SANDIA REPORT

SAND2009-1331
Unlimited Release
Printed March 2009

Some Guidance on Preparing Validation Plans for the DART Full System Models

Genetha A. Gray, Richard G. Hills, and Patricia D. Hough

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2009-1331
Unlimited Release
Printed March 2009

Some Guidance on Preparing Validation Plans for the DART Full System Models

Genetha A. Gray and Patricia D. Hough
Advanced Software R&D
Sandia National Laboratories
P.O. Box 969, MS 9159
Livermore, CA 94551

Richard G. Hills
Validation and Uncertainty Quantification
Sandia National Laboratories
P.O. Box 5800, MS 0828
Albuquerque, New Mexico 87185

Abstract

Planning is an important part of computational model verification and validation (V&V) and the requisite planning document is vital for effectively executing the plan. The document provides a means of communicating intent to the typically large group of people, from program management to analysts to test engineers, who must work together to complete the validation activities. This report provides guidelines for writing a validation plan. It describes the components of such a plan and includes important references and resources. While the initial target audience is the DART Full System Model teams in the nuclear weapons program, the guidelines are generally applicable to other modeling efforts. Our goal in writing this document is to provide a framework for consistency in validation plans across weapon systems, different types of models, and different scenarios. Specific details contained in any given validation plan will vary according to application requirements and available resources.

ACKNOWLEDGMENTS

The authors would like to acknowledge Mike Hardwick, Scott Klenke, Robert Mariano, Jim Stewart, Brian Adams, and Tony Giunta for their feedback on this document and for their insights into both the programmatic and practical aspects of V&V within the Advanced Simulation and Computing (ASC) program. In addition, the authors would like to thank Monica Martinez-Canales, on whose template for the WR1 this document is based. We also thank Tim Trucano for his thoughtful comments and insights, many of which are included in this work.

CONTENTS

FIGURES

TABLES

NOMENCLATURE

CTBT	Comprehensive Test Ban Treaty
DART	Design through Analysis Realization Team
DoD	Department of Defense
DOE	Department of Energy
FSM	Full System Model
I/O	Input/Output
IC	Initial Condition
IET	Integral Effects Test
M&S	Modeling and Simulation
NNSA	National Nuclear Security Administration
PCMM	Predictive Capability Maturity Model
PDE	Partial Differential Equation
PIRT	Phenomena Identification and Ranking Table
QMU	Quantification of Margins and Uncertainties
SA	Sensitivity Analysis
SQE	Software Quality Engineering
SRQ	System Response Quantity
SSP	Stockpile Stewardship Program
UQ	Uncertainty Quantification
V&V	Verification and Validation

1. OVERVIEW

In the mid-1990's, the United States established a policy of non-nuclear weapon testing to comply with the Comprehensive Test Ban Treaty (CTBT) [CTBTO, 1996; Clinton, 1995 & 1996]. Pre-CTBT, the US Weapons program relied on nuclear weapon (NW) testing data to evaluate nuclear explosive performance of new designs and to calibrate nascent nuclear design codes with nuclear test data. Post-CTBT, the nuclear weapons program shifted focus from designing new weapons to one of sustaining existing warheads for the indefinite future. As a result of the CTBT and the shift in focus, NNSA initiated the science-based Stockpile Stewardship Program (SSP) to develop and apply "improved technical capabilities to assess the safety, security, and reliability of existing nuclear warheads without the use of nuclear testing" [NNSA website] through the use of integrated science-based computer simulation.

The elimination of full system nuclear weapon testing and the increased reliance on computer simulation has changed the role of testing in assuring safety, security, and reliability of existing warheads. Testing is now focused on the subsystem and sub-physics levels, with an emphasis on the support and evaluation of full and sub-physics computational models for nuclear weapon systems. The advances in computer modeling capabilities have enabled simulation to become more integral and important part of the weapon safety, security, and reliability technical basis. To further the deployment of these capabilities, Sandia's Design through Analysis Realization Team (DART) established an effort to develop full weapon system analysis models. Validation of these Full System Models (FSMs) using experimental data is a critical component in insuring their adequacy.

1.1. Document Purpose and Scope

In order to assess Advanced Simulation & Computing (ASC) analysis code and computational model predictive credibility, Sandia has developed guidelines for developing verification and validation (V&V) plans [Trucano, *et al.*, 2002; Pilch, *et al.*, 2000]. The goal of verification is to determine if the equations underlying the computational model are being solved correctly. The purpose of validation is to quantify the degree to which a computational model is an accurate representation of the real world phenomena it seeks to represent. The validation process includes making reasonable comparisons between physical experiments and code calculations. Once validated, numerical models can be used to calculate design margins and quantify the uncertainty surrounding those margins. The fundamental idea behind V&V and quantification of margins and uncertainties (QMU) is that these processes will supplement current qualification activities to provide increased confidence in the weapon safety and reliability technical basis. This document presents some specific guidelines for preparing a validation plan.

The purposes of this document are to summarize some of the important points of the growing V&V literature and to provide general guidelines to Sandia's DART FSM teams. This document focuses on preparing the validation plans, not on the actual validation activities. Specifically, this document will be used by the FSM teams as a guide for FY08 validation planning for validation activities to be implemented in FY09 and beyond. There are currently 13 FSM teams working on 7 systems as follows:

- B61: Radiation Transport
- B83: Abnormal Mechanical & Abnormal Thermal
- W76-0: Abnormal Mechanical & Abnormal Thermal
- W78: Abnormal Thermal
- W80: Electrical & Abnormal Mechanical (W80-1)
- W87: Abnormal Mechanical & Abnormal Thermal
- W88: Abnormal Thermal & Electrical & Radiation Transport

Although this document was prepared specifically for the FSMs named above, we hope that the guidance is more widely applicable and assists in the validation planning activities of other systems.

The planned audience for this document consists of those who have a background in computer modeling of engineered systems but are new to model validation. The document provides a step by step process to be used for creating FSM validation plans. This document reiterates the general V&V plan guidelines presented by Trucano, *et al.* in SAND2002-0341, and it augments those guidelines with principles that have been developed since that time. In addition, it includes elements that are often included in validation plans but that are not explicitly recorded in existing guidance and elements that, in our practical experience, are discussed in the planning phases but sometimes go undocumented. The guidance presented in this document represents the minimum process that Sandia FSM validation planning efforts should follow. This document will be updated as more validation plans are created.

1.2. The Benefits of Validation

If a computational model is proven reliable under a range of operating conditions, then system performance and reliability at other points may be inferable using the model. Examples include

- other conditions of operation for which the model was not tested, including across a design space,
- other regimes of operation for which it is not possible to test (due to expense, time, lack of suitable instrumentation, or testable conditions),
- changes in a system that have undergone design changes – in this case, subsystems may need to be retested and the validation results of the system level model updated, and
- system level models validated under similar system conditions can be used to evaluate the environmental conditions for which a redesigned subsystem (internal component) should be tested.

In addition, model validation methodology can be used to

- evaluate whether the validation experiments are adequate for a specific application (e.g., design goals) and
- to assess the impact of uncertainty on the design of a validation experiment.

Validation benefits the following constituents:

- Customers - provides some idea of the quality/accuracy of the product for which they are paying and any simulation results upon which potential decisions will be based.
- Code groups and modelers - provides justification/confidence in claims of accuracy/predictivity for capabilities they are selling to customers.
- Management - can be used for resource allocation and investment decisions regarding which modeling capabilities are most lacking in accuracy/predictivity or are uncharacterized or undercharacterized in terms of validation evidence.
- Outside auditing or independent assessments - provides well documented evidence of quantitative and tangible procedures for validation.

1.3. The Relationship Between V&V, UQ, and QMU

Questions regarding the differences between V&V, uncertainty quantification (UQ), and quantification of margins and uncertainty (QMU) often arise. We provide here a very brief description of how the three are related to each other. V&V is a rigorous process for building confidence in the predictive capability of a model. UQ is one of the fundamental tools in executing the V&V process. In particular, it brings rigor to the analysis of both computational and experimental data and the comparison between the two. QMU is a process by which the predictive model is used to assess performance margins and associated uncertainties. UQ again is the tool that provides statistical rigor in the analysis of the performance requirements, the model predictions, and the comparison between the two. The evidence collected in the V&V process serves as a means of conveying the level of confidence in the QMU predictions. This relationship is illustrated in its simplest form in Figure 1.

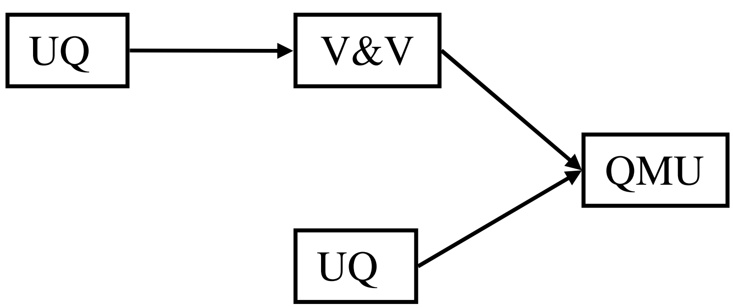

Figure 1 *This figure is a very simple illustration of the relationship between V&V, UQ, and QMU. UQ is a tool that is used to perform V&V and QMU. The evidence gathered during the V&V process supports the model predictions used for QMU.*

While further discussion of the QMU process is beyond the scope of the document, we do note that it is very similar to the V&V process. Details are described in [Pilch, *et al.*, 2006; Diegert, *et al.*, 2007]. Furthermore, we include some reference to margins in the V&V plan guidance with the hope of providing a forum for discussion of margin assessments across the range of weapon systems.

1.4. V&V is a Collaborative Process

Before launching into the remainder of this document, we first note a critical aspect of V&V projects. V&V is a *collaborative* process. It not only encourages, but requires participation from analysts, system engineers, code developers, and V&V specialists. For this reason, all parties should be engaged from the start. We strongly encourage the use of this planning process as a vehicle for that engagement. In particular, we suggest that the development of the V&V planning document include a process by which feedback and buy-in from all parties is requested, documented, and incorporated into the V&V plan. We work under the assumption that the analyst/modeler has the primary responsibility for developing the V&V plan and point out the partners we believe should be engaged at each stage.

2. VALIDATION PLAN STRUCTURE

The parts of a validation plan mirror the validation process. In this section, we describe the structure of a validation plan and its necessary components, methodologies, and participants. Note that while the overall structure of the validation plan will be the same for each FSM team, the details and complexity of each component will likely vary greatly. Furthermore, the completeness and maturity of each section may vary for any number of reasons. Thus, the V&V plan should be viewed as a living document that is updated as information documented in each section changes over time.

2.1. Overview of the Validation Process

Discussion of the complete validation process is outside the scope of this document; however, we provide a summary for reference. The validation process can be broken down into eight steps as follows:

*__Step 1:__ Identify validation requirements for the application
*__Step 2:__ Identify the critical physical phenomena
*__Step 3:__ Identify our ability to model the critical physical phenomena
*__Step 4:__ Identify the high priority validation experiments test, including critical design variables and validation metrics
Step 5: Experimental design for experiments identified in Step 4.
Step 6: Perform experiments
Step 7: Update model and perform model validation for experiments performed in Step 6
Step 8: Integrate results across validation hierarchy

Those steps marked with an asterisk are explicit planning activities and are discussed further in this document. The remaining steps will not be discussed, but FSM teams should be aware and mindful of them as validation plans are intended support the execution of these steps. FSM teams should seek out assistance, explanations and references from V&V experts as needed.

2.2. Plan Part 1: Application Requirements

Partners: modelers/analysts, system engineers, V&V specialists

Motivation: V&V is an inherently application-driven process, and thus, V&V activities should be carried out in a manner that covers all aspects of the intended use of the model (to the extent possible). In order to provide line of sight from the various elements of the V&V plan to the application itself, it is important to document the application requirements.

Key Elements:

- Stockpiler driver (e.g., W76-1 Life Extension)
- Qualification issues (e.g., Loss of Assured Safety in Abnormal Environment)
- Scenarios of interest (e.g., Fuel Fire)
- Performance characteristics to be studied (e.g., Weak Link/Strong Link Failure Delays)
- Margin definitions and requirements

Additional Comments on QMU: The Quantification of Margins and Uncertainties (QMU) is closely related to V&V and should be kept in mind throughout the planning process. The specifics of QMU planning and processes is outside the scope of this document; however, margin assessment inevitably falls into the category of validation requirements. Therefore, this section of the validation plan is ideal for documenting any performance requirements.

2.3. Plan Part 2: Phenomena Identification and Ranking

Partners: modelers/analysts, system engineers, code developers, V&V specialists
Reference: SAND2000-3101

Motivation: Phenomena identification and ranking tie the key stockpile, qualification, or other application requirements to the key physical phenomena being modeled. This activity provides the logical link between validation activities and the application requirements. In addition, it is extremely useful for prioritizing testing, including resources for physical tests, based on the expected impact to the application.

Key Elements:

- Phenomena Identification and Ranking Table (PIRT; see guidelines in Appendix A)
- Documentation of PIRT process (e.g., assumptions made, ranking criteria, etc.)

Additional Comments on the PIRT: The Phenomena Identification and Ranking Table (PIRT) includes all quantities of interest (QoI) and system response quantities (SRQ) and ranks, relative to a well-defined figure of merit, the relative importance of the physical phenomena that affect the SRQs. An initial assessment of the simulation adequacy for each of the “required” physical phenomena of interest is also included. Finally, any known requirements on the SRQs should be documented at this phase of planning.

Appendix A, included at the end of this document, gives the guidelines and descriptor definitions to be used in the PIRT. Phenomena importance and elements of simulation adequacy are ranked using “high”, “medium”, and “low” descriptors. Color coding adds a visual metric to the PIRT to assist in elucidating knowledge gaps that are used to prioritize V&V work. Green indicates no gap between phenomena importance and simulation element adequacy, yellow indicates a gap of one descriptor level (e.g., importance is high but adequacy is medium), and red indicates a gap of two descriptor levels (e.g., importance is high but adequacy is low). Note that PIRT rankings reflect current knowledge about the modeling and simulation structure and thus should be created with input from both the code and model developers.

2.4. Plan Part 3: Predictive Capability Maturity

Partners: modelers/analysts, system engineers, code developers, V&V specialists
Reference: SAND2007-5948

Motivation: Assessing the predictive capability of the model focuses on the extent to which model components have been calibrated, validated, and exercised in the context of scenarios of interest. Comparing a current assessment with required maturity levels provides a road map for model development. Furthermore, since model development is an iterative process, assessing predictive capability should also be an iterative process and provides a means of tracking progress in model development.

Key Elements:

- Predictive Capability Maturity Model (PCMM; See guidelines in Appendix B) – initial state
- PCMM – required end state
- Plan for version control of the model, including corresponding PCMM

Additional Comments on the PCMM: The PCMM appears to be very similar to the PIRT, but the two are quite different in intent and level of granularity. The PIRT is intended to identify and prioritize gaps in knowledge about the phenomena that affect quantities of interest for the intended application. It is used to prioritize validation activities and to drive the closure of the most important knowledge gaps. In contrast, the PCMM is a neutral mechanism intended to organize, summarize, and communicate evidence of the predictive capability of a given model for a given application. That evidence is accumulated by evaluating specific modeling elements that contribute to predictive capability. We recommend negotiation between analysts and system engineers to establish the required level of predictive capability for the intended application. Given such a requirement, gaps between the current and required levels can be identified and used to prioritize investment in advancing the predictive capability of the model. As with the PIRT, color coding can be used to elucidate these gaps.

The development of a simulation capability is an iterative process. Furthermore, limited resources and aggressive timelines often result in the undesirable situation of using a model for analysis while it is still under development. The PCMM should therefore be a living document, and each instantiation should be linked to a specific version of the associated model and code used to execute the model. The underlying assumption is that development of the model is managed using some from of version control. While there are numerous systems for version control, we recommend using one that is already in use and consistent with the requirements of the customer in order to avoid unnecessary duplication. This will allow clear identification of which version of the model is used in any given analysis, and it will allow easy accessibility and reuse of that version of the model. The end goal is a means of reproducing any given analysis and the level of confidence that can be associated with that analysis. Finally, we also note that a PCMM has been found to be useful by others to convey information about the maturity of computational models [NASA, 2008].

2.5. Plan Part 4: Verification

Partners: modelers/analysts, code developers, V&V specialists
References: SAND2000-3101

Motivation: Verification is a critical and highly related activity that should be addressed in validation planning efforts. Historically, there are two components to the verification: 1) code verification and 2) solution verification. We suggest including a third component we will describe as input verification. We assume that the development and execution of software quality plans are managed by code developers, so it is the application-specific components of verification activities that receive the most extensive treatment here.

Key Elements:

- Code Verification
  - References to code documentation (e.g., users manual, theory manual, software quality practices, etc.)
  - Reference to relevant verification test suites
  - Gaps in verification test suites
  - New verification tests needed
- Input Verification
  - Input decks to be checked
  - Mesh and other model properties to be checked
- Solution Verification
  - Type of mesh and mesh refinement strategy
  - Metrics for assessing mesh convergence
  - List of numerical parameters to be studied and ranges

Additional Comments on Code Verification: Code verification encompasses standard software quality engineering (SQE) practices. While this falls in the domain of code developers, it is a factor in simulation predictive credibility. As such, appropriate pointers and references warrant inclusion. More specifically, it is not necessary to document the SQE practices employed by code developers in the validation plan. Instead, references to documentation, such as the user's manual and SQE plan, are appropriate. In collaboration with code developers, the regression and verification test suite should be evaluated for adequacy with respect to the PIRT. Any gaps and a path forward for addressing those gaps and any commonalities with other model needs can also be included in the validation plan. To re-emphasize, this should be done in collaboration with code developers.

Additional Comments on Input Verification: Input verification is an area that falls somewhere between code verification and solution verification. This activity is intended to get at the question of whether or not the model was constructed as intended (denoted in the following sentences as "correctly"). For example, are the material properties correct? Was the mesh generated correctly? Are boundary conditions defined correctly? Are material properties mapped to the correct part of the mesh? Any activities one might refer to as "model checking" should be included in this section.

Additional Comments on Solution Verification: Solution verification is the responsibility of the analyst. A significant component of solution verification is a grid convergence study, which provides some basis for numerical accuracy assessment. Thus, one category of information to document in verification is candidate mesh refinement techniques, metrics to assess grid convergence for each quantity of interest, and gaps in the availability of appropriate metrics. In addition to grid convergence, solution verification can also include studying model sensitivity to code knobs (i.e., numerical parameters). A list of such parameters to which the model may be sensitive can be included here in the validation plan and prioritized relative to the PIRT and according to “expert judgment”. We caution that “expert judgment” can be highly subjective and is not the preferred approach to prioritizing parameters, but its use is usually inevitable due to resource limitations. The use of “expert judgment” should be explicitly identified, and the assumptions and evidence supporting it should be well documented along with the qualifications of the expert(s) so that the expertise can withstand scrutiny. We note that it may be the case that not all parameters are studied in the end; however, we encourage the inclusion of more parameters rather than fewer in the interest of documenting priorities and assumptions.

2.6. Plan Part 5: Validation Test Suite (VALTS)

Partners: modelers/analysts, system engineers, experimentalists, V&V specialists
Reference: SAND2002-0341

Motivation: The VALidation Test Suite (VALTS) is the set of physical experiments that will generate experimental data to which model predictions will be compared. The VALTS should logically tie to the PIRT. Ultimately, a full-fledged test plan will need to be developed, but in this initial planning document, the focus is on establishing the framework in which that test plan should be developed.

Key Elements:

- Illustration/description of system hierarchy
- Description of each tier of phenomena to be studied
- Identification of any legacy data sets
- Assessment of legacy data sets (i.e., is it suitable for use in validation or only for providing some initial insights?)
- High-level description of tests to be done for each tier at each level of the system hierarchy
    - Quantity or quantities of interest to be measured
    - Hardware configuration(s) of interest
    - Tie to PIRT
- Validation metrics (for comparing model prediction to test)

Additional Comments on the Hierarchical Approach: Because FSMs are highly complex, a hierarchical approach to validation is often required. In this case, validation activities begin with an investigation of partial physics models, component, or sub-system models and then work up to the full system. Such an approach may increase the validation process efficiency, decrease the time and resources required to complete the process, and result in greater insight into the issues identified in the PIRT and PCMM. While there may be cases in which it seems that a hierarchical approach to validation is not reasonable, it is more likely the case that one will not be able to fully understand the reasons why model and experimental data agree or disagree at the system level without drilling down to the lower levels of the hierarchy.

The validation plan should explicitly state the hierarchy to be used for validation. A hierarchical approach to validation must describe the process for moving up the hierarchical levels, which may not be immediately obvious. It should also be noted that as one progresses up the hierarchical chain, the number of validation tests will likely decrease while the complexity and fidelity of the model increase. Moreover, it will be harder to confirm that the right experiment was done and done correctly, and the results will become harder to interpret.

Additional Comments on Tiered Validation: Validation testing should be designed using the four tiered approach described in SAND2000-3101. Each tier explores a different aspect of how the physical model and the associated numerical model represent the real world:

- Tier I – Separable effects (or single phenomena).
- Tier II – Coupled effects of distinctly identified phenomena.
- Tier III – Integral phenomena, in which many coupled effects may be present.
- Tier IV – Readiness with respect to stockpile computing

Note that as the tier level increases, the complexity of the experiments required for the validation activity typically also increases. Tiers I through III are directly expressed in the PIRT while Tier IV is a qualification activity for the computing associated with the stockpile driver. Validation plans will not necessarily include tests from all four tiers. If Tier IV activities will be included, it is critical to get input from the weapons designers.

The hierarchy of a FSM should not be confused with the tiered approach to validation testing. It is possible to apply the tiered approach at each level of the hierarchy. In other words, each subsystem should be evaluated to determine which tiers are applicable, and validation tests should be designed accordingly. In the case that the hierarchical model only has a single level, each tier of tests will be performed for that one level.

Additional Comments on Data: Data is the main driver of all validation activities. The experiments required to collect the necessary data are expensive and often a limiting factor. Thus, a validation plan should be an assessment of alternative sources of data including shared data and legacy data.

The use of shared validation data for common or similar components, across multiple weapon system applications, can significantly reduce the costs associated with acquisition and analysis of data and can also help consolidate aspects of the validation process. This approach requires that those associated with multiple analysis and experimental teams work together to insure that the experiments are designed to characterize the model parameter ranges and experimental conditions required for all of the relevant systems. The use of a common approach to validation planning will greatly facilitate sharing data. The VALTS part of the validation plan should identify any possible places where data can be shared and with whom.

Often, some part of or the entire system of interest has been tested previously and that data may be available for use in upcoming validation activities. However, legacy data must be carefully evaluated before use. The experiments that produced such data may not have been designed to address the more rigorous requirements associated with model validation. The conditions of the experiments (including complete specification of boundary and initial conditions), the characterization for the important constitutive properties and their uncertainty, and the assumptions used to design the experiments and reduce the data are often not adequately documented. Even in this case, legacy data can still be useful in that the results can provide valuable qualitative insight as to the rankings associated with the PIRT, the conditions for which models should be tested, and can provide guidance for future experimental designs. The validation plan should include a list of the relevant legacy data and a plan for evaluating its usefulness in the upcoming validation activities.

Additional Commnets on Metrics: Validation metrics are the mathematical/statistical definitions of the quantities to be compared for the assessment of the model. Examples include differences between observed and predicted temperatures in a component, or the sum of squares of differences over a several measurements. These metrics may be based on quantities that are directly measureable, or based on quantities that require post processing of the experimental data. A validation metric must give a meaningful indication of the model's predictive performance within the application context and is often specific to the application of interest. [Oberkampf, *et al*., 2005]

Metrics are an important aspect of the VALTS as they identify and insure that important model parameters are quantified and provide a range of conditions over which to perform the experiments identified in VALTS. An emphasis on rigorous comparison insures that validation metrics are viable and useful given the realities associated with experimental work, the numerical models, and any given requirements. Decisions regarding which metrics to use should tie back to the qualification requirements and consider input from the DSW customer. Choosing validation metrics must be collaborative. The validation plan should include possible validation metrics. Note that at this time, the plan may include a specific metric or some reasonable possibilities which will be evaluated at the time that the validation activities are carried out.

2.7. Plan Part 6: Uncertainties

Partners: modelers/analysts, system engineers, experimentalists, V&V specialists
Reference: Helton (1994); Oberkampf, *et al*. (2002)

Motivation: At the heart of the validation activities is the comparison of experimental data and simulation results. Complicating this comparison is the fact that both contain uncertainties which must be quantified in order to make reasonable comparisons. Moreover, all significant uncertainties related to the FSM must be identified, characterized, reduced, or eliminated if possible. Identifying uncertainties and methods for quantifying them is part of the validation planning. At this stage, this list may not be exhaustive. It can often change during the course of the validation activities. However, it is important to make an initial list of uncertainties and how they relate to the PIRT, PCMM, and VALTS as part of the validation plan.

Key Elements:

- Model Uncertainties, Types, and Characterizations (to the extent known)
    - Material properties
    - Boundary and initial conditions
    - Empirical inputs
    - Model simplifications
    - Missing physics
- Experimental Uncertainties, Types, and Characterizations (to the extent known)

- Unit-to-unit properties and assembly
  - Measurement techniques, post-processing, and interpretation
  - Instrument calibration
  - Environmental conditions
  - Experimental biases
- Scenario Uncertainties, Types, and Characterizations (to the extent known)
  - E.g., drop angle
- Additional Parameters for Sensitivity Analysis
- Plan for Quantifying and Propagating Uncertainty

Additional Comments on Uncertainty Types: Two types of uncertainties are generally associated with the parameters associated with the validation of FSM. Aleatory uncertainty occurs due to random process such as unit-to-unit variability, material variability or dimensional variability and is characterized through probability density functions. Epistemic uncertainty (or incomplete knowledge) is associated with non-random effects, such as bias introduced by diagnostic calibration error, or non-physical parameters used to 'tune' models. As FSM become more complex, the ratio in the number of epistemic to aleatory parameters can increase. A validation plan must address these uncertainties.

Additional Comments on Sensitivity Studies: Quantifying uncertainties in both the experimental data and computational results usually requires one or more sensitivity studies. In the case of physical experiments, sensitivity analysis can be used to determine the effect on the data given changes in experimental conditions [Hu & Hamada, 2000; Montgomery, 2001; Ross, 1996; Saltelli, *et al*., 2000]. In the case of computational simulators, the sensitivity analysis focuses primarily on how changes in the input affect the output.

Uncertainties in data from experiments may be introduced by measurement and/or instrumentation calibration, experimental setup, and/or experimenter execution. In this case, results of sensitivity analysis are used to determine whether or not sensitivities are physically reasonable, are the result of measurement errors, or are a fault in the testing process. The experimental data needed to complete such a study is an important planning consideration.

For computational results, sensitivity analysis is particularly useful for determining the impact of simulation parameters. If a small change in a parameter results in relatively large changes to the simulation output, then the code is sensitive to choice of that parameter. This means that the parameter has to be determined very accurately in order for the code to be useful in accurately representing real world situations. In contrast, if a code is shown to be insensitive to a particular parameter, then that parameter can be chosen without a high level of accuracy. The validation plan should include the parameters to be studied and their characteristics (i.e. bounds, specification values, etc). Moreover, because sensitivity or insensitivity to a parameter can also indicate a problem with the code, the behavior of physics based parameters should be investigated to determine if similar behavior is observed experimentally.

Additional Comments on Hierarchical Uncertainty Propagation: In a hierarchical model with more than one tier, it is often necessary to conduct a more extensive study of how the simulation-related uncertainties are propagated through codes. The will help quantify the effect of the uncertainty and whether or not a particular uncertainty must be reduced or eliminated in order to meet specific code validation requirements. The validation plan should note what if any hierarchical UQ will need to be studied and any approaches that might be applied. For more information on this topic, see Helton (1994). In addition, Appendix D lists several tools and resources that are available to support both sensitivity studies and uncertainty quantification.

2.8. Plan Part 7: Resource Constraints and Risk Management

Partners: modelers/analysts, system engineers, code developers, experimentalists, V&V specialists

Motivation: Validation activities represent a significant undertaking and thus planning should include addressing the risks and outlining mitigation plans. The most common risks are associated with resource limitations, so we mention some of those specifically. Nonetheless, that should not preclude including any other known risks.

Key elements:

- Known Risks (if applicable)
    - Funding level
    - Experimental hardware availability
    - Required computational time
    - Test facility availability
    - Human resources
    - Other
- Risk Mitigation (if appropriate)
    - Identification of alternate resources
    - Identification of opportunities to combine resources with other teams
    - Prioritization of resources according to PIRT
    - Extended time lines
    - Other

Additional Comments on Resource Constraints: Validation is a resource intensive activity. Thus, there is always a tension between resource needs and availability. Understanding resource constraints, including available hardware and software, laboratory equipment and measurement capabilities, staff availability, and funding, is an essential part of validation planning. A good validation plan will outline available and needed resources. Note that this is also an opportunity to interact with other FSM teams in order to discuss sharing resources and funding.

This page intentionally left blank.

3. REFERENCES

[1] Diegert, K., Klenke, S., Novotny, G., Paulsen, R., Pilch, M. and T. Trucano. "Toward a More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Life Cycle – A Sandia Perspective." Technical Report SAND2007-6219, October 2007, Sandia National Laboratories, Albuquerque, NM 87185.

[2] Eldred, M.S., Brown, S.L., Adams, B.M., Dunlavy, D.M., Gay, D.M., Swiler, L.P., Giunta, A.A., Hart, W.E., Watson, J.-P., Eddy, J.P., Griffin, J.D., Hough, P.D., Kolda, T.G., Martinez-Canales, M.L. and P. J. Williams, "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 4.0 Reference Manual," Technical Report SAND2006-4055, October 2006. Updated September 2007 (Version 4.1). Sandia National Laboratories, Albuquerque, NM 87185.

[3] Helton, J. C., "Treatment of Uncertainty in Performance Assessments for Complex Systems," *Risk Analysis*, vol. 14, no. 4, p. 483-511, 1994.

[4] JMP®, SAS Institute Inc., SAS Campus Drive, Cary, North Carolina.

[5] Oberkampf, W. L., S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin, "Error and Uncertainty in modeling and simulation," *Reliability Engineering and System Safety*, vol. 75, p. 333-357, 2002.

[6] Oberkampf, W. L., and Barone, M. F., "Measures of Agreement Between Computation and Experiment: Validation Metrics," Technical Report SAND2005-4302, Sandia National Laboratories, Albuquerque, NM 87185.

[7] Oberkampf, W. L., Pilch, M., and T. G. Trucano, "Predictive Capability Maturity Model for Computational Modeling and Simulation," Technical Report SAND2007-5948, UUR, October 2007, Sandia National Laboratories, Albuquerque, NM 87185.

[8] Oberkampf, W. L., Trucano, T. G. and C. Hirsch, "Verification, Validation, and Predictive Capability in Computational Engineering and Physics," *Applied Mechanics Reviews*, vol. 57, no 5, p. 345-384, 2004. (Note: Also available as SAND2003-3769)

[9] Pilch, M., Trucano, T. G. and J. C. Helton. "Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper." Technical Report SAND2006-5001, Sandia National Laboratories, Albuquerque, NM 87185.

[10] Pilch, M., Trucano, T. G., Moya, J., Frohlich, G., Hodges, A. and D. Peercy, "Guidelines for ASCI verification and Validation Plans – Content and Format: Version 2.0." Technical Report SAND2000-3101, Sandia National Laboratories, Albuquerque, NM 87185.

[11] CTBTO, *Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization*. 1996.

[12] Clinton, W.J., *Statement on a Comprehensive Nuclear Weapons Test Ban*, in *Weekly Compilation of Presidential Documents*. 1995, 1995 Presidential Documents, GPO Access: Washington, D. C. p. 1383-1437.

[13] Clinton, W.J., *Statement on the Comprehensive Test Ban Treaty*, in *Weekly Compilation of Presidential Documents*. 1996, 1996 Presidential Documents, GPO Access: Washington, D. C. p. 1105-1149.

[14] NASA. "Standard for Models and Simulations," NASA-STD-7099, approved 7-11-2008.

[15] NNSA. *Defense Programs*. Available at http://www.nnsa.doe.gov/defense.htm.

[16] Trucano, T. G., Pilch, M. and Oberkampf, W. L., "General Concepts for Experimental Validation of ASCI Code Applications," Technical Report SAND2002-0341, Sandia National Laboratories, Albuquerque, NM 87185.

[17] Hu, C.F., and M. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization*. 2000: Wiley-Interscience.

[18] Montgomery, D.G., *Design and Analysis of Experiments*. 5th ed. 2001: Wiley.

[19] Ross, P.J., *Taguchi Techniques for Quality Engineering*. 2nd ed. 1996: McGraw-Hill.

[20] Saltelli, A., K. Chan, E. M. Scott, ed. *Sensitivity Analysis*. Wiley Series in Probability and Statistics. 2000, Wiley.

APPENDIX A: PIRT GUIDELINES

Note: This information is reprinted from SAND2002-1740.

Table 1 Guidelines for Importance Ranking

Descriptor	Definition
High	First order importance to metric of interest. Model adequacy, code adequacy, and validation adequacy should be at the "High Level."
Medium	Secondary importance to metric of importance. Model adequacy, code adequacy, and validation adequacy should be at least the "Medium Level."
Low	Currently believed to be of negligible importance to metric of interest. Not necessary to model this phenomena for this application.
Uncertain	Potentially important. Importance should be explored through sensitivity study or discovery experiments and the PIRT revised.

Table 2 Guidelines for Assessing Model Adequacy

Descriptor	Definition
High	A mature physics-based model or correlation-based model is available that is believed to adequately represent the phenomenon over the full parameter space of the application.
Medium	Significant discovery activities have been completed. At least one candidate model form or correlation form has emerged that is believed to nominally capture the phenomenon over some portion of the application parameter space.
Low	No significant discovery activities have occurred and model form is still unknown or speculative.
Strategy	Inadequacies are addressed through an explicitly stated strategy. This may include acceptance of the inadequacy, the parallel use of alternate plausible models, the use of stylized bounding models, or other documented strategies.

Table 3 Guidelines for Assessing Code Adequacy

Descriptor	Definition
High	The intended model is implemented in the code. An adequate regression suite is run routinely, and there are specific problems in the regression suite that test the implementation of the specified model. Verification problems have been run that test the correctness of the numerical implementation. Enabling code features are fully operational. There are no outstanding (reported) bugs or issues that can undermine usage of the model.
Medium	The intended model is implemented in the code. There is an inadequate regression suite or the regression suite does *not* specifically touch the phenomena of interest. The verification suite does not address the specific numerical implementation. Certain enabling code features are not fully functional. There are no outstanding (reported) bugs or issues that can undermine credibility of the proposed calculations.
Low	The intended model is not implemented in the code. The regression suite or the verification suite are inadequate. Certain enabling code features are not functional preventing the calculation from being run. There are outstanding code bugs or issues that must be resolved before model usage.
Strategy	Inadequacies are addressed through an explicitly stated strategy. This may include acceptance of the inadequacy, workarounds, or other documented strategies.

Table 4 Guidelines for Assessing Material Property Adequacy

Descriptor	Definition
High	Data-based [isotropic or orthotropic (if needed)] properties specific to the application material, with adequate [temperature] dependency over the [temperature] range of the application. This includes phase change properties if phase change is expected.
Medium	Application-specific material properties are available or missing properties can be estimated based on well-established theory. Deficiencies in [temperature] dependence, [orthotropic] behavior, or phase change behavior may exist relative to the parameter space of the application.
Low	Not able to estimate some or all of the material properties when a relevant database is not available.
Strategy	Inadequacies are addressed through an explicitly stated strategy. This may include estimated or inferred properties from similar materials, or other documented strategies.

Table 5 Guidelines for Assessing Validation Adequacy

Descriptor	Definition
High	Complete validation evidence to use the model for the intended application. Predictive uncertainties of the model or correlation are quantified over the full parameter space of the application or over the parameter space of the database and the degree of *extrapolation* to the application is quantified and justifiable. The database is relevant to the application.
Medium	Partial validation support for model use in the intended application. Some validation evidence exists, but there are known gaps. Non-statistical comparisons of experiment data such as tabular comparisons or data trace overlays are employed. The degree of extrapolation (if any) may not be quantified. The database may not be fully relevant to the application.
Low	Insufficient validation support for model use. No significant comparisons with experiment data or ad hoc comparison of experiment "pictures" with prediction. The database is not relevant to the application.
Strategy	Inadequacies are addressed through an explicitly stated strategy. This may include acceptance of the inadequacy, uncertainty quantification, or other documented strategies.

This page intentionally left blank.

APPENDIX B: PCMM GUIDELINES

Table 6 Scoring Guidance for PCMM Entries (See Table 3 on p. 38 of SAND2007-5948.)
Blue headings on the left column are required headings.

MATURITY / ELEMENT	**Maturity Level 0** Low Consequences, Minimal M&S Impact, e.g. Scoping Studies	**Maturity Level 1** Moderate Consequences, Some M&S Impact, e.g. Design Support	**Maturity Level 2** High-Consequence, High M&S Impact, e.g. Qualification Support	**Maturity Level 3** High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification
Representation and Geometric Fidelity What Features are neglected because of simplifications or stylizations?	• Judgment only • Little or no representational or geometric fidelity for the system and boundary conditions (BCs)	• Significant simplification or stylization of the system and BCs • Geometry or representation of major components	• Limited simplification or stylization of major components and BCs • Geometry or representation is well defined for major components and some minor components • Some peer review conducted	• Essentially no simplification or stylization of components in the system and BCs • Geometry or representation of all components is at the detail of "as built," e.g., gaps, material interfaces, fasteners • Independent peer review conducted
Physics and Material Model Fidelity How fundamental are the physics and material models and what is the level of model calibration?	• Judgment only • Model forms are either unknown or fully empirical • Few, if any, physicsinformed models • No coupling of models	• Some models are physics based and are calibrated using data from related systems • Minimal or ad hoc coupling of models	• Physics-based models for all important processes • Significant calibration needed using separate-effects tests (SETs) and integral-effects tests (IETs) • One-way coupling of models • Some peer review conducted	• All models are physics based • Minimal need for calibration using SETs and IETs • Sound physical basis for extrapolation and coupling of models • Full, two-way coupling of models • Independent peer review conducted
Code Verification Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results?	• Judgment only • Minimal testing of any software elements • Little or no SQE procedures specified or followed	• Code is managed by SQE procedures • Unit and regression testing conducted • Some comparisons made with benchmarks	• Some algorithms are tested to determine the observed order of numerical convergence • Some features & capabilities (F&Cs) are tested with benchmark solutions • Some peer review conducted	• All important algorithms are tested to determine the observed order of numerical convergence • All important F&Cs are tested with rigorous benchmark solutions • Independent peer review conducted
Solution Verification Are numerical solution errors and human procedural errors corrupting the simulation results?	• Judgment only • Numerical errors have unknown or large effect on simulation results	• Numerical effects on relevant SRQs are qualitatively estimated • Input/output (I/O) verified only by the analysts	• Numerical effects are quantitatively estimated to be small on some SRQs • I/O independently verified • Some peer review conducted	• Numerical effects are determined to be small on all important SRQs • Important simulations are independently reproduced • Independent peer review conducted
Model Validation How carefully is the accuracy of the simulation and experimental results assessed at various tiers in the validation hierarchy?	• Judgment only • Few, if any, comparisons with measurements from similar systems	• Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest • Large or unknown experimental uncertainties	• Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs • Experimental uncertainties are well characterized for most SETs, but poorly known for IETs • Some peer review conducted	• Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application • Experimental uncertainties are well characterized for all IETs and SETs • Independent peer review conducted
UQ and SA	• Judgment only • Only deterministic	• Aleatory and epistemic (A&E) uncertainties propagated, but	• A&E uncertainties segregated, propagated, and identified in SRQs	• A&E uncertainties comprehensively treated and properly interpreted

How thoroughly are uncertainties and sensitivities characterized and propagated?	analyses are conducted • Uncertainties and sensitivities are not addressed.	without distinction • Informal sensitivity studies conducted • Many strong UQ/SA assumptions made	• Quantitative sensitivity analyses conducted for most parameters • Numerical propagation errors are estimated and their effect known • Some strong assumptions made • Some peer review conducted	• Comprehensive SAs conducted for parameters and models • Numerical propagation errors are demonstrated to be small • No significant UQ/SA assumptions made • Independent peer review conducted

Table 7 PCMM Guidance provided in January 2008

PCMM Practice		Maturity Level 0 Low Consequence, Minimal M&S Impact, e.g. Scoping Studies	Maturity Level 1 Moderate Consequence, Some M&S Impact, e.g. Design Support	Maturity Level 2 High-Consequence, High M&S Impact, e.g. Qualification Support	Maturity Level 3 High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification
Representation and Geometric Fidelity (RGF) Are representation errors corrupting simulation conclusions?	Characterization (how close to as built are you representing the system)	• (unjustified) conceptual abstraction of the whole system	• Significant (unjustified) simplification or stylization of the system at the level of major elements	• Limited (unjustified) simplification or stylization of the system at the level of major and minor elements	• Geometry or representation of all elements is at the detail of "as built"
	Computation Error (what impact does imperfect RGF have on computation results)	• Judgment only, numerical errors introduced because of imperfect RGF not addressed	• Sensitivity to imperfect RGF explored for some System Response Quant. (SRQs)	• Numerical errors estimated for imperfect RGF for relevant SRQs	• Numerical errors for imperfect RGF rigorously quantified for all relevant SRQs
	Verification (is what you represented really what was built)	• RGF not verified, RGF simply used without verification that it represents the actual system as built	• RGF verified only by the analysts	• RGF independently verified	• RGF database formally managed and under configuration control so that information referenced from the database is certified to represent what is current reality
Physics and Material Model Fidelity (PMMF) How science-based and accurate are the physics and material models?	Science basis for models (how science-based are the models)	• Unknown model form: Calculations enabled through the use of "Knobs", which are surrogates for missing or unknown physics	• Empirical model form: Key dependencies derived predominantly from speculation or experimental observations. Parameters in the empirical model	• Physics-informed models: Key dependencies derived from fundamental theory and modified or augmented to account for model discrepancies. The	• Physics-based models: Key dependencies derived from fundamental or well accepted theory without need for modification or augmentation. Model

			calibrated to data	calibration of some model parameters may not be unique (i.e., the calibration of parameters in separate effects physics with data from integral tests)	parameters derived from theory or uniquely calibrated to relevant data.
	Model Accuracy (how accurate are the models)	• Judgment only, or model accuracy not addressed	• Qualitative assessment of model accuracy • Examples: Vugraph norms, tabular comparisons, or curve overlays	• Quantitative assessment of model accuracy *without* assessment of measurement uncertainty	• Quantitative assessment of model accuracy *with* assessment of measurement uncertainty • Explicit recommendation on how to quantify model form discrepancy or uncertainty in the target application
	Extrapolation (what is relevance of the validation database)	• Unknown extrapolation to the application parameter space	• Significant extrapolation to the application parameter space, which lies outside the validation parameter space	• Full or partial interpolation with significant extrapolation to some or all limits of the application parameter space	• Interpolation (i.e., full coverage) in the application parameter space
	Technical review (confirmation that the validation activities are relevant, adequate, and carried out in a quality manner)	• Judgment only, no technical review of the validation evidence	• Informal technical review or technical review from within the project team or stakeholder community only	• Formal technical review by Subject Matter Experts (SMEs) external to the project team or stakeholder community	• Formal technical review by SMEs external to the project team or stakeholder community • Formal technical review SMEs played an oversight and approval role of activities executed to fill important gaps in the validation evidence
Code Verification	Software Quality Engineering practices	• Judgment only, codes informally managed to	• Codes managed to repeatable and	• The SQE process in managed	• The SQE process is optimized

(CVER) **Are software errors or algorithm deficiencies corrupting the simulation results?**	(SQE: how mature are the SQE practices)	SQE practices or no documented SQE process requirements • Software process is characterized as ad hoc, and occasionally even chaotic	defined SQE practices • Repeatable: Basic project management processes are established to track cost, schedule, and functionality. • Defined: The software process for both management and engineering activities is documented, standardized, and integrated into a standard process for the organization and applied in a graded manner.	• Managed: Detailed measures of software process and product quality are collected. Both the software process and products are quantitatively understood and controlled.	• Optimized: Continuous process improvement is enabled by quantitative feedback from the process and from piloting innovative ideas and technologies.
	Software Quality Assessment (SQA: assurance that code development is managed to an appropriate level of process maturity)	• Judgment only, no assessment to SQE practices	• Self assessment and documentation of full or partial compliance to organizational SQE practices by code team • Self-assessments or formal assessments have identified compliance gaps	• Formal assessment and documentation of full compliance to organizational SQE practices by group external to the code development team	• Formal assessment and documentation of compliance to SQE practices and accreditation to an appropriate level of a nationally recognize set of SQE standards (e.g., CMMI, ISO9000, IEEE, etc) by team external to the code development team
	Test coverage (can the user be confident that the code is adequately tested for the intended application)	• Judgment only, minimal testing of any software elements	• Sustained unit and regression testing and/or limited scope Verification Test Suite (VERTS) routinely conducted with 75% coverage • Note: unit and regression problems track code drift and not necessarily code correctness	• Sustained VERTS re-run regularly w 75% F&C coverage and 75% coverage of all 2-way interactions of F&C • VERTS address *convergence* behavior to the *correct answer*	• Sustained VERTS re-run regularly w 75% coverage of F&C and all their interactions (2-way, 3-way, etc)

			• Here, VERTS address *comparison* (not convergence) to the *correct answer* • Coverage: Line, function, or feature and capability (F&C) coverage • Note: Coverage can be referenced to lines, functions, or F&C actually used in a simulation		
	Computation Errors (what is the numerical impact of undetected code or algorithm deficiencies on simulation results)	• Judgment only, numerical errors in computed System Response Quantities (SRQs) resulting from undetected code or algorithm deficiencies not addressed	• Sensitivity to numerical errors in computed SRQs resulting from undetected code or algorithm deficiencies explored for some SRQs	• Numerical errors in computed SRQs resulting from undetected code or algorithm deficiencies estimated for relevant SRQs	• Numerical errors resulting from undetected code or algorithm deficiencies rigorously quantified for all relevant SRQs
Solution Verification (SVER) **Are human procedural errors or numerical solution errors corrupting simulation conclusions?**	Numerical Solution Errors (what is the impact of numerical solution errors on relevant SRQs)	• Judgment only, numerical solution errors not addressed	• Sensitivity to discretization and algorithm parameters explored for some System Response Quantities (SRQs)	• Numerical errors estimated for discretization and algorithm parameters for relevant SRQs	• Numerical errors rigorously quantified for all relevant SRQs
	Input/Output Verification	• Input/output not verified	• Input/output verified only by the analysts	• Input/output data independently verified	• Input/output data independently verified, calculation results reproduced independently
	Technical Review (confirmation that the solution verification activities are relevant, adequate, and carried out in a quality manner)	• Judgment only, no technical review of the solution verification evidence	• Informal technical review or technical review from within the project team or stakeholder community only	• Formal technical review by Subject Matter Experts (SMEs) external to the project team or stakeholder community	• Formal technical review by SMEs external to the project team or stakeholder community • Formal technical review SMEs played an oversight and approval role of solution verification

					activities
Validation (VAL) **How accurate are the integrated physics and material models?**	Validation hierarchy (are you getting the right answer for the right reason) • IETs: Tests designed to look at limited interactions of physics, usually in stylized geometries • SSTs: Tests designed to look at full interaction of physics at the subsystem level(if appropriate) • FSTs: Tests designed to look at the full interaction of physics	• Judgment only, no comparisons of models to IETs, or SSTs, or FSTs	• Limited comparisons of models to validation data • Model comparisons to IETs only, or SSTs only, or FSTs only	• Systematic comparison of models to validation data • Two of the three: adequate IETs, adequate SSTs, or FSTs • Adequate refers to 75% coverage of 2-way interactions of dominant physics or 75% coverage of subsystem interactions	• Extensive comparison of models to validation data • Adequate IETs, and adequate SSTs, and FS tests
	Model Accuracy (how accurate are the models)	• Judgment only, or model accuracy not addressed	• Qualitative assessment of model accuracy	• Quantitative assessment of model accuracy with out assessment measurement uncertainty • Limits of applicability explicitly stated	• Quantitative assessment of model accuracy with measurement uncertainty • Limits of applicability explicitly stated • Explicit recommendation on how to quantify model form discrepancy or uncertainty in the target application
	Extrapolation (what is relevance of the validation database)	• Unknown extrapolation to the application parameter space	• Significant extrapolation to the application parameter space, which lies outside the validation parameter space	• Full or partial interpolation with significant extrapolation to some or all limits of the application parameter space	• Interpolation with full coverage in the application parameter space
	Technical review (confirmation that the validation activities are relevant, adequate, and	• Judgment only, no technical review of the validation evidence	• Informal technical review or technical review from within the project team or	• Formal technical review by Subject Matter Experts (SMEs) external to the project	• Formal technical review by SMEs external to the project team or stakeholder

	carried out in a quality manner)		stakeholder community only	team or stakeholder community	community • Formal technical review SMEs played an oversight and approval role of activities executed to fill important gaps in the validation evidence
Uncertainty Quantification (UQ) **What is the impact of variabilities and uncertainties on system performance and margins?**	Uncertainty Characterization and Interpretation (are uncertainties characterized, propagated, and interpreted in a manner consistent with their nature)	• Judgment only, uncertainties not addressed	• Aleatory and Epistemic (A&E) uncertainties, of potentially comparable magnitude, *not* segregated • A&E uncertainties characterized, propagated, and interpreted in a Bayesian framework	• A&E uncertainties, of potentially comparable magnitude, *are* segregated in their characterization, propagation, and interpretation • Or epistemic uncertainties clearly dominate • Epistemic uncertainties are characterized, propagated, and interpreted in a Bayesian framework	• A&E uncertainties, of potentially comparable magnitude, *are* segregated in their characterization, propagation, and interpretation • Or aleatory *or* epistemic uncertainties clearly dominate • Epistemic uncertainties are characterized, propagated, and interpreted in an "uncertainty preserving" manner
	Sensitivity Analysis (what uncertainties are important	• Judgment only, or the sensitivity of key SRQs to uncertain inputs are not addressed	• Qualitative sensitivity of key SRQs to uncertain inputs are explored with informal "what if" studies	• Sensitivity of key SRQs to uncertain inputs estimated quantitatively	• Sensitivity of key SRQs to uncertain inputs rigorously quantified
	Numerical Propagation Errors (how sensitive are UQ/SA results to numerical propagation errors due to a finite number of simulations)	• Judgment only, numerical (propagation) errors not addressed	• Sensitivity to numerical (propagation) errors estimated for UQ results	• Sensitivity to numerical (propagation) errors estimated for UQ/SA results	• Numerical (propagation) errors rigorously quantified for UQ/SA
	Aggregation of Evidence for Characterization of	• PI aggregates evidence for the characterization of	• PI aggregates evidence for the characterization of	• Semi-formal elicitation process for key uncertainties that	• Formal elicitation process for key uncertainties that

	Uncertainties	important uncertainties • The characterizations and the supporting evidence basis are not documented or disclosed by the PI or simply made up (e.g., material properties arbitrarily varied by +/- 10%) to execute an analysis	important uncertainties • The supporting evidence, aggregations, and characterizations are formally documented	involves a single SME for each uncertainty objectively represents the diverse perspectives of the broader community of SMEs • The supporting evidence, aggregations, and characterizations are formally documented	involves multiple SMEs who collectively represent a broad spectrum of perspectives • The elicitation process, supporting evidence, aggregations, and characterizations are formally documented • The elicitation results take on the stature of "defaults" to be used in future studies involving similar uncertainties
	Completeness (do we cast a broad enough net that all potentially significant sources of uncertainty or error a acknowledged and quantified or otherwise dealt with) Completeness: scenarios, IC/BC parameters, model and material property parameters, physical and statistical model forms (including alternate plausible models), R&G fidelity, undetected deficiencies in code or algorithms, numerical errors due to discretization or algorithm parameters, uncertainty propagation errors	• Very limited number of uncertainties addressed based largely on what is practical. The selected set of uncertainties likely does not envelop the dominate uncertainties	• Limited number of uncertainties addressed based largely on documented reasoning and judgment of analyst. Analyst asserts that all important uncertainties are addressed in the study based on judgment alone.	• Broad treatment of uncertainties. Although not exhaustive, the net is cast wider than what is thought to be important • The broader community has a say in what uncertainties are addressed	• Exhaustive treatment of uncertainties and errors. Broad consensus that dominant uncertainties have been addressed. The analysis speaks way more loudly than judgment in determining what is important • The broader community has a say in what uncertainties are addressed
	Strong Assumptions (do assumptions	• Strong assumptions not addressed or	• Many strong assumptions	• Some strong assumptions	• No strong assumptions

	undermine the rigor or accuracy of UQ/SA results) Strong Assumptions: linearity, dependencies, shape of distributions, consistency of measured property variability and underlying assumptions of physical model, BE+U	acknowledged			
	Technical Review (has the US/SA been executed in a rigorous and correct manner)	• Judgment only, no technical review of the UQ/SA framework, processes, execution, or results	• Informal technical review or technical review from within the project team or stakeholder community only	• Formal technical review by Subject Matter Experts (SMEs) external to the project team or stakeholder community	• Formal technical review by SMEs external to the project team or stakeholder community • Formal technical review SMEs played an oversight and approval role of UQ/SA activities
Documentation and Archiving **Are the work products reproducible?**	Documentation and Archiving (are the work products reproducible)	• Work is not reproducible, even in concept. Documentation does not exist, is inadequate, or is so informal that even the documentation cannot be retrieved on demand	• Work products are reproducible in concept, but not in detail • Documentation is retrievable on demand, but contains only enough detail to reproduce general results.	• Work products nearly reproducible by repeating the process • Detailed documentation of all representations, code versions, processes, models, model inputs, and assumptions at a detail that the work products could, in spirit, be exactly reproduced by *reconstructing* the entire process	• Work products fully reproducible on demand • Detailed documentation of all representations, code versions, processes, models, model inputs, and assumptions at a detail that the work products could, in spirit, be exactly reproduced by reconstructing the entire analysis process • Codes and all inputs files under configuration control and rerunable on

					demand • Original results are under configuration control for comparison to reproduced results

APPENDIX C: SOME VALIDATION TOOLS AND RESOURCES

As the FSM teams prepare their validation plans, we want to make them aware of some of the tools and resources available for the implementation. We also encourage the teams to specifically name appropriate tools for the activities included in their plans.

The DAKOTA (Design Analysis Kit for Optimization and Terascale Applications) software package, developed at Sandia, is a general purpose toolkit for the integration of commercial and in-house simulation capabilities with broad classes of systems analysis tools (REF: Eldred et al). DAKOTA can be used to support the propagation of random model parameter uncertainty through a model to provide estimates of the corresponding uncertainties in the model predictions. DAKOTA can support both first order sensitivity analysis and stratified sampling Monte Carlo analysis (eg., Latin Hypercube Sampling) on computers with parallel architectures.

JMP is a commercial graphical interface based package used to display and explore data. JMP can be used to support the design of experiments as well as to provide space filling designs (such as Latin Hypercube Sampling) for the propagation of random model parameter uncertainty though models. JMP also provides statistical analysis support for data exploration. Another commercial statistical data analysis tool for which Sandia has a site license is Minitab. It is particularly useful for sensitivity studies.

DART provides a number of tools that can be valuable in implementing V&V processes. WISDM is a database containing material properties. CUBIT and SIMBA provide a number of meshing capabilities and the ability to parameterize meshes and analysis code input decks. The APC provides simulation data management and archiving capabilities.

The V&V program uses the RMS system for archiving documentation. The NW SMU is rolling out the RPSS system for product realization.

V&V experts are available to advise and assist in both the development and implementation of V&V plans. Most reside in 1544, 8962, and 8964.

This page intentionally left blank.

DISTRIBUTION

4 Lawrence Livermore National Laboratory
Attn: N. Dunipace (1)
P.O. Box 808, MS L-795
Livermore, CA 94551-0808

2	MS9018	Central Technical Files	8944
2	MS0899	Technical Library	4536

Sandia National Laboratories