in that of all the other commercial nuclear power plants. Thus, for those initiating events which are generated "internally" to the plant, we can treat the matter statistically using Bayes' theorem to evaluate the evidence. For "external" initiating events such as earthquakes of various sizes, the recorded statistical evidence on earthquake occurrence is used together with the most up-to-date seismological models of the area.

In all cases, uncertainty is explicitly included in the calculations. The result, therefore, is a set of probability curves against the frequency of each initiating event. These curves are complete in the sense that they express the current state of knowledge about the initiating event frequencies based on all the evidence, experience, and intelligence available at this time. New evidence will, of course, change these curves, but for now, they complete the discussion of initiating event frequencies.

0.19.3 FREQUENCIES OF THE PLANT DAMAGE STATES--COMPLETENESS AT THE SYSTEM LEVEL

With the initiating event frequencies established, the next step is to determine the conditional frequency of each path through the plant event tree leading to a plant damage state. Each such path or plant event tree scenario is itself composed of numerous lower level scenarios representing the various ways in which systems and combinations of systems can be unavailable. The question of completeness thus comes up again at this level.

Each system, as suggested in Figure 0.9-1, is divided into a set of "components" in such a way that, if all the components work, the system must work. A fault tree is then prepared for the system to identify the minimum failure sets or "cutsets" for the system. The set of these cutsets is again complete in the sense that the system cannot fail unless all the components in at least one of the minimum cutsets fail. Thus, with reference to Figure 0.9-4 we have, in establishing our list of scenarios, thus far maintained completeness down through the component level. The next level in further detailing the list consists of identifying the "causes" or combinations of causes which could lead to component and system failure.

0.19.4 COMPLETENESS AT THE CAUSE LEVEL--COMMON CAUSE AND SYSTEMS INTERACTION

Since our cutsets are complete, any cause that results in a system failure must cause all the components in at least one cutset to fail. Potential causes for component failure may be grouped into three classes:

1. Random (independent) equipment failures
2. Independent human failures
3. Dependent or common cause failures.

The first two of these are independent, individual failures. The frequencies of such independent failures are explicitly evaluated in the systems analyses and event tree calculations. The data used to estimate these independent failure frequencies include failures from all causes, thus effectively incorporating the "other" category. Although the frequencies of these independent failures can be significant, as a result of conservative and redundant design, system failure leading to a damaging scenario is very unlikely. Since the frequencies of system failure due to independent causes are very low, the dependent or common cause failures become the most important failure modes. These are discussed next.

Scenarios resulting from coincident independent failures have a very low frequency because the scenario frequency is the product of the frequencies of the independent failures, each of which is itself low. For common cause events, this is not the case. "Common cause events" are single events having the potential to fail more than one safety function and possibly cause an initiating event simultaneously. Such events, therefore, merit much attention in a risk analysis. In the present study, common causes have been treated as follows:

1. Common Support Equipment Failures. Common power supplies, cooling systems, signals, etc., are modeled explicitly in both the event and the fault tree diagrams. That is to say, the electric power system has multiple states so that in the event tree there are more than two branches (actually eight are used) at the electric power system. In the fault tree diagrams for each other system the state of electric power is used as a boundary condition in the top structure of the tree. Thus, in effect, a different subfault tree is calculated for each system assuming different states of electric power (availability of buses). From this subfault tree the split fractions are calculated for the appropriate branch point in the event tree.

2. Common Human Errors.

   a. Dependent human actions such as a single maintenance man making the same mistake on two redundant pumps, two operators influencing each other such that they both make the same mistake, or sequential testing by one operator who makes the same mistake in each test are modeled explicitly for situations in which operator action is expected. The NRC Human Reliability Handbook provided the basis for some of these calculations as explained in Sections 0.15 and 1.5.1.

   b. Misinterpretation of plant conditions can lead to multiple inappropriate actions. Such errors are somewhat less likely than single mistakes. Event tree and system models have been used in this study for such errors and have included the number of operators available and their likely interactions.

   c. Sequentially compounded errors can initiate a transient, may destroy the effectiveness of mitigating systems, and will add to the confusion of the operator. Some sequences of this type show up in our data and some recent administrative changes provide a good means for recovery from such situations. In particular, the shift technical advisor and the staff of the technical support center can provide somewhat independent oversight to determine what has happened and how recovery might proceed. Furthermore, emergency operating procedures have been revised since the TMI-2 accident to help the operator zero in on a minimum set of plant conditions to ensure the core remains cooled even if the details of what has happened are unclear. Nevertheless, such events can happen and are included in our event tree analyses by modeling specific systems interactions and test errors. Furthermore the most often encountered effect of sequentially compounded errors is operator confusion; therefore, broad uncertainty bands are assigned to operator ability to provide and maintain necessary plant functions. Such bands account for the wide variations in operator capability and situation complexity.

3. Common Abnormal Environment. In Section 7, we explicitly model the contributions due to:

   a. Earthquake
   b. Fire
   c. Flood (Internal and External)
   d. Tornado and Tornado Missiles
   e. Aircraft Accidents
   f. Transportation of Hazardous Materials
   g. Turbine Missiles.

   Here, the likelihood of the entire range of each environment (not just the "qualified" range) is determined along with the susceptibility or "fragility" of equipment to those environments.

   Other abnormal environments with common cause potential (such as dirt, grit, freezing, humidity, etc.) were considered in each system analysis as appropriate. The ideas in References O-26 and O-27 were relevant in this context.

0.19.5 "OTHER" SCENARIOS

As can be seen from the previous sections, a systematic and deliberate effort has been made to identify and explicitly model all significant scenarios. Notwithstanding that, what other scenarios might there be which have not been explicitly included in this study?

1. There are abnormal environments other than those listed above. Some of these are known, e.g., volcanoes and meteor strikes, and some may be unknown and unimagined.

2. Common Design, Manufacturing, or Installation Defects. These could lead to common cause failures under normal environments (e.g., the Kahl failure). QA/QC programs help eliminate such problems. For plants like Zion that have operated for several years, such failures become extremely unlikely because the equipment has been operated, frequently tested, and changed by maintenance and replacement.

3. There is some possibility that common equipment defects (design, manufacturing, or installation) were not corrected by QA/QC procedures and could lead to common cause failure under the abnormal environmental conditions considered.

4. Common Wearout. Many factors contribute to wearout--abrasion, fatigue, corrosion, etc. If similar equipment sees similar operation, environment, etc., all items may fail due to wearout at nearly the same time. However, the chance that they all wear out within, say, one week following a specific initiating event, is vanishingly small when compared with other contributions to failure.

5. Plant Configuration. Changes in plant configuration can lead to excessive stress in operating components. Such failures were usually bounded in this analysis by assuming system failure when

equipment was required to operate outside of its design conditions. Initiating events caused by such failures are embedded in the initiating event data.

6. Miscellaneous, unimagined, oddball, human errors, or combinations of human error and equipment failure not otherwise included in the scenario list.

The "other" category of failure causes that was included in the event and fault trees is the study team's attempt to allow for these scenarios. The assigned frequencies represent our state of knowledge concerning the likelihood of these "odd" occurrences. Although we recognize that others may assign different frequencies to the "other" categories, we believe that our approach provides a rational framework in which we can discuss potential differences.

In addition to the "other" category we have acknowledged the limited amounts of information that are available in certain areas as well as our difficulties in assigning very low frequencies by broadening the generic distributions for failure and human error rates that are available in the literature. Moreover, the uncertainty bounds for fragility, seismicity, effects of fires on cables, etc., were quite broad.

Based on these considerations, the study team feels that the uncertainty bounds and the final curves are sufficient to cover possible contributions from the above "other" category.

0.19.6 COMPLETENESS OF THE CONTAINMENT SCENARIOS

The containment analysis is complete at the level of categorizing release mechanisms. All phenomena that have been proposed in the literature as well as many not previously considered have been examined and included in the analysis. It is possible that some "other" phenomena leading to these mechanisms exist but have not been imagined and accounted for in this study. It is the feeling of the study group that such phenomena could not have a more severe effect than those which have been included. Moreover, the state of knowledge of containment phenomena is sufficiently high in the group's judgment that the frequency of any such unimagined phenomena is assigned a negligible value compared to the scenarios explicitly included in the analysis.

0.19.7 COMPLETENESS OF THE SITE ANALYSIS

The portion of the scenarios which relates to events subsequent to release of radioactivity has been modeled by the computer program CRACIT. This model includes numerous (effectively, about 9,000) samples of actual weather scenarios in the Zion area.

From this work, an understanding has evolved of the conditions leading to most damage (i.e., movement of the plume to a populated area followed by rain at the worst possible time and inadequate sheltering or evacuation). The frequency of such conditions is uncertain, of course, and is represented by the uncertainty in the S matrix. The set of conditions as such, however, is felt to be effectively complete so that an explicit category "other" need not be added.

0.20 REFERENCES

0-1. Webster's New Collegiate Dictionary, G&C Merriam Company, 1977.

0-2. U.S. Nuclear Regulatory Commission, "Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants," WASH-1400 (NUREG-75/014), October 1975.

0-3. de Finetti, B., Theory of Probability, Vols. 1 and 2, John Wiley & Sons, N. Y., 1974.

0-4. Savage, L. J., The Fountain of Statistics, 2nd Ed., Dover Publications, N. Y., 1972.

0-5. Lindley, D. V., Introduction to Probability and Statistics from a Bayesian Viewpoint, Cambridge University Press, 1970.

0-6. Apostolakis, G., "Probability and Risk Assessment: The Subjectivistic Viewpoint and Some Suggestions," Nuclear Safety, 19(3): 305-315, May-June 1978.

0-7. Lewis, H. W., et al, "Risk Assessment Review Group Report to the U.S. Nuclear Regulatory Commission," NUREG/CR-0400, September 1978.

0-8. Kaplan, S., "A Matrix Theory Formalism for Event Trees Analysis," to appear in Risk Analysis, Vol. 1, 1981.

0-9. Kaplan, S., "On the Method of Discrete Probability Distributions in Risk and Reliability Calculation," to appear in Risk Analysis, Vol. 1, 1981.

0-10. Winkler, R. L., and W. L. Hays, Statistics: Probability, Inference and Decision, 2nd Ed., Holt, Rinehart & Winston, N. Y., 1975.

0-11. Green, A. E., and A. J. Bourne, Reliability Technology, Wiley Interscience, N. Y., 1972.

0-12. Hammersley and Handscomb, Monte Carlo Methods, Methuen, London, 1964.

0-13. Apostolakis, George, and Yum Tong Lee, "Methods for the Estimation of Confidence Bounds for the Top-Event Unavailability of Fault Trees," Nuclear Engineering and Design, 41, 1977, pp. 411-419.

0-14. Apostolakis, G., S. Kaplan, B. J. Garrick, and R. J. Duphily, "Data Specialization for Plant-Specific Risk Studies," Nuclear Engineering and Design, 56:321-329, 1980.

0-15. Kaplan, S., B. J. Garrick, and P. Bieniarz, "On the Use of Bayes' Theorem in Assessing the Frequency of Anticipated Transients," to appear in Nuclear Engineering and Design, 1981.

O-16. "IEEE Guide to the Collection and Presentation of Electrical, Electronic, and Sensing Component Reliability Data for Nuclear Power Generation Stations," IEEE STD-500, 1977.

O-17. Lichtenstein, S., B. Fischhoff, and L. D. Phillips, "Calibration of Probabilities: The State of the Art," in: J. Jungermann and G. de Zeeuw, Editors, Decision Making and Change in Human Affairs, D. Reidel Publishing Co., Dordrecht, Holland, 1977.

O-18. Slovic, P., B. Fischhoff, and S. Lichtenstein, "Facts Versus Fears: Understanding Perceived Risk," in Societal Risk Assessment, R. C. Schwing and W. A. Albers, Jr., Editors, Plenum Press, 1980.

O-19. Peterson, C., and A. Miller, "Mode, Median, and Mean as Optimal Strategies," J. Exp. Psych., 68:363-367, 1964.

O-20. Sullivan, W. H., and J. P. Poloski, "Data Summaries of Licensee Event Reports of Pumps at U.S. Commercial Nuclear Power Plants," January 1, 1972, to April 30, 1978, NUREG/CR-1205, EGG-EA-5044, January 1980.

O-21. Hubble, W. H., and C. F. Miller, "Data Summaries of Licensee Event Reports of Valves at U.S. Commercial Nuclear Power Plants," January 1, 1976, to December 31, 1978. NUREG/CR-1363, EGG-EA-5125, June 1980.

O-22. Poloski, J. P., and W. H. Sullivan, "Data Summaries of Licensee Event Reports of Diesel Generators at U.S. Commercial Nuclear Power Plants," January 1, 1976, to December 31, 1978, NUREG/CR-1362, EGG-EA-5092, March 1980.

O-23. Kaplan, S., "On a 'Two Stage' Bayesian Procedure For Determining Failure Rates From Experiential Data," to appear in IEEE Transactions on Power Apparatus and Systems, 1981.

O-24. Swain, A. D., and A. G. Guttmann, "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications," Draft Report, NUREG/CR-1278, October 1980.

O-25. Apostolakis, G., and S. Kaplan, "Pitfalls in Risk Calculations," Reliability Engineering, Vol. 2, 1981.

O-26. Rasmuson, D. M., N. H. Marshall, J. R. Wilson, and G. R. Burdick, "COMCAN II -- A Computer Program for Common Cause Failure Analysis," EG&G Idaho, Inc., Report TREE-1289, 1978.

O-27. Fleming, K. N., and P. H. Raabe, "A Comparison of Three Methods for the Quantitative Analysis of Common Cause Failures," Transactions, ANS Topical Meeting on Probabilistic Analysis of Nuclear Reactor Safety, Los Angeles, May 8-10, 1978.

0-28. Apostolakis, G., and M. Kazarians, "The Frequency of Fires in Light Water Reactor Compartments," paper presented at the ANS/ENS Topical Meeting on Thermal Reactor Safety, Knoxville, Tennessee, April 6-9, 1980.

0-29. Sideris, A. G., R. W. Hockenbury, M. L. Yeater, and W. E. Vesely, "Nuclear Plant Fire Incident Data File," Nuclear Safety, 20, pp. 308-317, 1979.

0-30. Siu, N. O., "Probabilistic Models of Compartment Fires," M. S. Thesis, University of California, Los Angeles, 1980.

0-31. Iman, R. L., J. C. Helton, and J. E. Campbell, "Risk Methodology for Geologic Disposal of Radioactive Waste: Sensitivity Analysis Techniques," NUREG/CR-0394, 1978.

0-32. Fleming, K. N., W. J. Houghton, and F. P. Scaletta, "A Methodology for Risk Assessment of Major Fires and Its Application to an HTGR Plant," GA-A15402, July 1979.

0-33. Aitchison, J., and J. A. C. Brown, The Lognormal Distribution, Cambridge University Press, 1963.

0-34. Hall, R. E., et al, "A Risk Assessment of a Pressurized Water Reactor for Class VII-VIII Accidents," NUREG-CR/0603, October 1979.

0-35. Kaplan, S., and B. J. Garrick, "On the Quantitative Assessment of Risk," Risk Analysis, Vol. 1, No. 1, March 1981.