Certainty in Stockpile Computing: Recommending a Verification and Validation Program for Scientific Software

James R. Lee

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 under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.
DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.
Certainty in Stockpile Computing:

Recommending a Verification and Validation Program

for Scientific Software

James R. Lee

Applied Physics Center
Sandia National Laboratories
P. O. Box 5800
Albuquerque, NM 87185-1179

Abstract

As computing assumes a more central role in managing the nuclear stockpile, the consequences of an erroneous computer simulation could be severe. Computational failures are common in other endeavors and have caused project failures, significant economic loss, and loss of life. This report examines the causes of software failure and proposes steps to mitigate them. A formal verification and validation program for scientific software is recommended and described.
Acknowledgments

The ideas presented in this report were developed during numerous discussions with dozens of staff members at Sandia who are in the trenches doing scientific computing every day. In particular, I thank Tim Trucano, Bill Oberkampf, Bob Cochran, and Len Lorence for their repeated discussions and suggestions. I am grateful to Jaime Moya, Tom Bickel, Jim Assay, and Wendell Kawahara for their insight into MAVEN, and to Paul Hommert and Bob Thomas for the cooperation and support of the ASCI Applications Program. I thank Gerry Yonas for sponsoring this study. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.
Contents

1. Introduction .................................................................................................................. 4
2. Sources of Computational Failure ............................................................................... 6
3. Principles of Verification and Validation ..................................................................... 12
4. Requirements .............................................................................................................. 14
5. Code Verification ...................................................................................................... 16
6. Model Validation ....................................................................................................... 19
7. Protocols for Stockpile Calculations ......................................................................... 21
8. Conclusions: Managing a V&V Program ................................................................... 23
9. References .................................................................................................................. 26
1. Introduction

The national laboratories increasingly rely on scientific computing to resolve technical issues that arise in managing the nuclear stockpile. This is the natural consequence of two powerful trends, one political and one technological. The first is the evolution of the nuclear weapons complex in the post cold war era. In the absence of a superpower threat that is visibly aggressive and hostile to the United States, there will be pressure to reduce the nuclear inventories and shrink the complex. The public will grow increasingly intolerant of risk, will not allow the resumption of nuclear testing, and will demand reductions in the cost of maintaining a nuclear deterrent. The second trend is the exponential improvement in computing technology. The goal of the Accelerated Strategic Computing Initiative (ASCI) is to further accelerate this development, leading to computers that run at tens to hundreds of teraflops.

The challenges of stockpile management in the twenty-first century will be many, and they are not widely appreciated by those outside the industry. A new weapon probably will not be designed for many years. This will push existing systems far beyond their design lifetimes, well past the limits of our practical engineering experience. The lack of nuclear tests means that system performance must be inferred from tests of the various subsystems and observations of degradation in the nuclear explosive package. Inevitably, subsystem deterioration eventually will force the replacement of components. In many cases, components will be redesigned from scratch because of sunset technologies. Some of the weapon technologies are quite exotic. They were developed specifically to meet the extreme design requirements of nuclear weapons. We will have to deal with the loss of specialized engineering expertise. For many critical components, most of the people who designed our newest system have left the labs or are near retirement. As we do this job, we must be ever mindful of the consequences that could result from errors in judgment regarding nuclear weapons.

The possibility that a revolution in computing will provide easy answers is enticing. Computer modeling is invisible. It doesn’t generate fire, smoke, noxious odors, hazardous waste, or loud noises. Protesters don’t line up at the gate in the morning when you run a
simulation. While the cost certainly is not insignificant, it is cheaper than digging tunnels at the Nevada Test Site and setting off nuclear devices. Virtually any question regarding the stockpile can be investigated on a computer. Aggressive efforts are underway to substantially improve our modeling capability in the areas of primary and secondary design, aging of materials, structural vibration, aerodynamics, fire simulation, shock physics, electromagnetic phenomena, production processes, electrical circuit design, and radiation effects. The only problem is that inferences drawn from computer simulations sometimes are wrong.

The danger that a serious mistake will be made in the design, construction, or handling of a nuclear weapon, based upon an erroneous computer simulation, is very real. In this report, I will examine the root causes that lead to failure in scientific calculations and propose steps to mitigate them. In particular, I will outline the essential elements of a program to verify and validate computer simulations. I will make recommendations for managing such a program in a systematic, comprehensive, and rigorous manner. It is not possible to eliminate all modeling errors and software bugs and to prove beyond all doubt that calculations are accurate. There is a great deal that we can do, however, to manage the risks inherent in computing. The goal is to demonstrate a level of certainty that is consistent with our responsibilities for managing the nuclear stockpile.
2. Sources of Computational Failure

Nearly all computer programs contain large numbers of bugs that, at least occasionally, lead to unintended behavior. Computers have become so ubiquitous that this fact is generally recognized. Most people are accustomed to dealing with fairly frequent manifestations of imperfection in their word processing, spreadsheet, and presentation software. Facing a deadline, they find that fonts are corrupted, tables get reformatted, e-mail from the boss disappears, and the operating system crashes. Life goes on. What is much less generally recognized is that safety-critical software sometimes fails with catastrophic consequences. These are programs that have been written with enormous attention to detail and subjected to extensive testing, and yet, they fail.

Examples of significant computational failures can be found in virtually every area where software has been used in critical applications. Faulty flight-control software has contributed to airplane crashes in the FA-18, the F-16, the Swedish Gripen, the RAF Tornado, the Fokker F.100, and the Airbus A320[1]. The inability of a Patriot missile to intercept the Iraqi Scud missile that killed 28 people during the Gulf War was caused by a floating point error[2]. A similar mistake caused the failure of the European Ariane 5 rocket. Software running on the navigation computer took into account the motion of the rocket swaying in the wind while initialization calculations were completed prior to launch. The calculations continued for a short time during launch, when the much greater horizontal velocity caused a variable to overflow. Detecting a trivial, irrelevant error, the computer shut down, because it had a twin backup. Unfortunately, the backup computer had just failed from the same problem. The result was characterized in French newspapers as a 37-billion franc ($7 billion) fireworks display. This quirk in the algorithm was benign to the Ariane 4, for which the software was developed, and the problem was not caught when the code fragment was reused[3]. People have been killed in accidents caused by industrial control software. In a near miss, a software upgrade caused a robot on an automobile production line to pick up a vehicle and drop it on top of the safety engineer’s hut[4]. Computer errors have caused financial chaos and economic loss. Montgomery Ward was shocked to discover that a data input mistake had caused an entire warehouse full of goods in Redding, California to disappear from its distribution
system. No trucks visited the warehouse to pick up or deliver goods for three years! The warehouse staff faithfully reported to work and was paid during this period[4]. Our own space program has had its share of problems. NASA was inconvenienced when programmers made gravity repulsive (Apollo 11) and forgot that the Earth moves (Gemini V), and the Mariner I mission was destroyed because a handwritten formula was incorrectly transcribed into the radar system software[1]. It could have been worse. Software errors have caused the North American Air Defense Command (NORAD) to go to full alert on at least three separate occasions, once when the Moon popped up over the horizon[4].

Scientific software presents several challenges beyond those that are common to all code development. The underlying phenomena usually are quite complex, as are the mathematical and numerical methods used to describe them. Applications frequently push the limits of available computer hardware, and new computing environments take time to become stable. Calculations depend upon the accuracy of input data. Scientific software is more susceptible than most to inappropriate use by those who are unfamiliar with the physical models and numerical methods. Hatton has published an extensive analysis of defects in scientific software from more than a dozen technical fields[5]. He concludes that scientific software typically is riddled with defects, and he suggests that calculations should be treated with severe skepticism.

The computing environment

The first documented bug was an unfortunate moth in Grace Hopper’s Mark I computer[6]. Ever since, a programmer’s first impulse has been to suspect the inner wiring of the computer as the cause of unexpected behavior in a program. On rare occasions, this is justified. The most notorious recent example was in the Intel Pentium Processor, in which the table used by the divide routine was not completely initialized. There are many other examples. The IBM System 360 did not correctly handle guard bits, which occasionally caused mistakes in subtraction[7]. The Intel 486 had flaws in the trigonometric functions[1]. Other computers have had problems with multiplication (returning 0.00482 for y = x*x when x = 6 × 10^-6) and with the square function (returning \( \text{sqr}(2.0) = 2 \))[4]. Intel devotes prodigious effort to validating the designs of their microprocessors, at a cost of several hundred million dollars per year. However, the complexity of modern processors has reached a level where complete validation of the design is essentially impossible. A totally bug-free processor may never be manufactured again. More than 50 errata have been found in the Pentium II since it was shipped to customers[8].

The development of massively parallel computers places extreme demands on the reliability of the individual components. Even if the mean time between failures (MTBF) of a processor-memory module is as long as 20 years, a computer constructed of 9000 modules can have a MTBF of no longer than about 20 hours. It easily could be shorter if hardware failures are dominated by failures of the interconnections between modules.
This could limit the effective computing capability more than processor speed or memory size[9].

A crucial aspect of the computing environment is the stability of the compiler. This is often a problem when programs contain statements, perhaps inadvertently, that are formally unspecified, undefined, or implementation dependent in the language standard. Code that does not manifest itself in a computational error in one situation can do so when the compiler is changed or different compile options are selected. Compilers for massively parallel computers are especially complex and relatively new, so they tend to be updated frequently.

Another issue is the reliability of function libraries. In general, routines found in standard libraries are very robust, but if there is an error it can affect many applications. These are seldom checked thoroughly by the end user.

**Coding errors**

The classic form of computer bug is the coding error, where unintended behavior is generated by a mistake in the logic or syntax of the program. There are many ways to make such errors, and they are much easier to create than to find.

Computer languages are full of features that can trap the unwary. For instance, C contains 22 constructions for which the program behavior is unspecified, 97 for which the behavior is undefined, 76 for which the behavior is implementation dependent, and 6 for which the behavior is locale dependent[4]. An example is the expression

\[ x = (i++) + a[i]; \]

Whether \( i \) is incremented before or after the \( i \)th element of \( a \) is selected is unspecified and will depend on the compiler. Other constructions are perfectly legal and well defined, but they are prone to error. Perhaps the most common C bug is the test

```c
if( i = j )
{
    do_something();
}
```

where the assignment operator \( = \) is used mistakenly instead of the equality operator \( == \). The fact that everyone knows this is a common error doesn’t stop it from happening.

Sometimes the program syntax is simply so complex that it invites error. The class hierarchy in C++ programs can become forbidding, especially when features such as multiple inheritance are used. What is perfectly obvious to the experienced programmer
creating the class structure may seem decidedly less obvious to her replacement, who is trying to modify the program and learn C++ concurrently.

Coding errors become more likely as the complexity of the structure and control logic of a program increases. Various metrics have been proposed for describing complexity. These include the number of decision points in a subroutine, the deepest level of nesting, the number of external variables in a subroutine, the number of static paths through a program, and the fan-in/fan-out product. While there is no consensus on what values signal danger, astoundingly high scores have been observed in commercial scientific software[5].

Formal programming standards, style guides, code analysis tools to detect static and dynamic faults, and independent software inspection are not widely used in the development of scientific software. The atmosphere for code teams working at the forefront of science can be tremendously exciting. Highly productive, clever teams with a “can do” attitude are the norm. Unfortunately, code development characterized by prodigious effort under deadline, caffeine consumption, and sleep deprivation is unlikely to produce bug-free software.

**Mathematical and numerical complexity**

If there is one characteristic that makes scientific calculations especially problematic, it is the exceptional complexity of the mathematical techniques and numerical methods that are embodied in the software. This is unavoidable. The only question is how to systematically recognize and mitigate the risks.

While each application may have a unique set of difficulties, they generally involve issues such as the numerical stability and convergence of iterative techniques, discretization error, complicated boundary conditions, estimation of uncertainties, and the accumulation of floating-point errors. Sometimes algorithms that improve computational efficiency make uncertainty estimation more difficult. Biasing techniques in Monte Carlo codes can suppress high-impact but low-probability histories called zingers, and convergence theorems may not exist for non-uniform grids. Oberkampf and Blottner have described some of the subtleties in applying Richardson’s extrapolation method to problems in computational fluid dynamics. They suggest that in many simulations, convergence is much slower than expected[10]. While great advances in computing can be enabled by the development of new mathematical and numerical techniques, the limitations and pitfalls of new methods are discovered over time. This is particularly relevant now, as massively parallel computers become available. The standard algorithms that have been optimized for serial computation frequently don’t run efficiently on the new platforms. New algorithms designed expressly for parallel processing are being researched intensively.
Code developers are not always expert in the techniques embodied in their software. If the project is large enough, there can be a separation of function between the development of numerical techniques and coding. Even in a small development team, subroutines developed by others may be borrowed. If the code developers are not familiar with the limitations of the algorithms, this can lead to trouble. The borrowed software also may contain outright errors. Code swapping propagates errors from project to project, and it limits the utility of code-to-code comparisons. Of course, reusing code also can enhance software reliability. If the software modules have been developed carefully and tested rigorously, reuse can be safer then redevelopment. A decision should be made mindfully after assessing the scope of the algorithms, the quality of the code, and the risks. The history of the borrowed code, the testing procedures that it has undergone, and any known problems or limitations should be understood fully and documented.

**Nature is complicated**

Except in rare instances, computer simulations are merely approximations of nature. Often, these approximations are crude. In developing a code, judgments are made about which physical phenomena to include and which to ignore. In the best case, an informed decision to exclude an effect is made after it is demonstrated to be unimportant to the intended application. In the worst case, its importance is unknown, but inclusion would make the problem intractable. Obviously, there is no accounting for effects that are unknown. Even when a phenomenon is included, a complete mathematical description may not be known. The best available model may have known deficiencies or limited applicability. Numerical accuracy in the solver will not help if the underlying equation is wrong.

In a typical application of simulation software, problem simplification is essential. The software user must systematically eliminate unnecessary detail to keep from overwhelming the capacity of the computer. A 100-teraflop computer will not render this step obsolete. If the user is not intimately familiar with the physical models and numerical algorithms in the code, essential detail may be discarded.

Simulations often use empirical models that depend on measured parameters. These include material properties and particle cross sections. In this case, the simulations can never be better than the database of parameters. A parameter may be valid only over some physical range, beyond which it changes value or the underlying model becomes invalid. For example, consider the elastic properties of a material. When the material becomes plastic, the coefficient of elasticity does not merely change. A completely different description of the material behavior is required. Extending a simulation beyond the regime where the models are valid can lead to results that are grossly incorrect. There also may be some intrinsic variability in a parameter, dependent upon manufacturing processes or sample histories that are unknown. Insuring the accuracy of data and the applicability of empirical models are daunting challenges in computer simulation.
Another is quantifying the impact of uncertainties in the data, especially when there are large numbers of parameters.

Sometimes nature is more than merely complicated. It can be chaotic\(^1\). The mathematical manifestation of chaos is instability in the differential equations that describe nature. Small changes in input parameters can cause large swings in results. Chaos is ubiquitous in nature. Normally, one of the design goals of an engineered system is to prevent chaotic behavior, but it cannot be avoided without understanding. The possibility of chaos presents two problems in computer modeling. The first is the chance that a non-linear phenomenon that drives instability has been discarded in simplifying the mathematical equations imbedded in the simulations. Non-linear terms that may be small corrections for one choice of input parameters can dominate the problem for other choices, and so these terms might mistakenly be thought insignificant and be neglected. The second problem is that an analyst may not be aware that the input variables lie in a chaotic region, even when the essential physics is included in the model. Unless the appropriate variations are made in the sensitive parameters, there may be no obvious clue to this behavior.

**Inappropriate use**

Distance, time, education, experience, organization, computing environment, retirement, and death may separate software users from developers. The result is simulation software being applied to problems that the developers never intended to support. The worst problems occur when an inexperienced analyst uses software that he doesn’t understand. Unless there is detailed oversight and review by experts who are thoroughly familiar with both the codes and the applications, there are many opportunities for egregious error. Essential details are ignored, parameters explore uncharted regions, algorithms compute but never converge, and long-dormant coding errors awaken with devastating, albeit unnoticed, effect. The results can seem entirely reasonable, and yet be utterly wrong.

This is another distinguishing feature of scientific simulation software. Bugs certainly exist in all types of codes, but when errors are generated in system software or typical commercial applications, the fact that something is wrong usually is fairly obvious. With scientific simulations, severe errors can be insidiously difficult to detect. The code can execute without crashing, but simply give answers that are numerically incorrect. As long as curves are smooth and plots look plausible, there may be no obvious warning sign that something is amiss.
3. Principles of Verification and Validation

A formal verification and validation (V&V) program is fundamental to successfully realizing the ASCI vision, which is to “shift promptly from nuclear test-based methods to compute-based methods”[12]. There is no other way to demonstrate convincingly that high-consequence software has been developed and applied without significant error. The ASCI goal is to “create the leading-edge computational modeling and simulation capabilities that are essential for maintaining the safety, reliability, and performance of the US nuclear stockpile and reducing the nuclear danger” (italics added). A computational failure that would compromise the safety, reliability, or performance of a nuclear weapon or increase the nuclear danger obviously is of high consequence.

My recommendation is that every code development project sponsored by ASCI shall maintain detailed, written V&V plans. These plans shall be updated frequently and subjected to rigorous peer review. There are four essential elements that must be addressed thoroughly:

1. Requirements
2. Code verification
3. Model validation
4. Protocols for stockpile computations

These will be described in the following sections of this report.

Clear accountability must be established for delivering calculations that satisfy the requirements and are demonstrably accurate. This accountability extends to all aspects of V&V. In most cases, the logical choice for the accountable individual will be the project leader of the code development team.

While full accountability rests individually with each development project, the V&V plans should build upon global efforts to achieve goals common to all code development. These would include establishing stable computing environments, maintaining
thoroughly tested and documented software libraries, promoting standard code development tools, configuration management tools, code architectures, and interfaces between codes, and developing programming standards and style guides. A carefully focused effort to leverage resources will be critical to a successful V&V program.

Code development projects differ greatly in scope, complexity, and purpose. The V&V program should be flexible enough to manage sensibly the whole spectrum of activities. Flexible does not mean out of date, undocumented, or not reviewed.
4. Requirements

The requirements for software development derive from our fundamental responsibilities in managing the nuclear stockpile. Code requirements must be tied explicitly and accurately to specific stockpile activities. If requirements are understated, software will not include the necessary features, be developed with the required rigor, or be ready in time to support the stockpile. If requirements are overstated, the resources needed to implement a comprehensive V&V program will not be bounded sufficiently. A V&V program cannot be managed rationally until the requirements have been thoughtfully negotiated and documented.

The questions that should be addressed in developing software requirements include the following:

1. Who are the customers?
2. What are the essential applications that must be supported? The truly critical applications should be differentiated from those that are merely important.
3. When must the software be ready to support key milestones in the weapons program?
4. What is the impact of having this software? What development or testing activities that were done previously can now be avoided?
5. What are the major technical risks in developing this software?
6. How accurate does this software have to be? What are the consequences if calculations done with this software are erroneous?
7. Which computer platforms must be supported?
8. Who will use the software? Will the code developers or those expert in the physical algorithms and numerical techniques be setting up all critical applications, or will this be done by novices? Developing bulletproof software for use by non-experts is extraordinarily challenging.
9. What other software is needed to support the critical applications? What interfaces exist with other codes?
The detail with which these questions may be answered will vary depending on the project. In many software development projects, the answers evolve as the scope is broadened to include more challenging or different applications. Considerable effort is warranted in developing the requirements in as much detail as possible. This will pay off in a more tightly focused and efficient V&V program.
5. Code Verification

Code verification means insuring that software correctly solves the equations that form the mathematical representation of nature in the program. Whether these equations constitute an accurate and complete description of nature is the subject of model validation. The V&V plan should describe the steps that are taken to minimize the introduction of errors, specify procedures for identifying and correcting mistakes, document software features that monitor the performance and convergence of the algorithms, discuss the quantification of uncertainty, and describe tests that are done to assess the reliability of the finished product. The scope of the verification effort should be consistent with the documented requirements.

Processes for code development

Finding and fixing errors is difficult, and so anything that is done to minimize bugs in the first place has great value. While there are some unique difficulties in developing scientific software, there are also concerns that have very much in common with software development in general. Substantial work has been done to identify processes for developing code that are seen to improve the quality of the product.

One of the most comprehensive models for software development is the Capability Maturity Model (CMM) from the Software Engineering Institute (SEI) at Carnegie Mellon University[13,14]. This is a five-level model, ranging from level 1 (initial) to level 5 (optimizing). The operational level of an existing software development process is determined with questionnaires, which are filled out independently by the development staff and management. The model presents a prioritized roadmap for incrementally improving the process. At each level, there are specific objectives that must be met to advance to the next level. There is a balanced approach in the areas of standards and procedures, organization, tools and technology, and process metrics. This is key, because optimizing a deficient process can be counterproductive. A strategy that is designed to address the challenges of a level-3 process, for example, may bring an immature process operating at level 1 (characterized as ad hoc and chaotic) to a grinding halt. The dominant
challenges facing a level-1 organization are management deficiencies. These must be solved to provide a solid foundation for further growth.

There are many specific processes that can improve software quality, and they should be considered carefully in constructing a V&V plan. A programming standard may be developed to document the reasons for choosing a particular programming language, to develop a subset of that language that avoids unsafe features, to set limits on the complexity of the code elements, and to isolate platform dependencies with function wrappers to enhance portability. A style guide may address conventions for naming identifiers, guidelines for comment statements, indentation standards, and other issues that tend to be more emotionally charged than those dealt with in the programming standard. Formalized code inspections are widely used in developing safety-related software. These are most valuable when human inspection is used in conjunction with software tools that automatically detect as many defects as possible. Code inspection procedures should guarantee adherence to the programming standard. Unit testing and regression testing may be done to specific standards for code coverage and the use of particular fault-detection tools. In some cases, individuals who are not part of the development team do the testing. Procedures for tracking bugs will keep them in sight until they are resolved. Change control formalizes the process of proposing improvements to software, evaluating them for technical feasibility, validating the proposal against written requirements, changing code components or adding new components, testing the code, documenting the changes, updating the production libraries, and releasing the software. Configuration control manages different versions of the software and allows calculations to be repeated much later in exact detail. This can be particularly important in managing the stockpile, where decades may elapse between modifications of a weapon. Thorough documentation is critical.

A V&V plan should describe the code development environment. It should specify the development platforms, the programming languages, the compilers, the code development tools, and the software libraries. The code development processes should be described in detail. Adherence to the processes should be transparent and permit auditing. There should be procedures defined for modifying the code development process.

It may be entirely appropriate that code development processes vary widely from one project to another. Judgement is needed in assessing which software-engineering techniques are most effective, and these may be project dependent. What is essential is that mindful choices be made that are consistent with the requirements, the choices are documented, they are followed, they are defended in peer review, and accountability is clear.

**Numerical methods**

The sophisticated mathematical models and numerical methods used in scientific software make code verification intrinsically difficult. It also is difficult to make
generalizations about how to do it well. There is enormous variety in the problems being considered and the approaches taken.

At the very least, the approach must be clear. The equations being solved and the algorithms used to solve them should be documented explicitly. This includes any auxiliary equations, such as equations of state, cross-section models, turbulence models, interpolation methods used in look-up tables, convergence criteria, and anything else that can influence the results.

Performance metrics should be built into the algorithms wherever possible. Quantities that are conserved exactly in nature, such as energy, momentum, and charge, are only conserved approximately in many simulations. These must be tracked during calculations. In iterative algorithms, parameters that indicate the degree of convergence must be tracked. In Monte Carlo codes, the statistical uncertainties must be tracked. Software should alert users when parameters in a simulation exceed validated limits. In short, any significant source of numerical error should be monitored during the calculation, including those caused by the auxiliary equations. Estimates must be made of the ensuing uncertainty in the final results of the simulation. This includes quantifying discretization errors caused by the spatial grid resolution, the magnitude of the time step, or linear approximations to differential equations.

It is a major challenge to perform integrated tests of the equation solvers, where many algorithms interact in a complicated way and the problem geometry is complex. When analytic solutions exist, they can be extremely valuable. Code-to-code comparisons also are valuable, especially when another code contains the same physical models but uses different numerical techniques. The widespread practice of borrowing software must be considered in evaluating the results of code-to-code comparisons. There is a high probability that two codes that solve the same problems have identical code incorporated in some of the algorithms. Software coverage and relevance to the critical applications are serious issues in developing a set of test problems.
6. Model Validation

Model validation means demonstrating that the equations used to describe nature are sufficiently complete and accurate to support the intended applications. A V&V plan should document the physical basis of the mathematical description, identify parameters which must be experimentally determined, assess the accuracy and integrity of existing databases, identify the need for additional or confirming data to support the requirements, and document the systematic comparison of the code with experiments at all levels of integration.

The physical phenomena that are included in the model should be described in detail. Fundamental assumptions that are implicit in the model should be documented. Any simplifications, approximations, or known limits of applicability should be described. Related phenomena that are not included also should be described, along with the justification for their exclusion.

If calculations depend on measured quantities, it must be demonstrated that these are known well enough to support the applications. If data already exists, their accuracy must be assessed. If not available, a plan to acquire necessary data must be developed. The range over which parameters are valid must be determined. Any code calibration that must be done experimentally in the normal course of using the code should be described.

There should be a comprehensive validation of the code with experiments. The first step is to isolate each phenomenon in turn and do simple experiments that confirm the basic models. These experiments should be very simple to validate the physics with few complicating factors. The next step is to incrementally increase the experimental complexity. Experiments where multiple phenomena are important, or which stress the two and three-dimensional capabilities of the code, should be done. Finally, the code must be tested in highly integrated experiments that are comparable in complexity to the critical applications. A V&V plan should describe experiments that have been done as well as further experimentation needed to acquire confidence. The relevance of the experiments to the intended applications must be established.
Resources necessary to implement the code validation should be identified. This includes experimental facilities, specialized diagnostics and test equipment, and experienced personnel. Estimates should be made of the funding and time needed to carry out the experiments. Program partners that are expected to sponsor the code validation should be identified.

Uncertainty in the measured parameters, limitations in the code calibration procedures, and simplifications and approximations in the physical models all will lead to uncertainty in calculations. A V&V plan should describe how these uncertainties are estimated.
7. Protocols for Stockpile Calculations

Both the promises and the dangers of code development are realized when software is used to perform stockpile calculations. This is when computer simulations provide the basis for decisions about nuclear weapons that either enhance or compromise the safety and security of the nation. These simulations should not be done in a haphazard, *ad hoc* fashion.

The purpose of defining protocols for stockpile calculations is to insure that software is used properly in making decisions about nuclear weapons. Even codes that have been verified and validated thoroughly may be misused. All simulations have limitations, and these must be understood and considered.

Appropriate circumstances for doing stockpile calculations should be designated. The user must have the necessary expertise. Test problems must be run successfully on the computing platform. Code verification and validation must demonstrate that the application is within the scope of the software. In some cases, further experiments will be necessary to address validation issues that are specific to the problem at hand.

The entire simulation process should be documented in reproducible detail. This includes the version of the code used in the calculation, the computing platform, and the compiler. In most simulations the analyst makes a series of problem simplifications. These should be documented and justified. In some cases, the calculations must be linked to configuration control in the production of weapon components. This prevents production changes from invalidating analyses without due consideration.

Extreme care is warranted for the most critical applications. Various measures should be considered to reduce risk:

1. Use codes that have been formally accredited to be suitable for particular applications.
2. Formally accredit analysts to insure that they have the necessary expertise to use these tools for particular applications.

3. Perform independent simulations by separate teams of analysts. This provides a check on the problem simplification and engineering judgment that is inherent in computer simulations. There is additional benefit if the simulations are done with different codes or separate computing environments.

4. Scrutinize the simulations with peer review.

5. Audit the simulation process to guarantee that test problems are successfully completed, configuration control is maintained, necessary information is archived, and documentation is completed.

A V&V plan should describe and justify the protocols for doing stockpile calculations. The scope and formality of the protocols must be consistent with the documented requirements.
8. Conclusions: Managing a V&V Program

Demonstrating that calculations are correct is difficult. The consequences of misleading simulations could be severe. There are many opportunities for error, and much must be done to gain confidence. A V&V program is unlikely to succeed unless it has high priority and receives sustained attention.

Establish accountability

The first step is to establish accountability unambiguously. This is not as trivial as it might seem, since there are numerous players in a comprehensive V&V program. They include the software developers, the analysts who use the codes for specific applications, the customers who are responsible for the applications, scientists who do validation experiments, those who provide experimental facilities, and the numerous managers who fund various activities or provide people to implement the program. Splitting accountability is a bad idea. It guarantees that something important will not get done.

In most cases, the project leader of the code development team is in the best position to succeed in the long run. Knowledge and continuity are the deciding factors. This individual should know as much as anyone about the details of the software implementation and will have ready access to those who develop algorithms or run applications. Moreover, a project leader tends to be involved with a particular code for an extended period of time. Management structures, funding chains, and weapon programs with milestones are shifting sand, but there are project leaders who have experience exceeding three decades in developing a particular code or class of codes. Of course, accountability passes through the management chain of the project leader. The management responsibility can move as organization charts evolve, without undue disruption.

The scope of accountability must be defined carefully, since the project leaders usually do not control all of the resources needed to implement all aspects of a comprehensive V&V plan. However, it is reasonable to hold them accountable for the following:
1. Negotiate the requirements for the software.
2. Develop and implement the code verification plan.
3. Develop the code validation plan.
4. Become an advocate for the code validation experiments. Negotiate partnerships with programs that have a vested interest in the software to provide experimental resources. Bring shortfalls to the attention of management.
5. Develop the protocols for stockpile calculations.
6. Write the V&V plan, and keep it up to date.
7. Lead the defense of the V&V plan and its implementation in periodic peer reviews.
8. Demonstrate the accuracy and quantify the limitations of calculations that are done to support the nuclear stockpile.

These all can be done as an integral part of developing software. They should be supported by the same source of funds used to develop the code. This will increase the cost of developing code, probably substantially. However, I believe that this will be more efficient and more successful than attempting to graft a V&V effort onto code development, funded separately and managed externally.

Peer review

Rigorous peer review is a critical element of a V&V program. No other process is as efficient in exposing gaps in planning or deficiencies in execution. All aspects of V&V should be scrutinized in peer review, from defining requirements to stockpile computations.

Members of a peer review panel should encompass a breadth of experience. They should include experts in the physics incorporated in the code, experts in the applications of the code, customers, program managers, program partners, and those with experience in high-consequence computing. The perspective of people from outside Sandia can provide valuable insight in a peer review.

Metrics

There are several metrics that could be used by executive management to assess progress in a V&V program. Initially, these could include the fraction of code development projects for which V&V accountability has been established, the fraction that have written V&V plans, and the fraction that have been peer reviewed. Milestones could be defined for increasing these fractions to 100% within a defined period.
Another metric could be the establishment of program partnerships to provide experimental resources. This is an area where management involvement is likely to be especially helpful.

**Understand the risks**

My expectation is that an aggressive V&V program can make a dramatic difference in refining software requirements, improving software quality, developing better physical models, and minimizing application errors. I do not expect that any V&V program can make the risk of computational failure zero.

The scientists who are creating the revolution in modeling and simulation understand this viscerally. Managers have to work at it. A V&V program can reduce and quantify computational uncertainties, but it cannot make them disappear. Even as we celebrate the new computational tools, some small degree of humility is warranted as we apply them to the sobering task of managing the nuclear weapons stockpile.

There will be some instances where we are forced to make stockpile decisions based largely or entirely on computations, with little or no supporting test data. These cases demand special attention and should be the highest priority of the V&V program. There are many other problems that can be addressed through a variety of computational and experimental approaches. We should avoid becoming completely dependent upon calculations, except under the most compelling circumstances and after exhaustive consideration of the risks and potential consequences.
9. References

<table>
<thead>
<tr>
<th>Distribution:</th>
<th>R. J. Eagan, 1000</th>
<th>M. C. Hinckley, 2254</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-MS-0513</td>
<td>T. J. Allard, 1002</td>
<td>C. L. Knapp, 2262</td>
</tr>
<tr>
<td>I-MS-1427</td>
<td>S. T. Picraux, 1100</td>
<td>D. J. Beyer, 2263</td>
</tr>
<tr>
<td>I-MS-1056</td>
<td>B. L. Doyle, 1111</td>
<td>R. D. Monson, 2265</td>
</tr>
<tr>
<td>I-MS-0960</td>
<td>J. Q. Searcy, 1400</td>
<td>D. P. Green, 2265-1</td>
</tr>
<tr>
<td>I-MS-0953</td>
<td>W. E. Alzheimer, 1500</td>
<td>R. G. Miller, 2266</td>
</tr>
<tr>
<td>I-MS-1452</td>
<td>J. G. Harlan, 1500</td>
<td>T. R. Harrison, 2271</td>
</tr>
<tr>
<td>I-MS-0953</td>
<td>T. J. Cutchen, 1501</td>
<td>V. E. Byfield, 2282</td>
</tr>
<tr>
<td>I-MS-0613</td>
<td>D. H. Doughty, 1521</td>
<td>W. D. Williams, 2300</td>
</tr>
<tr>
<td>I-MS-0614</td>
<td>D. E. Mitchell, 1522</td>
<td>T. B. Lucy, 2301</td>
</tr>
<tr>
<td>I-MS-0614</td>
<td>K. R. Grothaus, 1523</td>
<td>D. L. Eilers, 2304</td>
</tr>
<tr>
<td>I-MS-0613</td>
<td>P. C. Butler, 1525</td>
<td>M. J. Mundt, 2314</td>
</tr>
<tr>
<td>I-MS-0665</td>
<td>L. J. Azvedo, 1541</td>
<td>S. M. Kohler, 2334</td>
</tr>
<tr>
<td>I-MS-0665</td>
<td>R. L. Pettit, 1542</td>
<td>B. D. Shafer, 2335</td>
</tr>
<tr>
<td>I-MS-0665</td>
<td>J. M. Simons, 1544</td>
<td>P. A. Molley, 2336</td>
</tr>
<tr>
<td>I-MS-1452</td>
<td>J. A. Merson, 1552</td>
<td>M. K. Lau, 2338</td>
</tr>
<tr>
<td>I-MS-1453</td>
<td>F. H. Braaten, Jr., 1553</td>
<td>W. H. Schaedla, 2343</td>
</tr>
<tr>
<td>I-MS-1454</td>
<td>L. L. Bonzon, 1554</td>
<td>D. R. Weiss, 2343</td>
</tr>
<tr>
<td>I-MS-0515</td>
<td>F. M. Bacon, 1561</td>
<td>R. M. Axline, 2344</td>
</tr>
<tr>
<td>I-MS-0521</td>
<td>T. J. Young, 1567</td>
<td>B. C. Walker, 2345</td>
</tr>
<tr>
<td>I-MS-1079</td>
<td>A. D. Romig, Jr., 1700</td>
<td>M. B. Murphy, 2346</td>
</tr>
<tr>
<td>I-MS-1083</td>
<td>P. S. Winokur, 1727</td>
<td>D. J. Rigali, 2400</td>
</tr>
<tr>
<td>I-MS-0525</td>
<td>P. V. Plunkett, 1734</td>
<td>W. E. Williamson, 2414</td>
</tr>
<tr>
<td>I-MS-1434</td>
<td>G. E. Pike, 1800</td>
<td>C. M. Hart, 2500</td>
</tr>
<tr>
<td>I-MS-0333</td>
<td>W. C. Moffatt, 1803</td>
<td>J. H. Stichman, 2600</td>
</tr>
<tr>
<td>I-MS-0457</td>
<td>G. N. Beeler, 2000</td>
<td>D. J. Gangel, 2601</td>
</tr>
<tr>
<td>I-MS-0457</td>
<td>W. J. Tedeschi, 2001</td>
<td>R. J. Longoria, 2611</td>
</tr>
<tr>
<td>I-MS-0429</td>
<td>R. D. Andreas, 2100</td>
<td>T. R. Perea, 2612</td>
</tr>
<tr>
<td>I-MS-0435</td>
<td>F. F. Dean, 2102</td>
<td>L. J. Dalton, 2615</td>
</tr>
<tr>
<td>I-MS-0427</td>
<td>W. R. Reynolds, 2103</td>
<td>D. W. Plummer, 2643</td>
</tr>
<tr>
<td>I-MS-0453</td>
<td>D. L. McCoy, 2104</td>
<td>R. E. Kreutzfeld, 2645</td>
</tr>
<tr>
<td>I-MS-0475</td>
<td>R. C. Hartwig, 2105</td>
<td>V. P. Salazar, 2663</td>
</tr>
<tr>
<td>I-MS-0447</td>
<td>J. O. Harrison, 2111</td>
<td>D. E. Ryerson, 2664</td>
</tr>
<tr>
<td>I-MS-0487</td>
<td>J. D. Mangum, 2121</td>
<td>M. E. Partridge, 2665</td>
</tr>
<tr>
<td>I-MS-0486</td>
<td>M. E. Bleck, 2123</td>
<td>J. W. Hole, 2671</td>
</tr>
<tr>
<td>I-MS-0436</td>
<td>G. L. Maxam, 2147</td>
<td>J. A. Wilder, Jr., 2674</td>
</tr>
<tr>
<td>I-MS-0479</td>
<td>P. A. Sena, 2151</td>
<td>S. C. Roehrig, 2691</td>
</tr>
<tr>
<td>I-MS-0482</td>
<td>A. B. Cox, 2161</td>
<td>R. J. Detry, 2900</td>
</tr>
<tr>
<td>I-MS-0483</td>
<td>T. D. Hernandez, 2165</td>
<td>R. L. Hagengruber, 5000</td>
</tr>
<tr>
<td>I-MS-0481</td>
<td>M. A. Rosenthal, 2167</td>
<td>W. M. Knauf, 5001</td>
</tr>
<tr>
<td>I-MS-0481</td>
<td>K. D. Meeks, 2168</td>
<td>J. F. Ney, 5100</td>
</tr>
<tr>
<td>I-MS-9005</td>
<td>J. B. Wright, 2200</td>
<td>J. B. Woodard, 5131</td>
</tr>
<tr>
<td>I-MS-9032</td>
<td>M. A. Dremalas, 2211</td>
<td>C. G. Thomas, 5132</td>
</tr>
<tr>
<td>I-MS-9014</td>
<td>W. G. Wilson, 2221</td>
<td>J. L. Moya, 9105</td>
</tr>
<tr>
<td>I-MS-1223</td>
<td>M. W. Callahan, 5133</td>
<td>J. R. Garcia, 9106</td>
</tr>
<tr>
<td>I-MS-1393</td>
<td>J. A. Ang, 5134</td>
<td>W. Hermima, 9111</td>
</tr>
<tr>
<td>I-MS-1237</td>
<td>J. L. Meinhardt, 5134</td>
<td>D. K. Gartling, 9111</td>
</tr>
<tr>
<td>I-MS-1237</td>
<td>R. K. Wilson, 5135</td>
<td>P. R. Schunk, 9111</td>
</tr>
<tr>
<td>I-MS-1237</td>
<td>A. K. Hays, 5136</td>
<td>A. C. Ratzel, 9112</td>
</tr>
<tr>
<td>I-MS-1223</td>
<td>D. M. Dominguez, 5138</td>
<td>R. J. Cochran, 9113</td>
</tr>
<tr>
<td>I-MS-1233</td>
<td>W. I. Klein, 5139</td>
<td>S. N. Kempka, 9113</td>
</tr>
<tr>
<td>I-MS-9003</td>
<td>D. L. Crawford, 5200</td>
<td>B. F. Blackwell, 9113</td>
</tr>
</tbody>
</table>
1-MS-0820  P. Yarrington, 9232
1-MS-0439  D. R. Martinez, 9234
1-MS-0439  K. F. Alvin, 9234
1-MS-0439  M. S. Eldred, 9234
1-MS-0439  J. R. Red-Horse, 9234
1-MS-1155  J. Polito, Jr., 9300
1-MS-1155  W. Beezhold, 9300
1-MS-1165  D. M. Rondeau, 9301
1-MS-1165  G. A. Zawadzkas, 9302
1-MS-1170  M. A. Robertson, 9309
1-MS-1168  W. A. Boyer, 9321
1-MS-1156  R. Tachau, 9322
1-MS-1392  V. Gabbard, 9324
1-MS-1153  M. T. Buttram, 9330
1-MS-1166  G. J. Scrivner, 9332
1-MS-1166  C. Drumm, 9332
1-MS-1166  W. C. Fan, 9332
30-MS-1179  J. R. Lee, 9341
1-MS-1179  W. P. Ballard, 9341
1-MS-1179  L. J. Lorence, 9341
1-MS-1179  J. L. Powell, 9341
1-MS-1179  J. A. Halbleib, 9341
1-MS-1179  G. D. Valdez, 9341
1-MS-1179  D. E. Beutler, 9341
1-MS-1106  A. W. Sharpe, 9342
1-MS-1159  M. A. Hedemann, 9344
1-MS-1187  J. L. Porter, 9573
1-MS-1186  T. A. Mehlhorn, 9574
1-MS-1196  R. J. Leeper, 9577
1-MS-1178  J. J. Ramirez, 9580
1-MS-1178  P. S. Raglin, 9581
1-MS-9999  V. S. Roberts, 9591
1-MS-1002  P. J. Eicker, 9600
1-MS-1005  R. W. Shaum, 9601
1-MS-1003  R. D. Robinett, III, 9611
1-MS-1008  J. C. Fahrenholtz, 9621
1-MS-1010  M. E. Olson, 9622
1-MS-1004  R. W. Harrigan, 9623
1-MS-1125  A. K. Miller, 9652
1-MS-1006  P. Garcia, 9671
1-MS-1007  A. T. Jones, 9672
1-MS-0428  W. C. Nickell, 12300
1-MS-0829  K. V. Diegert, 12303
1-MS-0829  B. M. Rutherford, 12303
1-MS-0635  G. W. Kuswa, 12365
1-MS-0867  J. L. Martinez, 14000
1-MS-0868  K. G. McCaughey, 14400
1-MS-1179  R. P. Kensek, 9341
1-MS-9018  Central Technical Files, 8940-2
2-MS-0899  Technical Library, 4916
1-MS-1167  E. F. Hartman, 9343
1-MS-0619  Review & Approval Desk
          for DOE/OSTI, 15102

1  Mike Jones
   MS B259, X-PA
   LANL
   Los Alamos, NM 87545

1  Jay Boettner
   MS P915, DX-DO
   LANL
   Los Alamos, NM 87545

1  Stephen Lee
   MS F663, X-CI
   LANL
   Los Alamos, NM 87545

1  Rendell Carver
   MS P940, DX-3
   LANL
   Los Alamos, NM 87545

1  Tom Bement
   MS F600, TSA-1
   LANL
   Los Alamos, NM 87545
1 Linda Verdoorn Powers  
DP18  
Department of Energy  
19901 Germantown Road  
Germantown, MD 20874

1 Chris Keane  
Department of Energy  
DP18  
19901 Germantown Road  
Germantown, MD 20874

1 Rose Mary Baltrusaitis  
Group XTA, MS B220  
LANL  
Los Alamos, NM 87545

1 Gary Carlson  
LLNL  
P. O. Box 808, L-170  
Livermore, CA 94551

1 David Luginbuhl  
DP 51  
Department of Energy  
1000 Independence Ave. SW  
Washington, DC 20585

1 Kirk Levedahl  
LLNL  
P. O. Box 808, L-016  
Livermore, CA 94551

1 John Cerutti  
MS D413, X-HM  
LANL  
Los Alamos, NM 87545

1 Randy Christensen  
LLNL  
P. O. Box 808, L-066  
Livermore, CA 94551

1 M. Peterson-Schnell  
MS B295, CIC-12  
LANL  
Los Alamos, NM 87545

1 John Reynders  
MS B260, CIC-DO  
LANL  
Los Alamos, NM 87545

1 Steven Ashby  
LLNL  
P. O. Box 808, L-561  
Livermore, CA 94551

1 Ken Hanson  
MS P940, DX-3  
LANL  
Los Alamos, NM 87545

1 Richard Klein  
LLNL  
P. O. Box 808, L-012  
Livermore, CA 94551

1 William Reed  
DP 51  
Department of Energy  
1000 Independence Ave., SW  
Washington, DC 20585

1 Robert Benjamin  
MS P940, DX-3  
LANL  
Los Alamos, NM 87545

1 Tom Adams  
LLNL  
P. O. Box 808, L-098  
Livermore, CA 94551

1 Bob Tomlinson  
MS B287, CIC-8  
LANL  
Los Alamos, NM 87545

1 Allan McPherson  
MS B287, CIC-ACL  
LANL  
Los Alamos, NM 87545