Nuclear Safety Design Principles & the Concept of Independence: Insights from Nuclear Weapon Safety for Other High-Consequence Applications

Jeffrey D. Brewer*
Sandia National Laboratories, Albuquerque, NM, USA

Abstract: Insights developed within the U.S. nuclear weapon system safety community may benefit system safety design, assessment, and management activities in other high consequence domains. The approach of *assured nuclear weapon safety* has been developed that uses the Nuclear Safety Design Principles (NSDPs) of *incompatibility*, *isolation*, and *inoperability* to design safety features, organized into subsystems such that each subsystem contributes to safe system responses in *independent* and *predictable* ways given a wide range of environmental contexts. The central aim of the approach is to provide a robust technical basis for asserting that a system can meet quantitative safety requirements in the widest context of possible adverse or accident environments, while using the most concise arrangement of safety design features and the fewest number of specific adverse or accident environment assumptions. Rigor in understanding and applying the concept of independence is crucial for the success of the approach. This paper provides a basic description of the *assured nuclear weapon safety* approach, in a manner that illustrates potential application to other domains. There is also a strong emphasis on describing the process for developing a defensible technical basis for the independence assertions between integrated safety subsystems.

Keywords: System Safety Design, Safety Assessment, Independence, Nuclear Weapon Safety.

1. INTRODUCTION

Insights developed within the U.S. nuclear weapon system safety community may benefit system safety design, assessment, and management activities in other high consequence domains. The approach of *assured nuclear weapon safety* has been developed that uses the Nuclear Safety Design Principles (NSDPs) of *incompatibility*, *isolation*, and *inoperability* to design safety features, organized into subsystems such that each subsystem contributes to safe system responses in *independent* and *predictable* ways given a wide range of environmental contexts. The *assured nuclear weapon safety* approach strives toward use of a concise arrangement of safety design features and a limited number of specific adverse or accident environment assumptions. Simplicity of safety features, passive safe responses, and a systematic allocation of basic features among engineered features and human actions[1] are emphasized in the implementation, and an innovative *inside out* process for hazard identification, which is described in this paper, is also applied throughout iterative system design phases to support the process of NSDP integration. In essence, this approach claims to be an efficient method for engineering bounded system safety-related responses.

Appropriate independence assertions are essential given that multiple safety subsystems, each providing safe responses for all relevant environments in independent and predictable ways must be integrated into the system to form a basis for meeting stringent qualitative and quantitative safety requirements. Overreliance on the concept of independence for asserting levels of safety without providing a sufficient technical basis is tempting, and must be avoided. In addition, it is recognized that humans do a poor job both of conceiving the many ways things may fail and estimating

* Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. This paper is designated as SAND2014-1832C at Sandia National Laboratories. The author may be contacted at: jdbrewe@sandia.gov.

[1] The primary human actions to consider are those designed to provide an unambiguous indication of intent to achieve a nuclear detonation. Other human actions include those related to ensuring safety during weapon production, assembly, testing, transportation, maintenance, and disassembly.

probabilities for conceivable failures [1-3]. Particularly in the domain of low probabilities and high stakes it is helpful to move beyond the common 'model' and 'parameter' uncertainty approach and consider opportunities for flaws across the spectrum of theories, models, and calculation techniques [3]. For high-consequence systems it can be argued that the approach taken ought to be more *possibilistic* than *probabilistic* [2]. Therefore, the foundations for any (hopefully few) probabilistic statements regarding safety must be clear and defensible.

Particular emphasis in this paper is placed on implementing the NSDPs using independence assumptions founded upon the distinctions of functional, temporal, and physical differences. Although the concepts of function, time, and physical properties do not provide mutually exclusive categories for describing dependence, they are proposed as helpful concepts when striving to increase the independence of safe responses to ensure the control of hazards in high-consequence applications. The motivation for using all three of these conceptual distinctions is increased in situations where data are sparse or non-existent for demonstrating that a focus on only one or two is sufficient. The paucity of data is typical, thus far (fortunately), for extremely high-consequence events where observing data may involve observing massive disasters.

For example, no inadvertent nuclear weapon detonations have been observed. This is due in part to the care with which such systems are handled, but the weapons must still meet stringent requirements if an accident were to occur—and there are many ways in which credible accidents may occur. As another example, no pandemic illness resulting from an accidental release of a biological organism from a research facility has been observed. No doubt this has been aided by the lack of man-made or natural disasters in the vicinity of such facilities, but the facilities must still be able to provide a safe response to these externally imposed conditions if such an event were to occur.

This paper provides a basic description of the *assured nuclear weapon safety* approach, incorporating the NSDPs, and describes the process for developing a defensible technical basis for the independence assertions between integrated safety subsystems. The contribution of functional, temporal, and physical differences to achieving independence are described. Clear and distinct definitions for *common-cause failure* and *common-mode failure* (with examples) are provided. Attributes of *predictability* that strengthen the technical basis for safe system responses are given. It is hoped that insights presented here, developed within the U.S. nuclear weapon system safety community over the course of more than 50 years, may benefit system safety design, assessment, and management activities in other high consequence domains.

2. NUCLEAR WEAPONS

To elucidate the concepts in this paper, it is beneficial to: (1) describe some specific notion of a system, (2) identify the hazards to safety associated with the system, and (3) develop a grouping strategy for sources of energy that can contribute to the release of the hazards. In this discussion the system of interest is a nuclear weapon. A nuclear weapon is a device in which the explosion results from the energy released by reactions involving fission or fusion (of atomic nuclei) [4]. It is not unusual for the total energy release of a nuclear weapon, given that it is detonated in the intended mode, to be expressed in the range of hundreds of kilotons or even in the vicinity of a megaton of trinitrotoluene (TNT) equivalent, i.e., the energy equal to exploding two billion pounds of TNT. The greatest hazard[2] to safety that a nuclear weapon can pose is an inadvertent nuclear detonation[3] (IND), especially an IND that approaches the designed yield for the weapon—this may well be the archetype of a high consequence safety hazard.

---

[2] In addition to a nuclear detonation, hazards to safety include high explosive detonation/deflagration as this will disperse special nuclear material. There are certainly other hazards to consider with the materials inside nuclear weapons (e.g., fire, toxic chemicals, and various radionuclide concerns); however, nuclear detonation and special nuclear material dispersal are the generally considered the greatest hazards.

[3] Nuclear Detonation – An energy release through a nuclear process, during a period of time on the order of 1 microsecond, in an amount equivalent to the energy released by detonating 4 or more pounds of TNT [5].

The simplified conceptual nuclear weapon design of interest here is that of a sealed pit, implosion type nuclear weapon. In this design, a mass of fissile material (the pit) is surrounded by high explosives (HE) and a detonation system is also included that initiates the HE. When the HE detonation system is operated, the explosives compress the fissile material until a nuclear detonation results. Figure 1, A notionally represents the weapon configuration before HE detonation and Figure 1, B notionally represents the configuration immediately after HE detonation, just prior to the nuclear explosion.

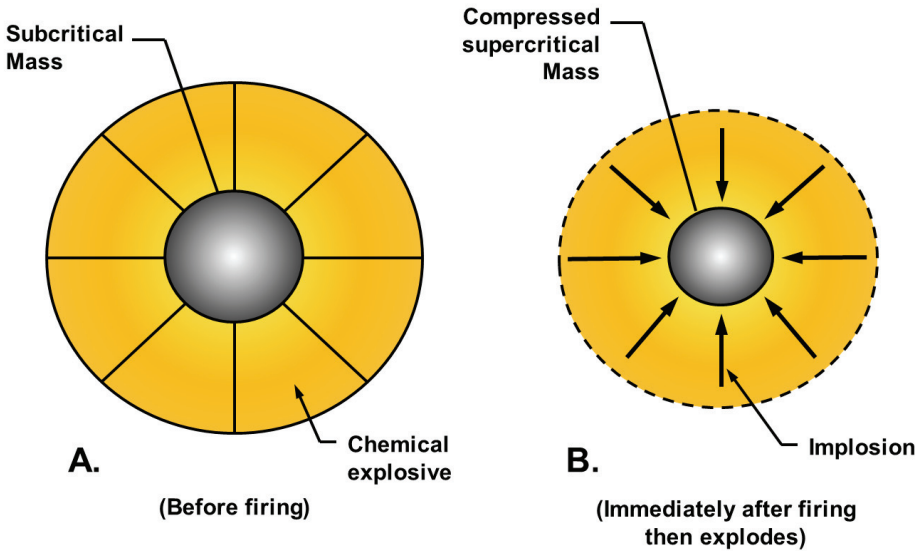

Figure 1: Illustration of sealed pit, implosion assembled weapon.

In the context of a nuclear weapon, a helpful grouping strategy for sources of energy that can contribute to the release of the hazards to safety (i.e., IND or dispersal of special nuclear material) is that of front, side, and back door energy scenarios. These scenario groups are associated with distinct energetic materials and/or distinct collections of energy storage/production devices that are essential for producing a nuclear detonation or dispersal of nuclear material when the weapon is fully assembled or partially assembled either due to adverse or accident environment exposure or due to assembly or dismantlement actions. It is important to note that this grouping strategy for the nuclear weapon example is focused on distinguishing the *energy sources* associated with release of the hazard(s). It is not focused on distinguishing all of the *specific pathways* through which energy may travel. The specific energy pathways are considered, as needed, during the safety design/analysis process relative to the level of assembly of the weapon and a decomposition of types of adverse or accident environments. The main benefit of this approach can be described in simple terms: "if the *energy sources* respond safely in an adverse or accident environment, then consideration of the possible *energy pathways* is not required."

Here are descriptions of the front, side, and back door energy scenarios for the nuclear weapon example: *Front Door Energy Scenarios* involve operating the HE detonation system using the intended energy storage/production devices designed to operate the HE detonation system or another internal energy storage/production device that is compatible with operating the HE detonation system. For example, the intended energy storage/production devices for operating the HE detonation system are certainly compatible with operating it and achieving a nuclear detonation. However, another energy device may also be able to provide an energy input to the HE detonation system along an unintended energy pathway.

Side Door Energy Scenarios involve operating the HE detonation system in any way that does not involve the intended energy storage/production devices. For example, a lightning strike to a damaged weapon may provide sufficient energy to operate the HE detonation system. *Back Door Energy Scenarios* involve direct initiation of the high explosive material required to achieve a nuclear

detonation; the HE detonation system is either irrelevant or simply plays a secondary role in achieving a nuclear detonation. For example, a shock environment or thermal environment that causes initiation of the explosives would be a back door energy scenario. Mixed-door scenarios are also possible such as a case in which a lightning strike event is able to provide the energy needed to energize intended energy storage/production devices and operate the HE detonation system. As soon as HE is assembled with the fissile material then back door energy scenarios are possible. When the HE detonation system is included then both back door and side door energy scenarios are possible. Finally, once the intended energy storage/production devices are installed then back, side, and front door energy scenarios are possible (see Figure 2).

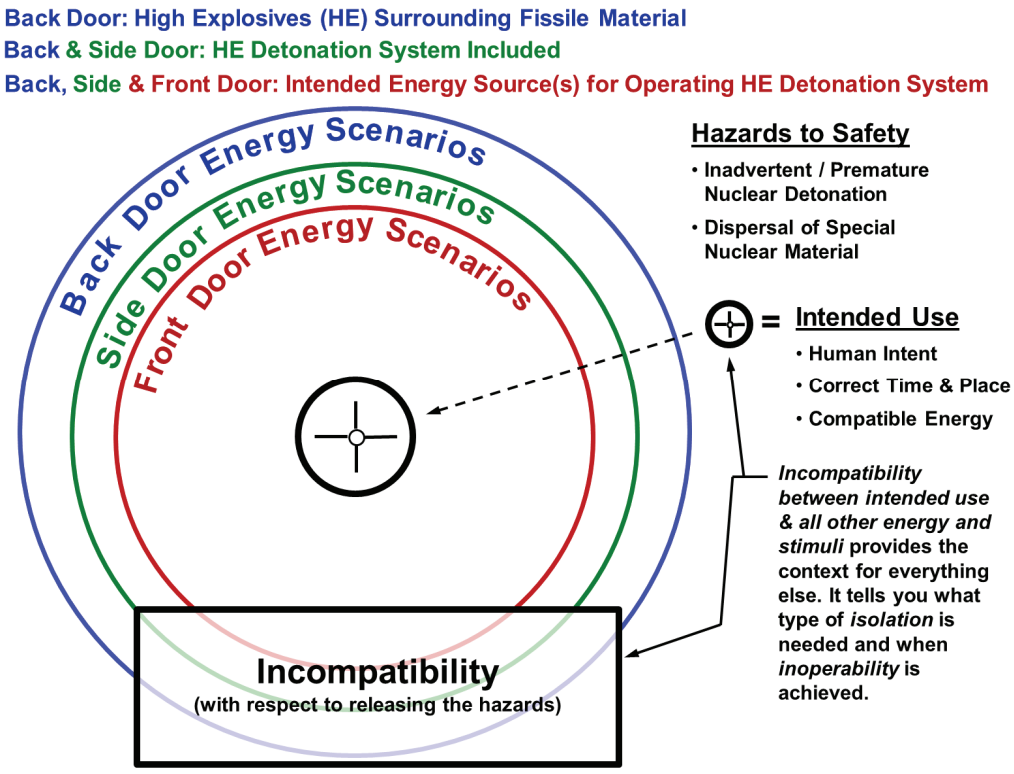

Figure 2: Illustration of energy scenarios for nuclear weapon example.

The front, side, and back door energy scenario taxonomy for the nuclear weapon example was inspired by the *inside out* (IO) approach to identifying hazards to safety. The inside out approach involves: (1) identifying the simplest configuration of system elements that present a safety hazard of concern, (2) evaluate all hazards to safety and the possible energy types, energy sources, energy pathways, and information that may facilitate unsafe consequences, (3) add the next design feature to the simplest configuration, (4) evaluate hazards, energy, and information as in step 2, and (5) repeat the process until the full system design configuration has been analyzed. The inside out (IO) approach, when repeated at various times in the system design process, is an excellent means of identifying hazards and incorporating an arguably complete set of features to control the hazards throughout the lifecycle of the system. The IO approach proves beneficial for deciding how to manage hazards during the production process of a weapon system, as well as assessing levels of safety when the weapon system is fully assembled or partially assembled either due to adverse or accident environment exposure or due to assembly or dismantlement actions. The same approach can be used when managing the radionuclide hazards of a nuclear power reactor or biological hazards of a research facility.

Nuclear weapons are indeed ‘special’ due to their potentially devastating consequences in terms of loss of life, injury, and damage to property and the environment. Therefore, highly stringent safety

requirements, including *stringent numerical inadvertent nuclear detonation safety requirements* are appropriate. U.S. Department of Defense (DoD) Directive 3150.02 [6] and U.S. Department of Energy (DOE) Order 452.1D [5] both reflect this perspective. The following nuclear detonation safety design requirements have been established by the U.S. DOE for nuclear weapons delivered to the U.S. DoD (DOE O 452.1D section 4.f.(3)(a)) [5]:

1. Normal Environment[4]. Prior to receipt of the enabling input signals and the arming signal, the probability of a premature nuclear detonation must not exceed one in a billion (1E-09) per nuclear weapon lifetime.
2. Abnormal Environment[5]. Prior to receipt of the enabling input signals, the probability of a premature nuclear detonation must not exceed one in a million (1E-06) per credible nuclear weapon accident or exposure to abnormal environments.
3. One-Point Safety. The probability of achieving a nuclear yield greater than 4 pounds of TNT equivalent in the event of a one-point initiation of the weapon's high explosive must not exceed one in a million (1E-06).

It is noted that the normal environment requirement is for the lifetime of a weapon, and the abnormal environment requirement is given an accident or exposure, and that both requirements are for a single weapon.

The only way to confidently assert that such stringent numerical nuclear detonation safety requirements have been met is to use a decomposition approach in which multiple, independent elements work together in the system to provide the desired level of safety. DoD 3150.2-M [4], U. S. Navy and U. S. Air Force nuclear safety requirements documents (e.g., AFMAN 91-118 [7]), DOE O 452.D [5], and requirements used at the U. S. nuclear weapon laboratories all reflect this perspective.

3. ASSURED NUCLEAR WEAPON SAFETY APPROACH

Sandia National Laboratories (SNL) has established a nuclear safety design philosophy, called assured nuclear weapon (NW) safety, that is used to ensure that the quantitative safety requirements for U. S. nuclear weapons are met or exceeded. This section provides the author's 'summary perspective' of several key aspects of that philosophy. The foundation for the safety design philosophy consists of the three irreducible nuclear safety design principles (NSDPs) of incompatibility, isolation, and inoperability. *Incompatibility* is the use of energy or information that will not be duplicated inadvertently. *Isolation* is the predictable separation of weapon elements from compatible energy. *Inoperability* is the predictable inability of weapon elements to function. Incompatibility is the dominant NSDP as it provides the context for both isolation and inoperability features. That is, unless you know what energy and stimuli are compatible with operation of all or part of the nuclear weapon, you will not know what to isolate, nor will you know the point at which inoperability is achieved. The definition of predictability is 'certain[6] to happen based on knowledge and experience.' To be "predictable," knowledge and experience should be based on attributes that are identifiable, analyzable, testable, controllable, and verifiable[7].

---

[4] Normal Environment – In DOE operations, the environment in which nuclear explosive operations and associated activities are expected to be performed. In DoD operations, the expected logistical and operational environments, as defined in a weapon's stockpile-to-target sequence and military characteristics, that the weapon is required to survive without degradation in operational reliability [5].

[5] Abnormal Environment – In DOE operations, an environment that is not expected to occur during nuclear explosive operations and associated activities. In DoD operations, as defined in a weapon's stockpile-to-target sequence and military characteristics, those environments in which the weapon is not expected to retain full operational reliability [5].

[6] In practical applications 'certain to happen' is often implicitly understood to mean that a safe response is known to occur with a high degree of confidence.

[7] *Identifiable* – capable of being distinguished and named, specifically, the safety-critical characteristics and parameters of the safety critical components and other elements. *Analyzable* – amenable to available analysis techniques. *Testable* – can be demonstrated by measurement from experimentation and test activities.

3.1. Four Step Process of Implementing the Nuclear Safety Design Principles

The four step process of implementing the NSDPs is as follows[8]:

1. Develop a nuclear weapon that is **incompatible** with all forms and levels of energy except the correct sequence of intended, authorized, and unambiguous energy and stimuli.

   Begin with examination of incompatibility with respect to back door energy scenarios from the point at which the HE is assembled with the fissile material. Then examine incompatibility with respect to back and side door energy scenarios from the point that the HE detonation system is included. Then examine back, side, and front door energy scenarios from the point that energy storage/production devices capable of operating the HE detonation system are added.

2. For any part of the nuclear weapon that is compatible with unintended energy or stimuli, provide **isolation** from that energy or stimulus that could lead to an accidental explosion of any kind, and/or provide a reversible **inoperability** feature to eliminate or minimize exposure to safety hazards. The inoperability feature must be incompatible with all forms and levels of unintended energy and stimuli.

3. For any isolation feature that also blocks intended HE detonation system energy, provide a reversible **isolation** feature (a.k.a., a *stronglink*) to allow only intended energies to propagate to the HE detonation system. The stronglink must be **incompatible** with all forms and levels of energy and stimuli except the correct sequence of intended, authorized, and unambiguous energy and stimuli.

4. For any of these stronglinks, isolation features, reversible inoperability features, or incompatibilities that are subject to failure, provide an irreversible inoperability feature (a.k.a., a *weaklink*) that passively renders the nuclear weapon **inoperable** and incapable of producing an accidental explosion of any kind before such failure.

   For example, prior to melting through an energy isolation barrier, the material within a weaklink changes its physical state due to the rise in the temperature and renders the system irreversibly inoperable.

For a specific system, a high-level, concise expression of what will be isolated, inoperable and/or incompatible to provide assured safety is generated and captured in a *safety theme*. A *safety theme implementation* is a detailed explanation of how independent safety subsystems and associated safety design features are integrated into a system to provide assured safety [2]. The safety design features should provide their safety function in an *inherently safe*[9] or *passively safe*[10] manner. *Active safety*

Controllable – can be produced in a repeatable fashion. *Verifiable* – critical parameters can be shown to have met their requirements.

[8] This version of the process was adapted from an elegant articulation of it made by a Principal Member of the Technical Staff at SNL, Robert G. Hillaire, Ph.D., PE — provided via personal communication to the author.

[9] An *inherent safety feature* achieves its safety function by the elimination of a specific hazard by means of the choice of material and design concept. Therefore, no changes of any kind, such as internally or externally caused changes of physical configuration can possibly lead to an unsafe condition. An inherent safety feature represents conclusive or deterministic safety, not probabilistic safety [adapted from ref. 8].

[10] A *passive safety design feature* provides a safety function that does not depend on external mechanical and/or electrical power, signals, or forces to operate. That is, passive safety design features provide their safety function without having to sense/detect an undesirable environment and then actuate to a safe state. Passive safety design features rely on natural laws and properties of materials. It is important to note that passive safety design features, while not relying on human action or other power, can fail due to mechanical or structural failure or willful human interference. Thus, passive safety is not synonymous with inherent safety or absolute reliability [adapted from ref. 8]. One example of a passive safety design feature would be driver restraints in a

design features[11] should only be used when inherent or passively safe features cannot be used. In other words, the safety theme implementation should be designed to rely mainly upon *first principles of physics and chemistry*. These principles are the fundamental characteristics inherent in physics and chemistry that provide a predictable response when the subsystem in question is subjected to specified stimuli [2]. Nuclear weapons tend to be amenable to the incorporation of inherent or passive safety features given that they are "single-use" devices which are designed to spend most of their existence in a passive or dormant state, i.e., they are unpowered and stored in secured areas[12]. A largely passive system such as a nuclear weapon may be contrasted with an active system such as a commercial passenger aircraft that exercises its repeatable engineered functions (e.g., takeoff, safe flight, landings) many times. An active system will necessarily incorporate more active safety design features during its operational life.

Figure 3 provides a graphical illustration of how the NSDPs may be applied in a safety theme for controlling the hazards associated with the notional nuclear weapon. Some features may provide isolation or inoperability only for particular energy scenarios. Others may be built into the weapon in a safe state and must be actively transitioned to an unsafe state to allow intended use. Other features such as weaklinks (WL) may be built into the weapon in an unsafe/operable state, but they will irreversibly transition, in a passive manner, to a safe/inoperable state if certain adverse or accident environments occur. Isolation features such as stronglinks (SL) may be able to transition between safe and unsafe states upon receipt of unambiguous information resulting from human actions.

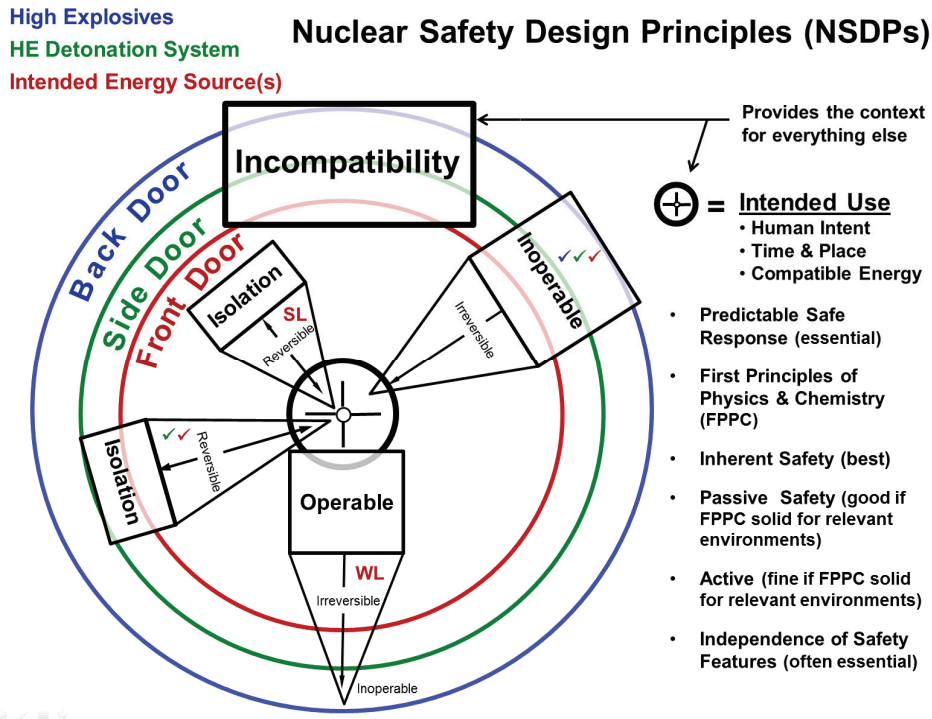

Figure 3: Illustration of NSDP application to energy scenarios for weapon example.

---

race car provide their safety function without having to change state in an accident. Another example would be an energy storage/production device in a weapon that becomes irreversibly inoperable due to a physical material state change when the weapon is subjected to a thermal environment such as in an aircraft fuel fire.

[11] An *active safety design feature* is anything that does not provide its safety function in an inherent or passive manner. That is, an *active safety design feature* must successfully sense/detect an undesirable environment **and** then actuate to a safe state (e.g., an airbag deployed in a motor vehicle crash); **or** it must continue to **act** in a safe manner (e.g., the functioning engines on an aircraft maintaining safe flight).

[12] Ensuring the safety of nuclear weapons includes designing systems which strive to meet stringent IND requirements given AEs while simultaneously striving to prevent weapons from exposure to AEs [6].

3.2. Levels of System Safety & Assessment of the 'Level' for a Specific Weapon Design

In the practice of implementing the SNL safety design philosophy, there are three levels[13] of safety that can be identified: (1) the Safety Design Ideal, (2) SNL Assured Safety, and (3) SNL Asserted (Implemented) Safety. The three levels are described below:

1. **Safety Design Ideal —** Isolation of compatible energy from detonation-critical components of an operable weapon until after the weapon becomes irreversibly inoperable is assured by first principles of physics and chemistry for all levels of relevant environments.

2. **SNL Assured Safety —** Isolation of compatible energy from detonation-critical components of an operable weapon until after the weapon becomes irreversibly inoperable to the levels required in a sufficiently predictable manner. Adequate requirements and technical bases exist to support compliance with system level requirements. This is the highest possible level of system safety that can be implemented given currently available knowledge, experience, and technology. This approach to engineering system safety changes over time with increases in knowledge, experience, and available technology.

3. **SNL Asserted (Implemented) Safety —** this is the same as assured safety in that system level requirements are asserted to be met, but with specific exceptions noted as necessary. This is the level of system safety that is actually achieved in light of "as safe as reasonably practicable" (ASARP) considerations that are balanced against security, reliability, operational capabilities, and overall resource constraints (e.g., time, money, skilled personnel). This level of safety may be expected to change over time due to factors such as aging, maintenance, operational conditions, and increases in knowledge, experience, and available technology.

The SNL Asserted (Implemented) Safety for a specific weapon is asserted in the combined engineering judgment of SNL. This 'combined engineering judgment' means the informed and authoritative judgment of executive management, based on the best technical information and opinions of relevant subject matter experts, including independent assessment and others who may have differing professional opinions about a technical solution or its adequacy. Two major parties involved in this judgment process include the *weapon project* (WP) team that is responsible for designing and producing the weapon, and the *independent nuclear safety assessment* (INSA) team that is an independently funded and managed group of experts that reports assessment results to SNL executive management. The independence and autonomy of the INSA team from the WP team is an essential element for ensuring that a complete and balanced body of evidence exists for informing the authoritative judgment of executive management.

4. THE CONCEPT OF INDEPENDENCE

Since IND safety requirements are so stringent (i.e., < 1E-09 per weapon lifetime and < 1E-06 per abnormal environment exposure), multiple safety subsystems are incorporated into NW systems to avoid reliance on a single safety subsystem. Accounts of those who were present at the time these stringent criteria were developed (circa 1968) have stated that the reason for selecting these numbers was that they reflected the following perspectives: (1) the intention to never have an inadvertent nuclear detonation (IND); (2) the intention to ensure that the probability of an IND is vanishingly small; and (3) the belief that *independent* components and subsystems could be engineered such that combinations of them could attain those numerical system safety levels [2]. The U. S. NW safety community has determined that two safety subsystems are the optimal number relative to the quantitative abnormal environment (AE) safety requirements and three are optimal[14] relative to the quantitative normal environment (NEs) safety requirements [9].

---

[13] This section provides the author's 'summary perspective' of the levels of safety that may be identified within the implementation of the SNL nuclear safety design philosophy.

[14] Optimal here indicates a balancing of safety, security, reliability, and operational considerations.

If two safety subsystems each provide a safe response to all AEs, independently, such that the failure to provide a safe response is < 1E-03 per abnormal environment exposure for either subsystem, then the probability that the overall system would respond safely to all AEs could be asserted to be bounded by the product of the subsystem responses. The reductionist approach of designing multiple safety subsystems, each of which are believed to *independently* contribute to achieving the numerical system safety requirements, while mathematically expedient, requires great confidence in the degree of independence achieved. Note also that when decomposing inadvertent nuclear detonation safety subsystem requirements down to the components that comprise a subsystem, the approach has typically been to provide a technical basis that demonstrates an assurance that each component will provide their contribution to the safety function of the subsystem without fail to a level of < 1E-04 to <1E-05 per weapon lifetime or per abnormal environment exposure as appropriate. It is recognized that it is very difficult to provide assurance that the < 1E-03 numerical requirements for a safety subsystem have been met, let alone the < 1E-04 or 1E-05 assignments given to elements or components within a safety subsystem. In fact, it is not possible to amass quantitative data that supports such assertions with a high degree of statistical confidence across all relevant environmental conditions. In other words, it is not possible to conceive of all possible AEs, nor can all environments which can be conceived be tested exhaustively in a repeated fashion to generate overwhelming statistical certainty of weapon response[15] [2]. Thus, since it cannot be rigorously 'proved' that requirements are met, a technical basis is demanded to support assertions that requirements are met.

4.1. Definition & Discussion of Independence and Related Concepts

In recent years, the author of this paper has worked to improve understandings of the concept of independence and how it applies to the *assured nuclear weapon safety* approach. This section summarizes some of that work and may be of use in designing safety features and developing a technical basis to justify safety-related independence assumptions associated with those features. This is the definition of independence that has emerged:

Definition of Independence: The occurrence of a state of one or more things does not provide any information regarding the likelihood of occurrence of another state of one or more things, or sequences thereof.

Discussion: The concept of independence provides meaning only when describing the attributes of the relationship between two or more states,[16] or sequences of states, of one or more things.[17] This independence concept is a fundamental pillar in the domains of probability theory and statistics [2, 10-16]. In probability theory or statistics settings, these "states" of "things" are typically described as events, or outcomes of a process or experiment [14]. Therefore, the definition of independent events would be—the occurrence of one event does not provide any information regarding the likelihood of occurrence of another event. Mathematically, this may be written in the form[18]: event (A) and event (B) are independent events if $P(A|B) = P(A)$ and $P(B|A) = P(B)$, thus $P(A \cap B) = P(A)P(B)$. Recall that in general, $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$, given $P(A) \neq 0$, $P(B) \neq 0$. To reiterate, independent events are defined when the occurrence of one event does not provide any information regarding the likelihood of occurrence of another event, *because* $P(A|B) = P(A)$ and $P(B|A) = P(B)$. It is important to realize that independence is defined by the absence of information.

---

[15] It should be noted that with respect to the aforementioned 'one-point safety' requirements the probability distribution functions describing weapon response behavior are known; therefore, probabilities can be calculated. However, the relevant normal and abnormal environmental contexts extend beyond situations in which the one-point safety probabilities represent bounding probabilities for weapon response.

[16] "State" is defined here as a mode or condition of being [17]. State may be interpreted here to be a single state or a sequence of states.

[17] The definition of "thing" ranges from ideas and concepts to physical objects or processes [17].

[18] Notation: P(A) is the probability of event A occurring. P(A|B) is the conditional probability of "A given that B has occurred." $P(A \cap B)$ means the probability of "A intersected with B," i.e., all points common to both subset "A" and subset "B," or simply the probability of both event A and event B occurring.

Note that the formal definitions of independence are founded upon verification using a sufficient amount of observed data for all relevant contexts. They do not provide much assistance for verification in situations where data are sparse or non-existent [2]. "***Declaring*** events independent for reasons other than those prescribed in (the formal definition) is a necessarily subjective endeavour. In practice, all we can do is look at each situation on an individual basis and try to make a ***reasonable judgment*** (emphasis added) as to whether the occurrence of one event is likely to influence the outcome of another" [15, p. 74].

When discussing independence the terms of "common-cause failure" and "common-mode failure" are often used in a synonymous fashion. However, in the development of a technical basis for independence assertions it is beneficial to give these terms the distinct meanings provided below.

Definition of Common-Cause Failure: Failures involving multiple design attributes of a system that fail as a result of the same causal event or the same type of causal event. The mode of failure across the design attributes may be different (see also the definition for common-mode failures below).

Discussion: Consider a situation in which a flooding event results in the submersion of two components within a subsystem. One component fails to maintain an isolation barrier due to a chemical reaction involving component materials and water. Another component fails to operate due to an electrical short. Both components failed due to a common cause yet the modes of failure were different. Examples of common cause failures include mechanical, electrical, thermal, chemical, biological, moisture, or radiological insults, or failures of common energy connections, materials, power sources, barriers, maintenance, calibration, testing, design, manufacturing, or operational practices. The distinction in terminology between "same causal event" and "same type of causal event" given in the common-cause failure definition is important, since the "same type of causal event" may occur at vastly different times to multiple design attributes within the same system. An environmental example of this situation would be chemical corrosion that slowly leads to penetration of the isolation barrier of one subsystem. Months or years later, the same chemical corrosion process (having breached one barrier) now penetrates the isolation barrier of another subsystem that is nested within the first subsystem. In this case, both the causal events leading to failure were common and the modes of failure were common for the two barriers. A human-related example would involve the performance of the same type of incorrect maintenance or test activity, due to an incorrectly written procedure, on two redundant components within the same system during maintenance or testing events occurring months or years apart. In this case, as in the previous example, the redundant components fail the same way (common-mode) due to the same type of maintenance or testing event (common-cause).

Definition of Common-Mode Failure: Failures involving multiple design attributes of a system that fail in a similar manner as a result of the same causal event or the same type of causal event.

Discussion: Consider a situation in which a combined thermal and crush environment causes two energy isolation barriers associated with two subsystems to rupture in a similar manner. Both components failed in the same mode (rupture) due to the same cause (thermal and crush environment). Common-mode failures can occur across different types of system design attributes or across identical, redundant components such as redundant switches relying on the same conductive electrical energy pathway. Note that the concept of "failure" here refers to a failure to perform an intended function. Failure does not always involve damage to or destruction of an object. For example, two different types of electrical switches may fail in different ways due to a single shock environment. One may violently disassemble (shatter) and another may simply actuate to a closed state. The actuated switch was not damaged, but it may now be in the incorrect state relative to the system-level function it is intended to perform.

To summarize, the distinction made here between common-cause and common-mode makes it easy to conceptually separate the modes of failure manifested across multiple parts of a system given exposure to the same cause. In other words, parts of a system may break or fail to perform their intended

function in different ways given exposure to the same adverse, accidental, or otherwise abnormal environment. In many cases this distinction is not necessary, such as when a power plant maintenance worker uses the same faulty procedure to maintain or test a series of identical valves and leaves them all in the open state when they are normally closed valves. The cause of the failure and the mode of the failure are the same for the series of identical valves. In some cases the distinction between common-cause and common-mode is beneficial, as in the case where the same electrical energy insult (a lightning strike event) subjects the system to high-voltage source with a fast rising current. One part of the system fails due to a minor electrical short, another part of the system deflagrates, another part of the system melts, and another part of the system containing a solenoid actuates to an improper state due to the electromagnetic flux of the event. All of these parts of the system experienced different modes of failure due to a common-cause environment.

Now that some key definitions have been provided, let's return to a central question: How can one develop a rigorous, defensible technical basis justifying independence assertions in situations where data are sparse or non-existent? Independence assumptions can be founded upon the distinctions of *functional*, *temporal*, and *physical* differences. Although the concepts of function, time, and physical properties do not provide mutually exclusive categories for describing dependence, they are proposed as helpful concepts when striving to increase independence between states of one or more things.

Functional independence[19] between two or more states of one or more things is increased when the states are achieved by functions that use different types of energy, logical relationships, materials, mechanisms, and/or methods of operation. Therefore, across a wide range of unintended energy or stimuli best described as functionally variant[20], the independence of the state changes observed between two or more states of one or more things subjected to the same energy or stimuli may be anticipated to increase as the functional differences increase between the possible states of the one or more things. Stated differently, to unintentionally remove energy isolation provided by functionally independent safety features, an accident environment would be required to inadvertently accomplish the same overarching objective (e.g., unintentionally remove electrical isolation barriers) via different means. An example of functional independence in the engineered hardware domain would be to have both hydraulic mechanical actuators and electric motor actuators to control aircraft flight control surfaces. They use different types of energy and mechanisms to accomplish the same overarching goal so a failure in one system will not affect the other and the potential for common-mode and common-cause failures is minimized.

Temporal independence between two or more states of one or more things is increased by manipulating the time when states may be achieved. Therefore, across a wide range of unintended energy or stimuli best described as time variant[21], the independence of the state changes observed between two or more states of one or more things subjected to the same energy or stimuli may be anticipated to increase as the time-related differences[22] increase between the possible states of the one or more things. Typically, but not always, this involves maximizing the time separation of states. Stated differently, to unintentionally remove energy isolation provided by temporally separated safety features, an accident environment would be required to inadvertently exhibit stability for longer periods of time to use design features of the functional system as a significant contributor to compromising a safety subsystem. Temporal independence may involve *minimizing* the time of exposure of energy or information to the system, or *maximizing* the time separation of packets of

[19] Functional independence as used here is basically synonymous with the term *diversity* as may be used when describing engineered safety systems [18]. It is acknowledged that true diversity is difficult to achieve due to common-cause/common-mode failures; this often results in redundancy but not diversity in practice.

[20] For example, thermal energy, high or low voltage electrical energy, direct current or alternating current electrical energy, mechanical shock or vibration, spurious digital messages or data words on a communication bus, corrosive chemicals.

[21] For example, large fluctuations in the rate at which energy or stimuli is imparted to the one or more things.

[22] For example, time-related differences between states could be achieved by limiting times of mechanical operation, using timer-based functions, and manipulating speeds at which digital information can be communicated.

energy, enabling stimuli, or packets of information provided to the system, or some combination of each approach depending upon the environmental context.

Physical independence between two or more states of one or more things is increased when the states must be achieved on different sides of one or more barriers or there are significant intervening structures.[23] Therefore, across a wide range of unintended energy or stimuli best described as physically variant[24], the independence of the state change responses observed between two or more states of one or more things subjected to the same energy or stimuli may be anticipated to increase as the physical differences increase between the possible states of the one or more things. Physical independence is very similar to functional independence and can also be distinguished using the above examples. However, physical independence emphasizes maximizing separation imposed by various types of barriers or structures.

It is important to note that the functional, temporal, and physical separation strategies for increasing independence are different from the concept of redundancy as commonly used when describing engineered safety systems. The concept of redundancy typically involves providing multiple safety-system components or subsystems of the same type in series (e.g., isolation valves) or parallel (e.g., emergency coolant supply valves) as a hedge against failure modes which manifest primarily due to random, independent failure mechanisms associated with the redundant items [18]. Redundancy may be used, for example, when the time between failures of critical design attributes in the redundant components is exponentially distributed and there are no known common-cause or common-mode failure mechanisms (e.g., those associated with aging or wear) that may degrade the redundant components simultaneously.

Ideally, a technical basis for assuring independence between events would be established that is not greatly affected by all three factors of function, time, *and* physical properties, but achieving such a technical basis is difficult for a system that is designed to execute specific goal-directed behaviors such as a nuclear weapon system[25]. Examples of design features that can create unsafe dependencies between safety subsystems in accident environments include: conductor assignments in cables, choices of materials, printed wiring board layouts, computer programming algorithms, types of power supplies, switch designs, collocated isolation barriers, types of enabling stimuli. The designed-in tendencies toward particular responses resulting from such design features are not random. They are neither equally likely nor independent *a priori*, and it is not clear how to generate a credible technical basis for asserting that such independent responses would result across a wide range of adverse, abnormal, or otherwise unintended environments.

---

[23] Here a "barrier" means the presence or absence of any type of matter or energy within a defined space. For example, an air gap maintained by a predictable barrier would itself be a physical barrier to some types of phenomena, a steel plate barrier would be a barrier to some types of phenomena, and separation of a certain distance across the vacuum of space would be a barrier to some types of phenomena. Note however, that barriers used to implement a nuclear weapon safety theme must be predictable across the range of relevant environments. For example, wire separation between the input and output of a safety device maintained merely by a thin layer of non-conductive insulation does not provide predictable separation since there are failure mechanisms which may readily defeat this type of separation (e.g., certain types of vibration-induced damage that may occur in normal environments).

[24] For example, energy or stimuli that varies with respect to the spatial location at which it is applied to the one or more things.

[25] If the only goal was to provide a uniform distribution of well-defined and easily achieved outcomes the design activity becomes much simpler, but still not trivial. For example, it is possible to design a machine that randomly selects numerical "events" within a certain range—these are called lottery machines. Designing a properly functioning lottery machine requires skill, but it is far more difficult to design a lottery machine that is guaranteed to provide you with one particular sequence of ball draws for all "operationally desired" environmental conditions in a highly reliable fashion, and also provides completely random ball draws for all other environmental conditions.

4.2. Practical Tools to Aid in Generating and/or Identifying Independence Features

The process of testing independence assertions should include development of specific propositions, structured in both positive and negative frames of reference, which are tested using theoretical, analytical, and experimental models to provide sufficient knowledge and experience to defend the claim of predictable response in normal and accident environments. The practical tools below provide a method for constructing independence-related propositions to guide construction of a defensible technical basis for independence assertions.

When applying the definition of independence to a specific analysis, it is helpful to complete the following statement employing a *positive frame of reference*:

_____ is independent of _____ with respect to _____.

For example, the energy isolation features of safety subsystem 1 are independent of the energy isolation features of safety subsystem 2 with respect to exposure to a high voltage accidental electrical environment.

For example, the steel isolation barriers for subsystem 1 provide a safe response independently of the aluminium isolation barriers for subsystem 2 with respect to exposure to the chemical _____.

Altering the terminology to create a *negative frame of reference* and requiring open-ended responses for justification is also helpful:

Inadvertent operation of safety interlock 1 does not lead to inadvertent operation of interlock 2 due to the following independence attributes:

- Enablement energy is different (hydraulic versus electric)
- Enablement stimuli is different (two different uncorrelated patterns of 12 binary pulses)
- …

Unintentionally circumventing the _____ does not increase the likelihood of unintentionally removing energy isolation provided by the _____ due to the following attributes of independence:

- …

In accordance with the functional, temporal, and physical decomposition of the independence concept, technical justifications should be required in each of those three areas:

The attributes of independence between the _____ and the _____ with respect to _____ which are best described as (**functional, temporal, physical)** include the following:

- ...

For example, the attributes of independence between the subsystem 1 and subsystem 2 with respect to energy isolation which are best described as **temporal** include the following:

- Viscous damped interlock in subsystem 1 requires 10 seconds to actuate, the spring pin interlock on subsystem 2 re-latches if not operated within 300 ms.
- Subsystem 1 interlock must operate before subsystem 2 interlock

For example, the attributes of independence between interlock 1 and interlock 2 with respect to isolating movement of material which are best described as **physical** include the following:

- Interlock 1 is located 25 meters from interlock 2
- The power supply cable for interlock 1 travels through the east-west cable tray, the power supply for interlock 2 travels through the north-south cable tray.

In addition, one question that can aid in probing for dependencies between safety features is:

"If the electrical or information enabling energy or stimuli were changed, would one or both of the safety features need to be modified to accommodate the change to maintain a safe response in a predictable manner?"

If both safety features must be changed then it may represent a lack of independence between the features.

This same type of question can be adapted to investigate independence-related impacts of any change involving the key elements of the safety theme that implement any of the NSDPs.

Another essential question that probes for independence between an intended weapon environment experienced in normal use and any other environment experienced in an accident is:

"If the mechanical enabling energy or stimuli were changed, what design features need to be modified to accommodate the change to maintain a safe response in a predictable manner?"

If any safety features must be changed, then it may indicate pre-storage of some portion of enabling stimuli information that should only originate from the intended human actions, or only from the physical stimuli experienced by the weapon during normal anticipated operations; thus it is *information or energy that should never be pre-stored* in the system.

In summary, independence assertions can be founded upon the distinctions of *functional*, *temporal*, and *physical* differences. Thus the recommended approach is to implement independence between safety subsystems using functional diversity and the separation of energy and information both physically and temporally. In other words, achieve independence by preventing common-cause and common-mode failures when implementing the relevant NSDPs (incompatibility, isolation, inoperability) by scrutinizing the safety subsystems in terms of functional, temporal, and physical differences. This insight is particularly valuable in situations where a technical basis supporting independence assertions must be created when data are lacking to sufficiently demonstrate compliance with the formal definition of independence.

5. CONCLUSION

This paper provided a basic description of the *assured nuclear weapon safety* approach that uses the Nuclear Safety Design Principles (NSDPs) of *incompatibility*, *isolation*, and *inoperability* to design safety features, organized into subsystems such that each subsystem contributes to safe system responses in *independent* and *predictable* ways given a wide range of environmental contexts. Simplicity of safety features, passive safe responses, and a systematic allocation of basic features among engineered features and human actions are emphasized in the implementation, and an innovative *inside out* process for hazard identification, which was described in this paper, is also applied throughout iterative system design phases to support the process of NSDP integration. The central aim of the *assured nuclear weapon safety* approach is to achieve a robust technical basis for asserting that a system can meet stringent quantitative safety requirements in the widest context of adverse or accident environments, while using the most concise arrangement of safety design features and the fewest number of specific adverse or accident environment assumptions. In essence, this approach claims to be an efficient approach for engineering bounded system safety-related responses.

Particular emphasis in this paper was placed on implementation of the NSDPs using independence assumptions founded upon the distinctions of *functional*, *temporal*, and *physical* differences, which are proposed as helpful concepts when striving to increase the independence of safe responses across multiple safety design features within a system. Clear and distinct definitions for *common-cause failure* and *common-mode failure* (with examples) were provided, as were the attributes of *predictability* that strengthen the technical basis for safe system responses. It is hoped that insights presented here, developed within the U. S. nuclear weapon system safety community over the course

of more than 50 years, may benefit system safety design, assessment, and management activities in other high consequence domains.

Acknowledgements

In summarizing the 'assured nuclear safety concept' the author has condensed hard-won insights earned by many over a span of more than 50 years. Numerous individuals have helped the author understand the development of U. S. nuclear weapon safety, the most important have been two of the 'founding fathers'—Stanley D. Spray and the late J. Arlin Cooper. With respect to recent refinements in the concept of independence and application of the nuclear safety design principles, thanks are due to the author's cadre of colleagues/sounding boards at SNL and across the nuclear security enterprise. The author expresses particular thanks to the seven reviewers of an earlier version of this paper.

References

[1] C. Perrow, "*Normal Accidents: Living with High-Risk Technologies,*" Princeton University Press, 1999, Princeton, NJ.
[2] J. D. Brewer. "*The concept of independence in weapon safety: foundations and practical implementation guidance (SAND2009-4216C)*", Proceedings of the 27th International System Safety Conference, System Safety Society, 2009, Huntsville, AL.
[3] T. Ord, R. Hillerbrand, and A. Sandberg. "*Probing the improbable: methodological challenges for risks with low probabilities and high stakes*", Journal of Risk Research, 13, pp. 191-205, (2010).
[4] DoD, "*Department of Defense Manual 3150.2-M: DoD Nuclear Weapon System Safety Program Manual,*" Assistant to the Secretary of Defense for Nuclear and Chemical and Biological Defense Programs, United States Department of Defense, 1996, Washington, DC.
[5] DOE, "*Department of Energy Order 452.1D: Nuclear Explosive and Weapon Surety Program,*" United States Department of Energy, 2009, Washington, DC.
[6] DoD, "*Department of Defense Directive 3150.02: DoD Nuclear Weapons Surety Program,*" United States Department of Defense, 2013, Washington, DC.
[7] USAF, "*Air Force Manual 91-118: Safety Design and Evaluation Criteria for Nuclear Weapon Systems,*" United States Air Force Safety Center (AFSC/SEWN), 2011, Kirtland AFB, NM.
[8] IAEA, "*IAEA-TECDOC-626: Safety related terms for advanced nuclear plants,*" International Atomic Energy Agency, 1991, Vienna, Austria.
[9] P. D'Antonio. "*Surety Principles Development and Integration for Nuclear Weapons (SAND98-1557)*", High Consequence Operations Safety Symposium II, Sandia National Laboratories, 1998, Albuquerque, NM.
[10] A. J. Duncan, "*Quality Control and Industrial Statistics*," Richard D. Irwin Inc., 1974, Homewood, IL.
[11] I. Miller and J. E. Freund, "*Probability and Statistics for Engineers*," Prentice-Hall, 1977, Englewood Cliffs, NJ.
[12] W. W. Hines and D. C. Montgomery, "*Probability and Statistics in Engineering and Management Science*," John Wiley & Sons, 1990, New York, NY.
[13] E. Kreyszig, "*Advanced Engineering Mathematics, 7th ed.*," John Wiley & Sons, 1993, New York, NY.
[14] W. J. Conover, "*Practical Nonparametric Statistics*," John Wiley & Sons, 1999, New York, NY.
[15] R. J. Larsen and M. L. Marx, "*An Introduction to Mathematical Statistics and Its Applications*," Prentice-Hall, 2001, Upper Saddle River, NJ.
[16] D. C. Montgomery, "*Design and Analysis of Experiments*," John Wiley & Sons, 2001, New York, NY.
[17] Merriam-Webster, "*Definitions of 'state' and 'thing'*," Merriam-Webster Online Dictionary <http://www.merriam-webster.com/dictionary/state>., Retrieved April 30, 2010.
[18] R. A. Knief, "*Nuclear Engineering: Theory and Technology of Commercial Nuclear Power, 2nd ed.*," American Nuclear Society Inc., 2008, La Grange Park, IL.