Unclassified, Unlimited Release
SAND2008-xTBDxC
SAND2008-1330P

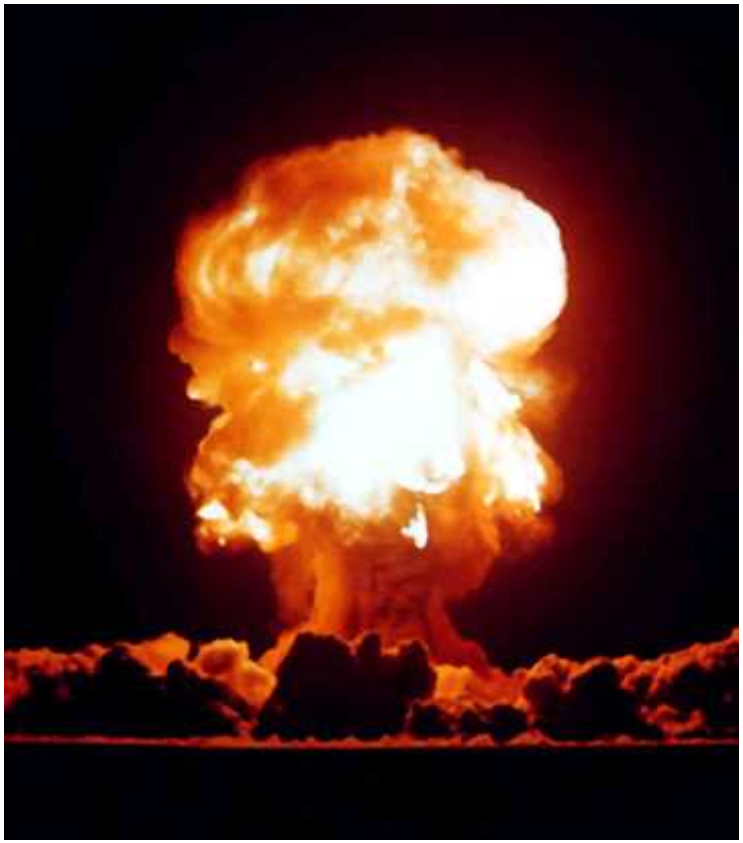
Verification & Validation:
Measured Credibility, on Demand, for Medical Applications

Martin Pilch PhD, PMP
Deputy Program Manger
Advanced Simulation and Computing Program
(Former) SNL/ASC/V&V Program Element Manager
mpilch@sandia.gov

Presented at:
FDA/NIH/NSF Workshop on Computer Methods for Cardiovascular Device Design and Evaluation
Bethesda, Maryland
March 18-19, 2008

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.

We Share a Need to Make High Consequence Decisions Informed by M&S



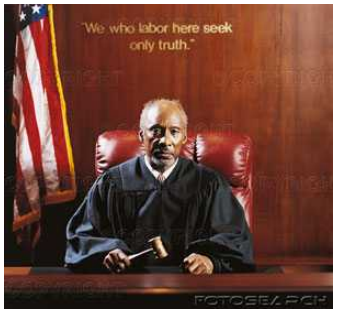
Performance, safety, security, and design decisions for the US nuclear weapons stockpile

Cardiovascular device design and evaluation

Begin With The End In Mind

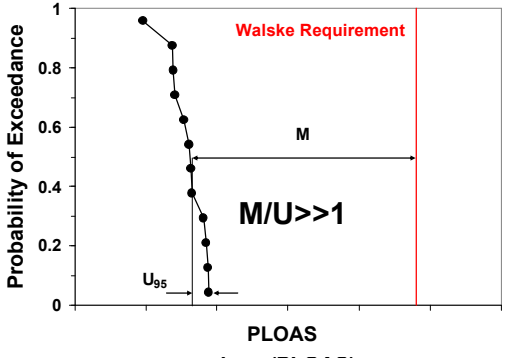
Decisions should balance testing, simulation, and the credibility of M&S that generated the simulation results

Weapons Safety in a Fuel Fire

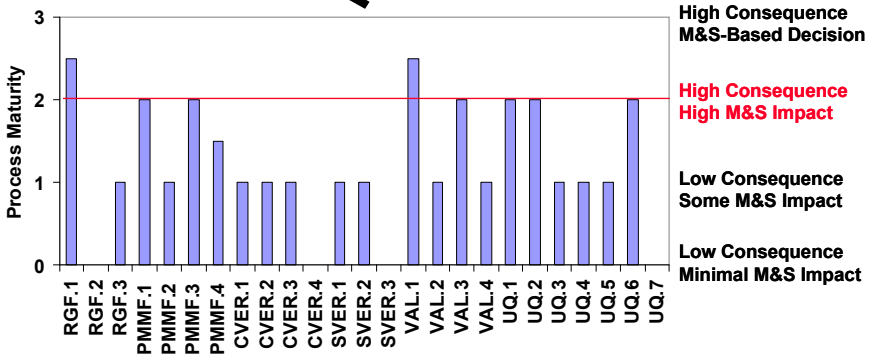


M&S supports risk-informed decisions:

PLOAS<10^{-6}?



Quantified Margins and Uncertainties



Credibility That is Measured and Communicated

What Makes M&S Results Worthy of Confidence?

Processes that support *Credible Predictive Capability*

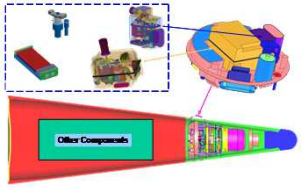
1. **RGF: Representation or geometric fidelity**
2. **PMMF: Physics and material model fidelity**
3. **CVER: Code verification**
4. **SVER: Solution verification**
5. **VAL: Validation**
6. **UQ: Uncertainty quantification**

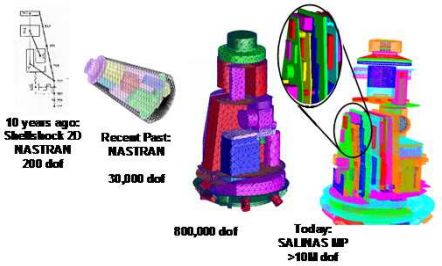
I will present 25 specific practices with examples drawn from a wide variety of M&S applications

Representation or Geometric Fidelity

Are representation errors corrupting simulation results?

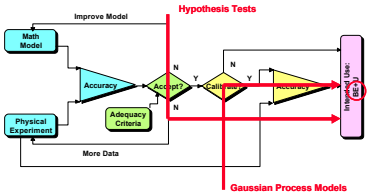
- **Characterize RGF**
  - How close to "as built" you are representing the system?
- **Quantify computation errors**
  - What impact does imperfect RGF have on simulation results?
- **Verify representation or geometry**
  - Is what you represented really what was built?





Physics and Material Model Fidelity

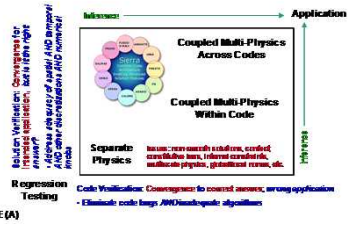
How science-based and accurate are the physics and material models?



- Characterize science basis for the models
  - Are the "models" best described as "knobs", empirical correlations, physics-informed, or fundamental physics?
- Quantify model accuracy
  - How accurate are the models?
- Assess the degree of interpolation or extrapolation
  - What is the relevance of the underlying databases?
- Perform technical review
  - Verify that the physics models are relevant, adequate, and executed in a technically sound manner

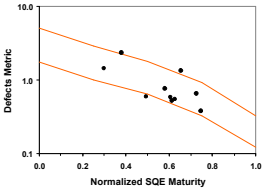
Code Verification

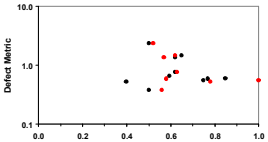
Are software errors or algorithm deficiencies corrupting simulation results?

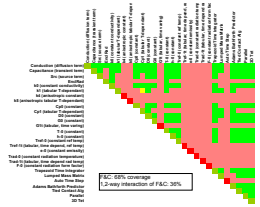


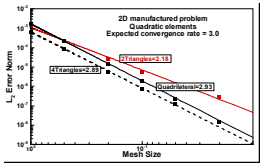
- **Apply good SQE processes**
  - Do you have a mature code development process?
- **Assess SQE processes**
  - Verify that codes are developed with an appropriate level SQE maturity?
- **Provide adequate test coverage**
  - Can the user be confident that the code is adequately tested for the intended application?
- **Quantify computation errors**
  - What is the impact of undetected code or algorithm deficiencies on simulation results?

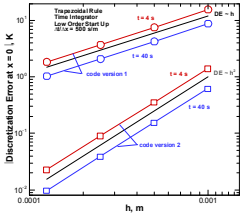
~~Code1:Code2 Comparisons~~





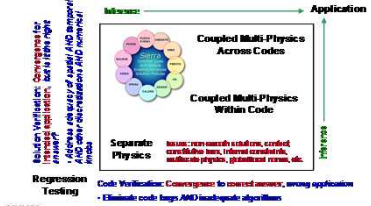




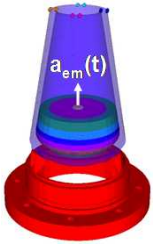


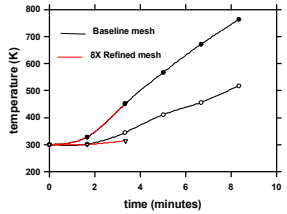
Solution Verification

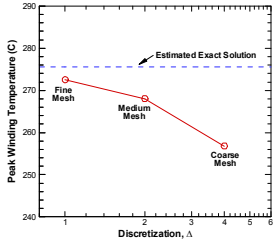
Are human procedural errors or numerical solution errors corrupting simulation results?

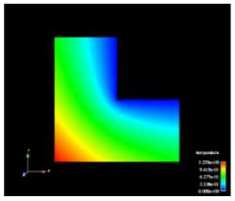


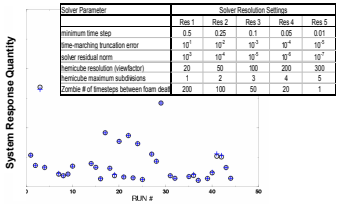
- **Quantify numerical solution errors**
  - **What is the impact of numerical solution errors on relevant system response quantities (SRQs)**
- **Verify all simulation inputs and outputs**
  - **Have we corrupted simulation results with incorrect inputs or post processing errors?**
- **Perform technical review**
  - **Verify that the solution verification activities are relevant, adequate, and executed in a technically sound manner**

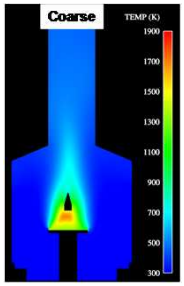


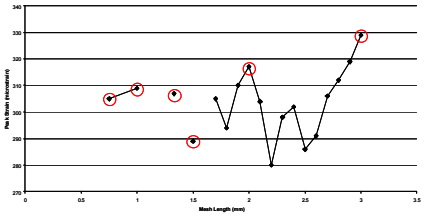








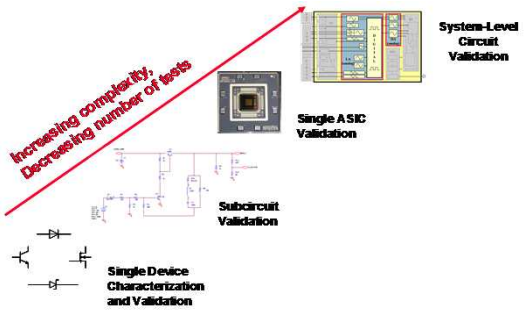




Validation

How accurate are the integrated physics and material models?

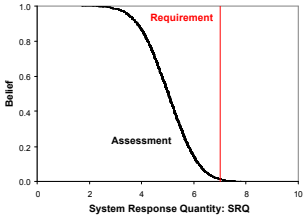
- Apply a validation hierarchy
  - Are you getting the right answers for the right reasons?
- Quantify model accuracy
  - How accurate are the models?
- Assess the degree of interpolation or extrapolation
  - What is the relevance of the underlying databases?
- Perform technical review
  - Verify that the validation activities are relevant, adequate, and executed in a technically sound manner

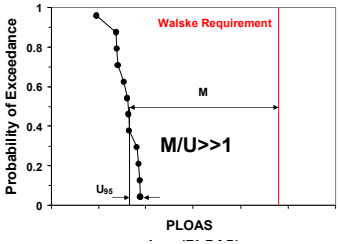


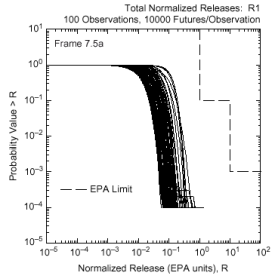
Uncertainty Quantification

What is the impact of variabilities and uncertainties on system performance and margins?

- Characterize "uncertainties" and provide a proper interpretation
  - Are uncertainties characterized, propagated, and interpreted in a manner consistent with their nature?
- Perform sensitivity analysis
  - What input uncertainties dominate output uncertainties?
- Quantify numerical propagation errors
  - How sensitive are UQ/SA results to numerical propagation errors (finite number of simulations)?
- (To be continued)







Uncertainty Quantification (Cont.)

What is the impact of variabilities and uncertainties on system performance and margins?

- Assess completeness
  - Do we cast a broad enough net that all potentially significant sources of uncertainty or error are quantified or otherwise dealt with?
- Avoid strong assumptions
  - Do strong assumptions corrupt the accuracy of UQ/SA results?
- Perform technical review
  - Verify that UQ/SA activities are relevant, adequate and carried out in a technically sound manner

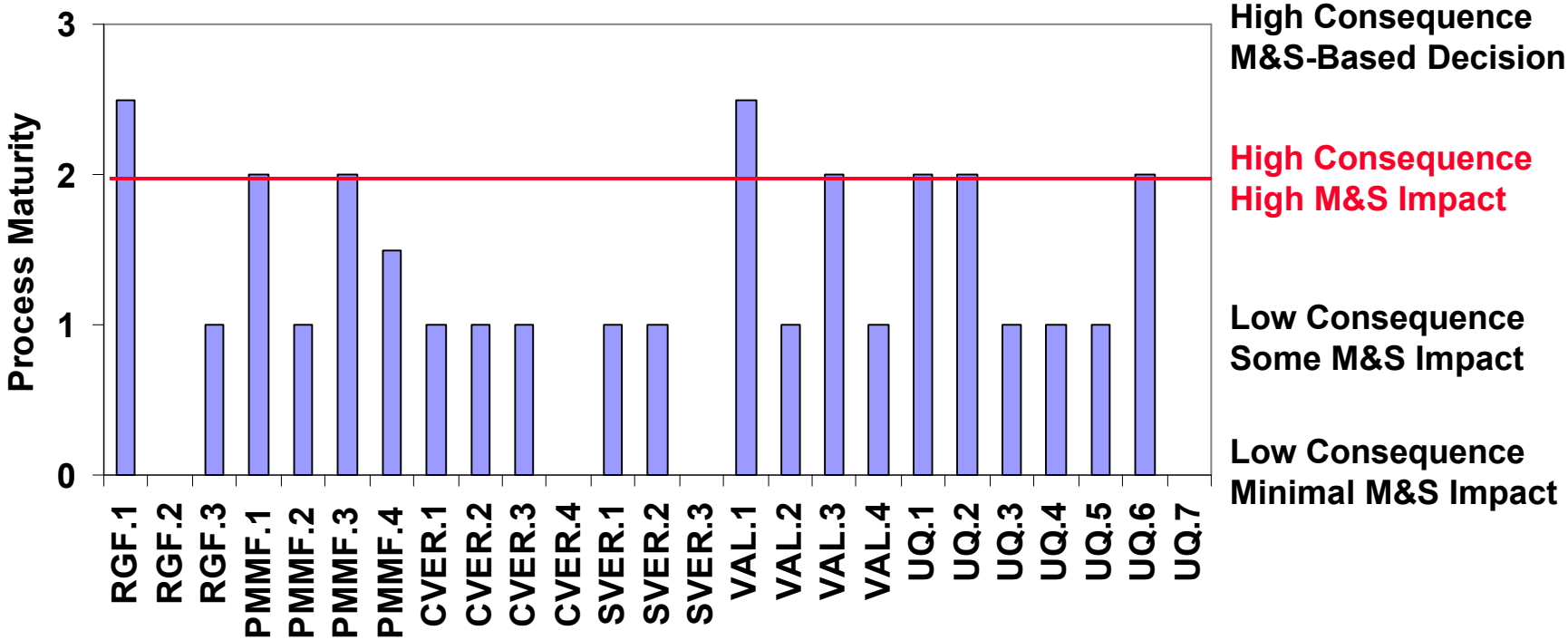
How Much is Enough?
The Predictive Capability Maturity Model (PCMM)

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
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
	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
	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
Physics and Materi... Fidelity (PMM... How science-based and acc... physics and material m...				
Code Verification (CVER) Are software errors or algorithm deficiencies corrupting the simulation results?			• 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.	controlled.
	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 offull or partial compliance to organizational SQE practices by code team • Self-assessments or formal assessments have identified compliance gaps	• Formal assessment and documentation full compliance to organizational SQE practices by group 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	• Sustained VERTS re-run regularly w 75% F&C coverage and 75% coverage of all way interactions of F&C • VERTS address *convergence* behavior t...

Complete Table Available Upon Request:
mpilch@sandia.gov

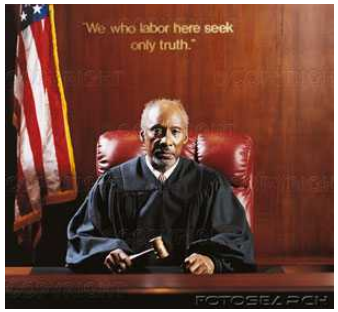
Key Concepts Described Further:
Oberkampf, Pilch, Trucano, *Predictive Capability Maturity Model for Computational Modeling and Simulation*, SAND2007-5948, Oct 2007

The PCMM Can Be Used to Measure and Communicate the Credibility of Simulation Results in the Context of a Specific Application



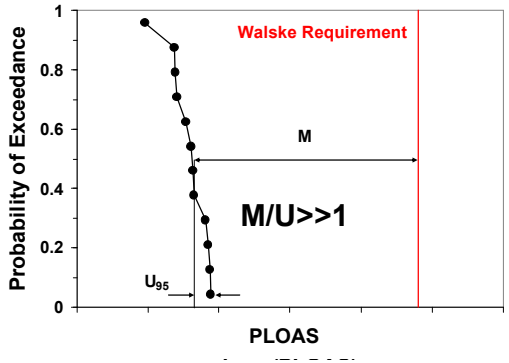
Risk-Informed Decisions for High Consequence Applications

Weapons Safety in a Fuel Fire

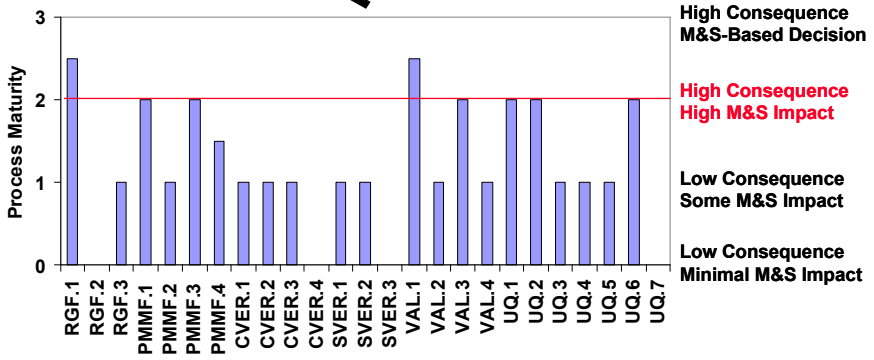


M&S supports risk-informed decisions:

PLOAS<10^{-6}?



Quantified Margins and Uncertainties



Credibility That is Measured and Communicated

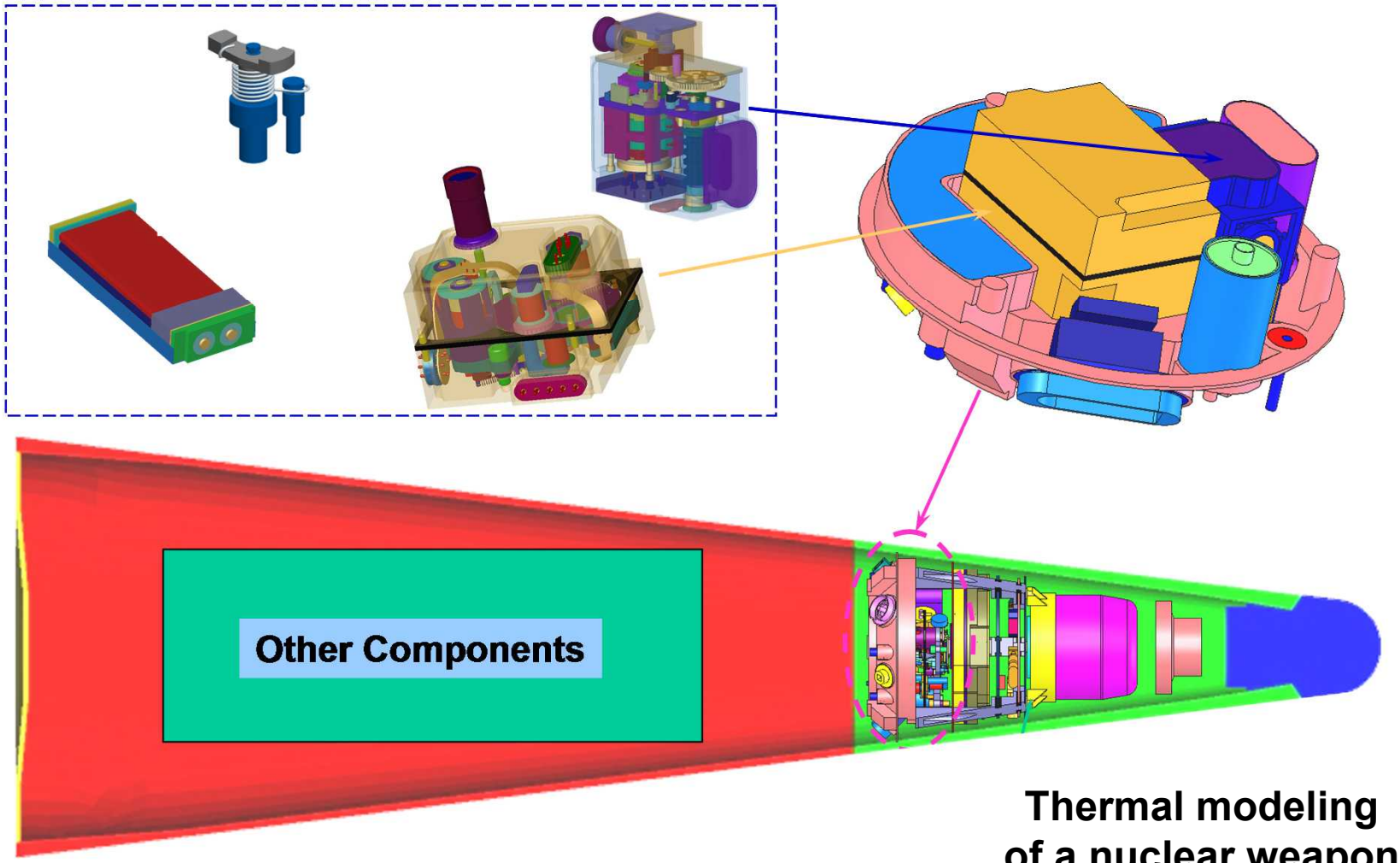
You Can't Know If You Don't Ask!

"Due diligence means asking *all* the questions, even if you don't think you'll like the answers."

Hyperlinked Pages

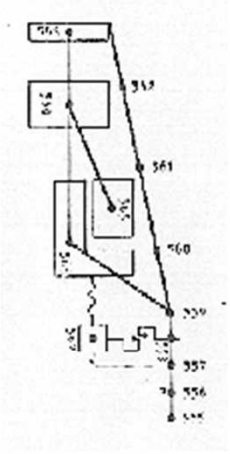
- Listed in order as they appear in main body of talk

Modern Computing Platforms Enable "~As Built" Geometric Fidelity

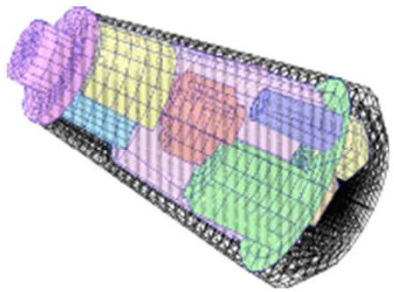


Thermal modeling of a nuclear weapon

Modern Computing Platforms Enable "~As Built" Geometric Fidelity

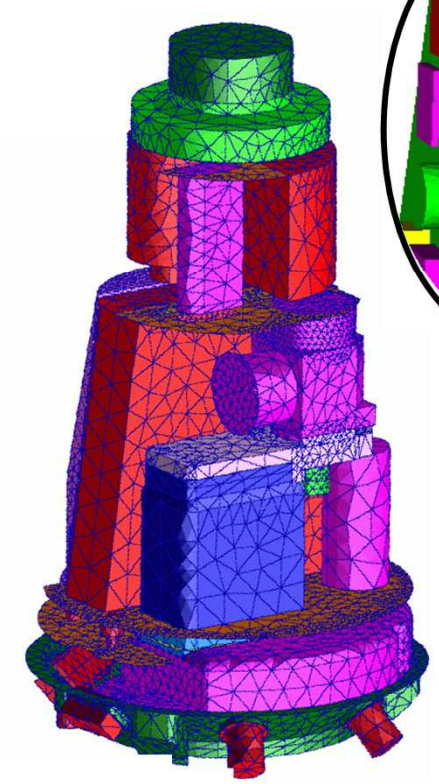


10 years ago:
Shellshock 2D
NASTRAN
200 dof

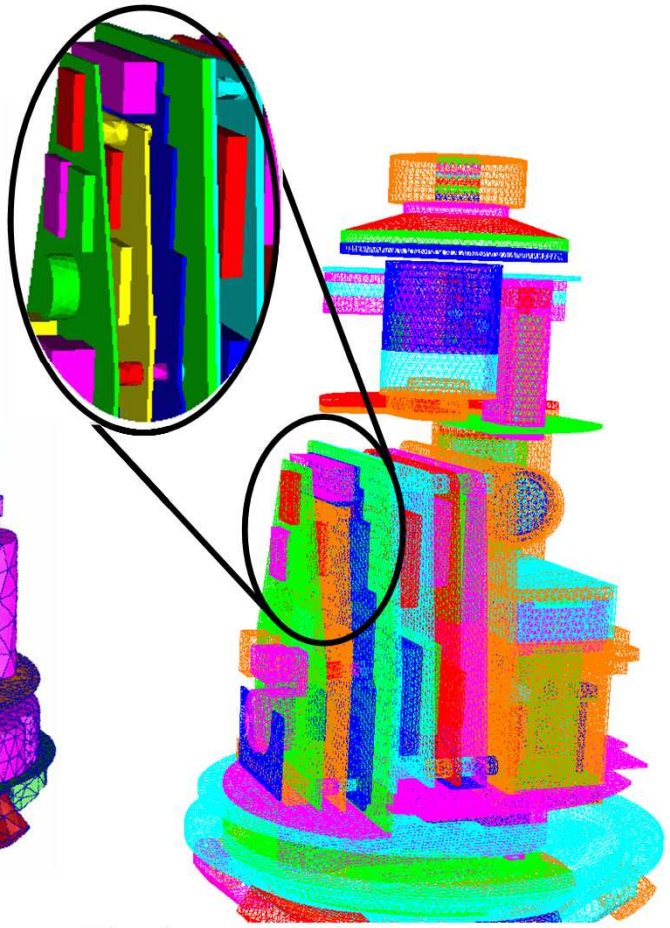


Recent Past:
NASTRAN

30,000 dof



800,000 dof

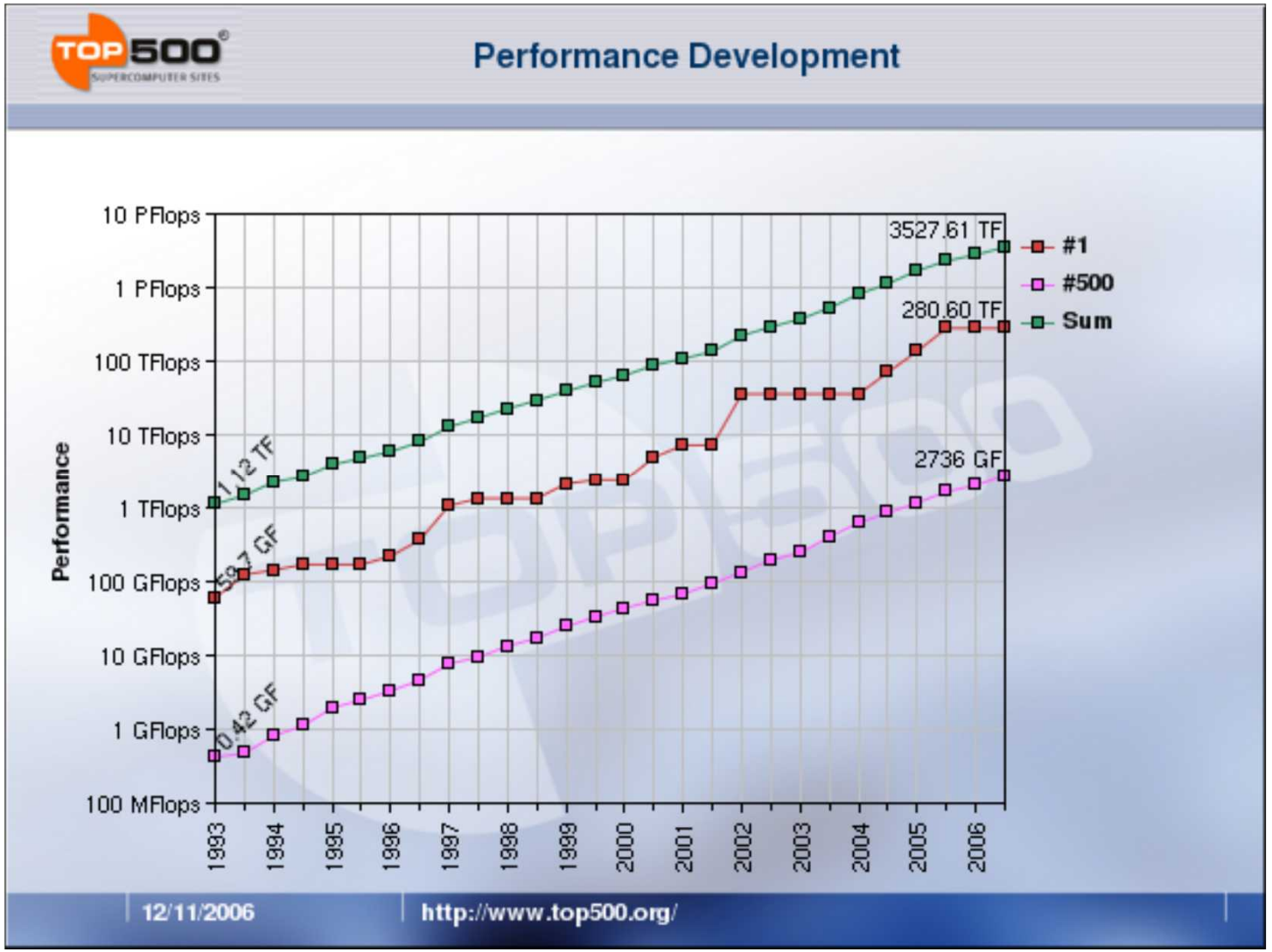


Today:
SALINAS MP
>10M dof

Structural dynamics

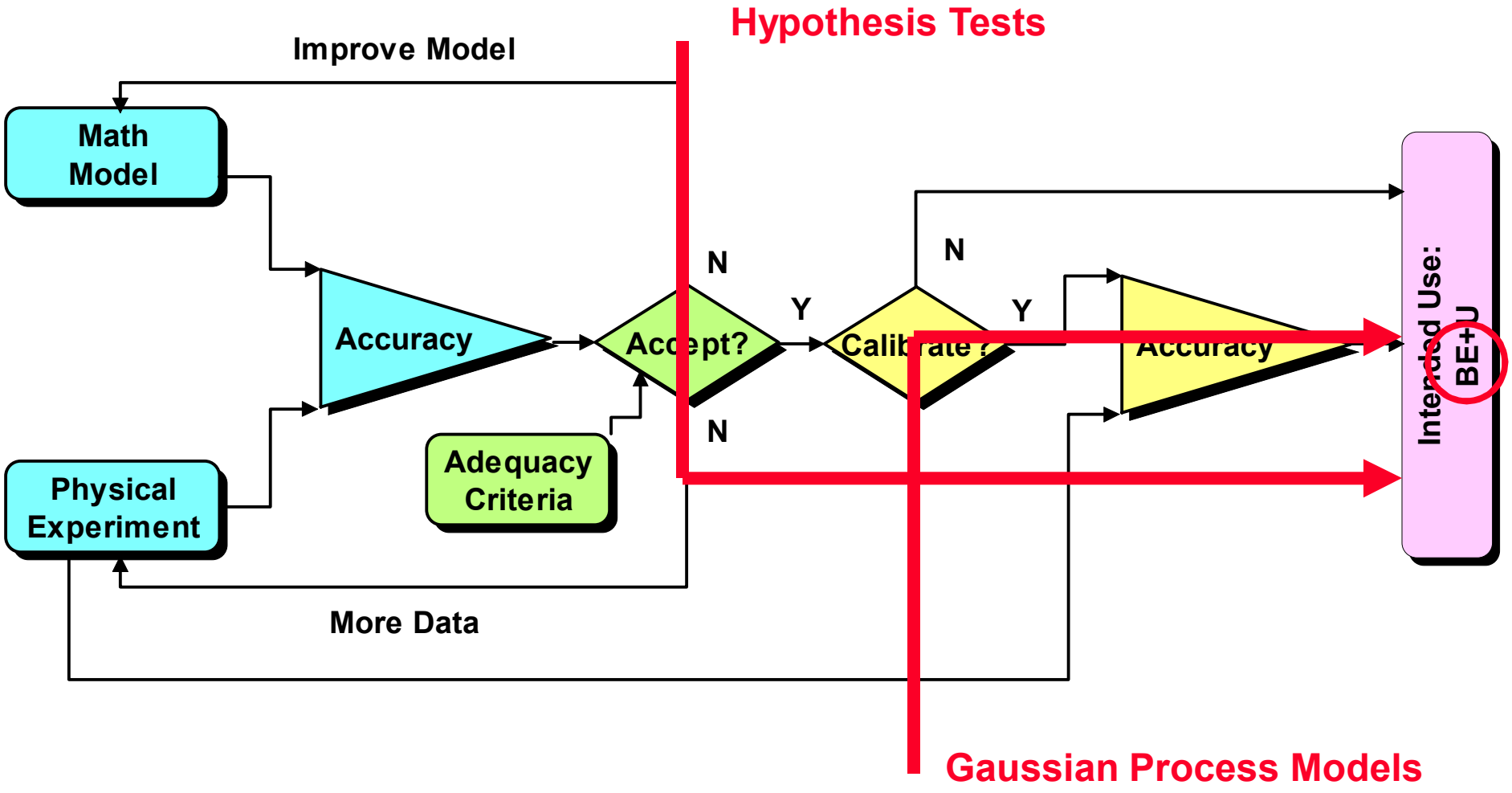
If Not Today, Then Imagine the Future

Computing Speed - Dec. 2006



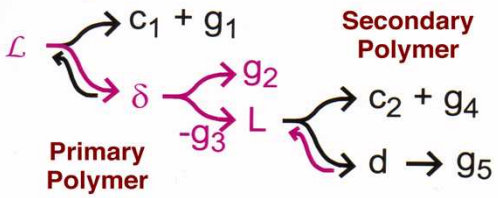
Validation is Assessment Supporting BE+U

Calibration is not Validation



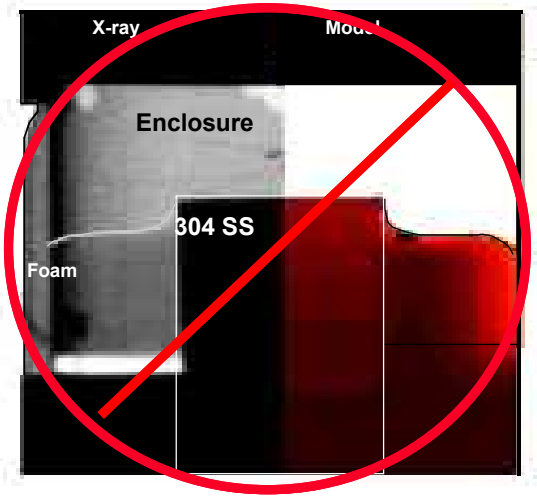
Validation is Statistical

Vugraph Norms Are Not Adequate

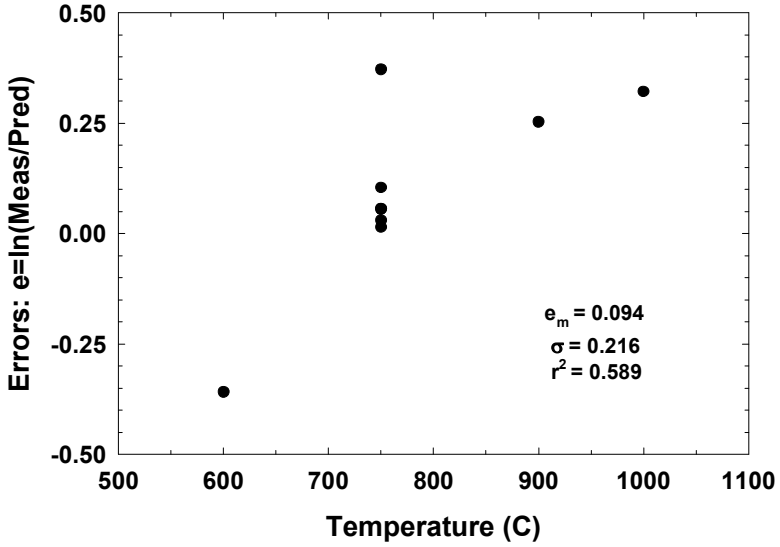


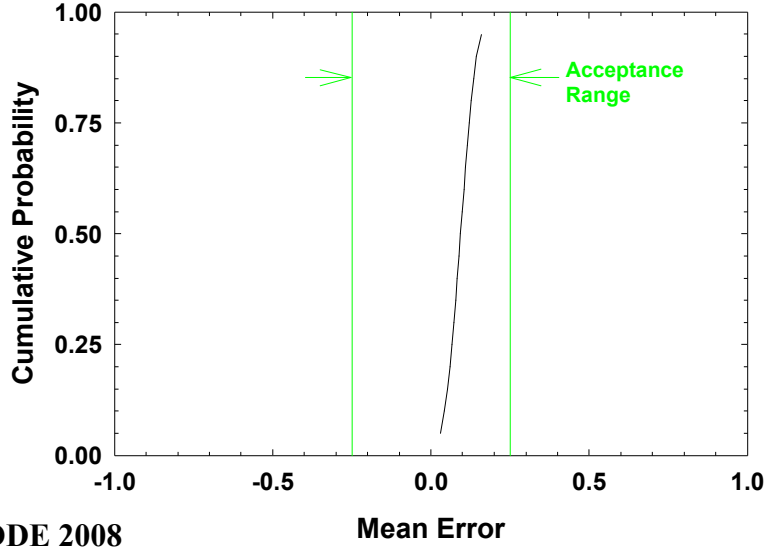
Physics-informed model

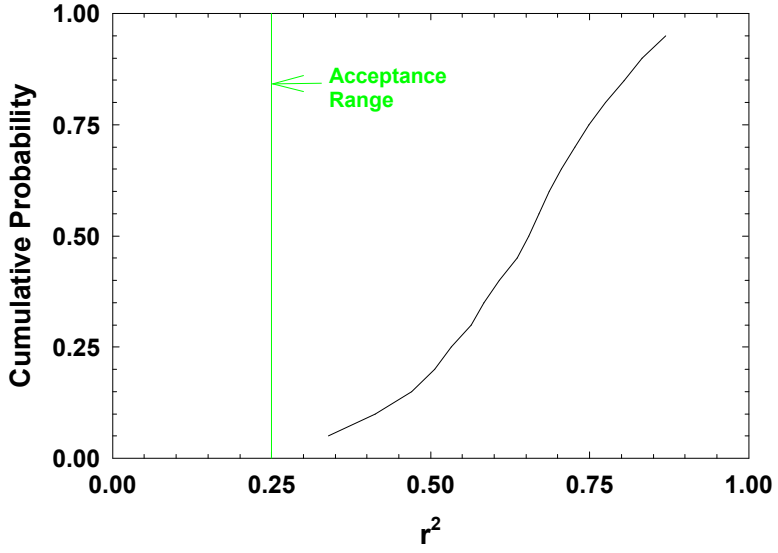
Vugraph Norm



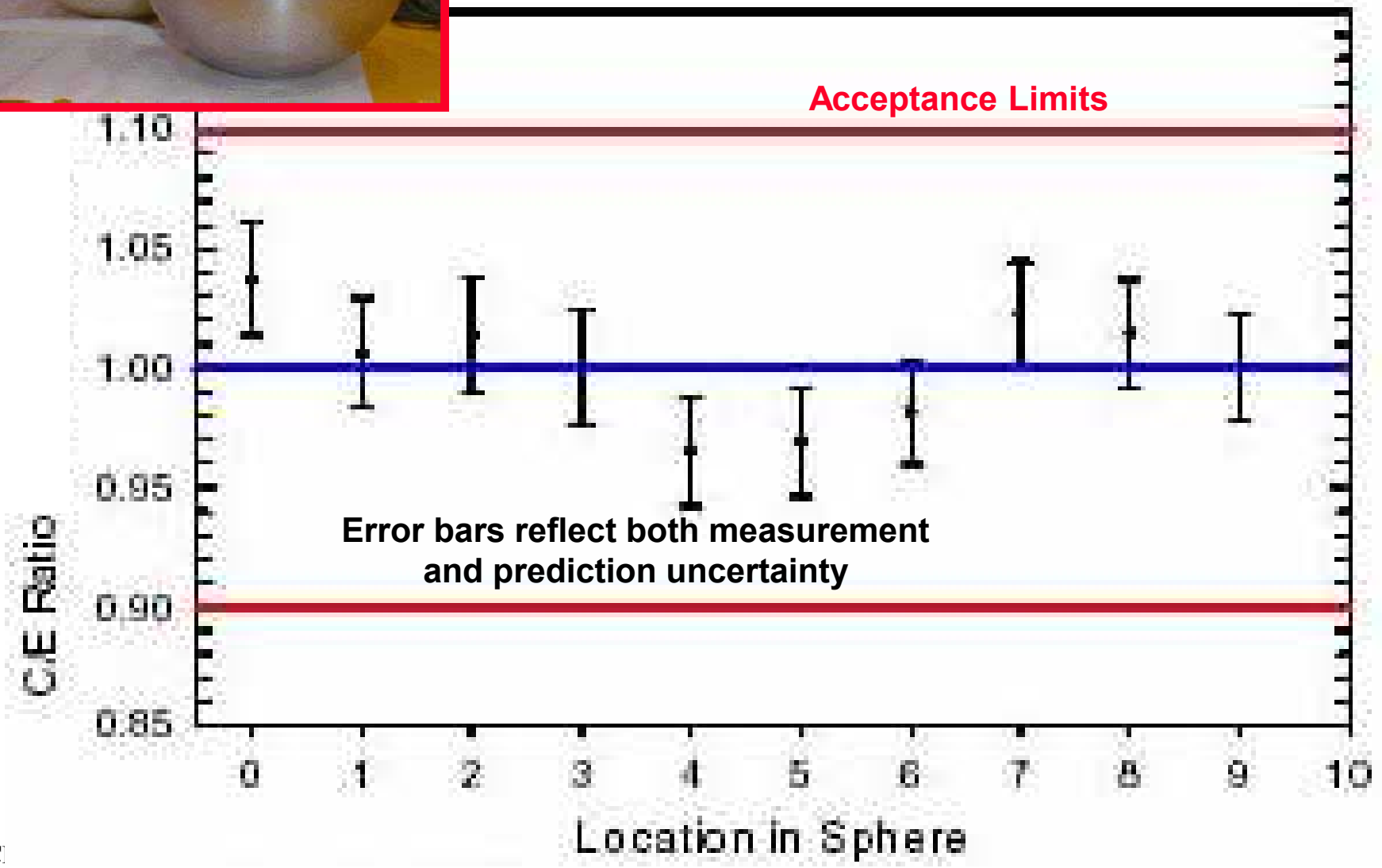
T-dot BC is more appropriate than as-tested T-fixed







Neutron Attenuation in Test Objects



Attributes of Code and Solution Verification

Demonstrating Convergence to Correct Answer for the Intended Application

Solution Verification: Convergence for intended application, *but is it the right answer?*

• *Address adequacy of spatial AND temporal AND other discretizations AND numerical knobs*

Inference → Application

Coupled Multi-Physics Across Codes

Coupled Multi-Physics Within Code

Separate Physics

Issues: non-smooth solutions, contact, constitutive laws, internal constraints, multiscale physics, global/local norms, etc.

Inference ↑

Regression Testing

Code Verification: Convergence to correct answer, *wrong application*

• Eliminate code bugs *AND* inadequate algorithms

Code to Code Comparisons Are a Poor Substitute for Formal Verification

Code Comparison Principle (CCP)

Code 1 = assessed code Code 2 = benchmark code

$$\|\text{Code 1} - \text{Truth}\| \le \|\text{Code 1} - \text{Code 2}\| + \|\text{Code 2} - \text{Truth}\|$$

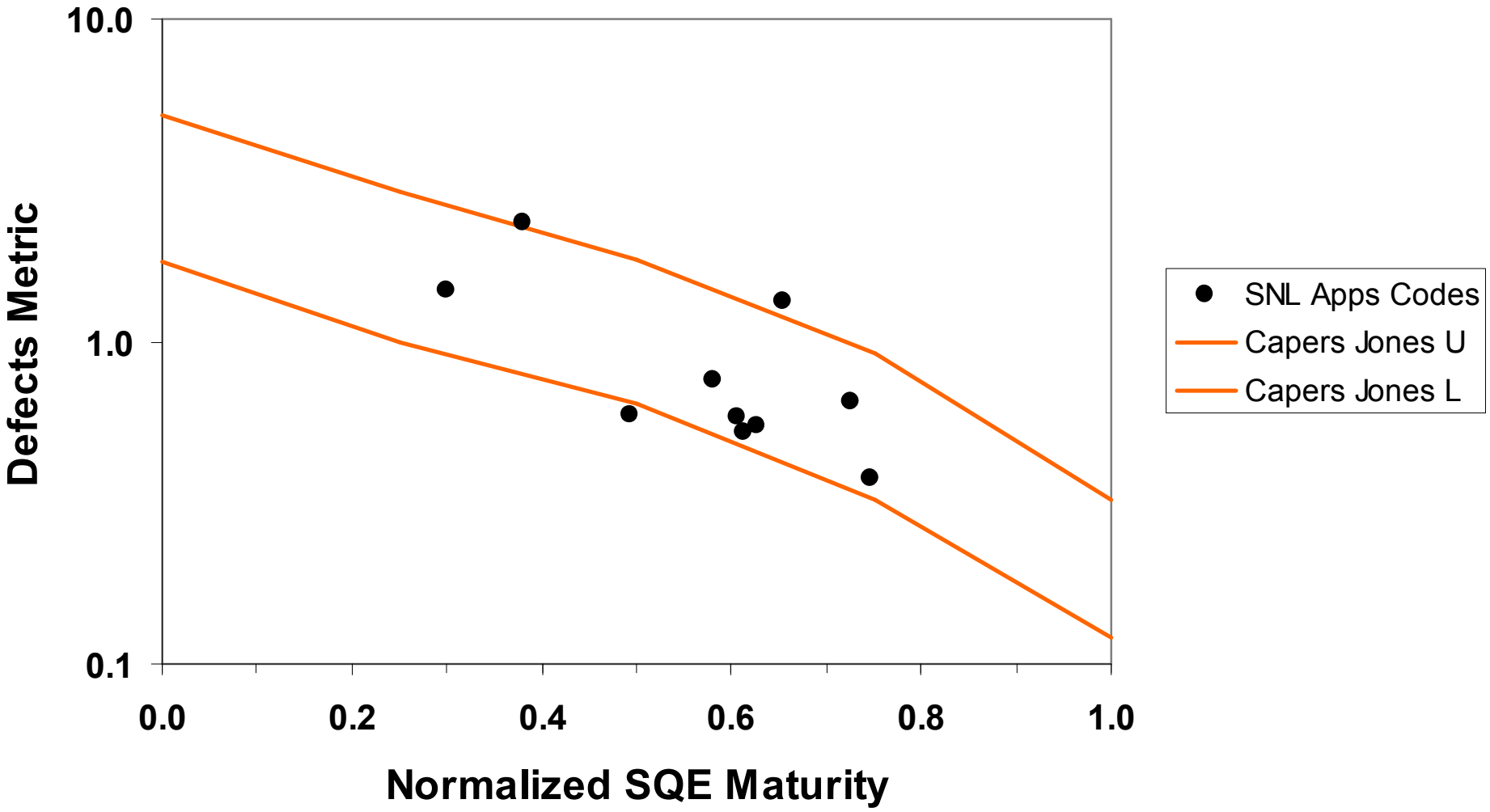
$$\therefore \|\text{Code 1} - \text{Code 2}\| \cdot$$

What if this term is not negligible?

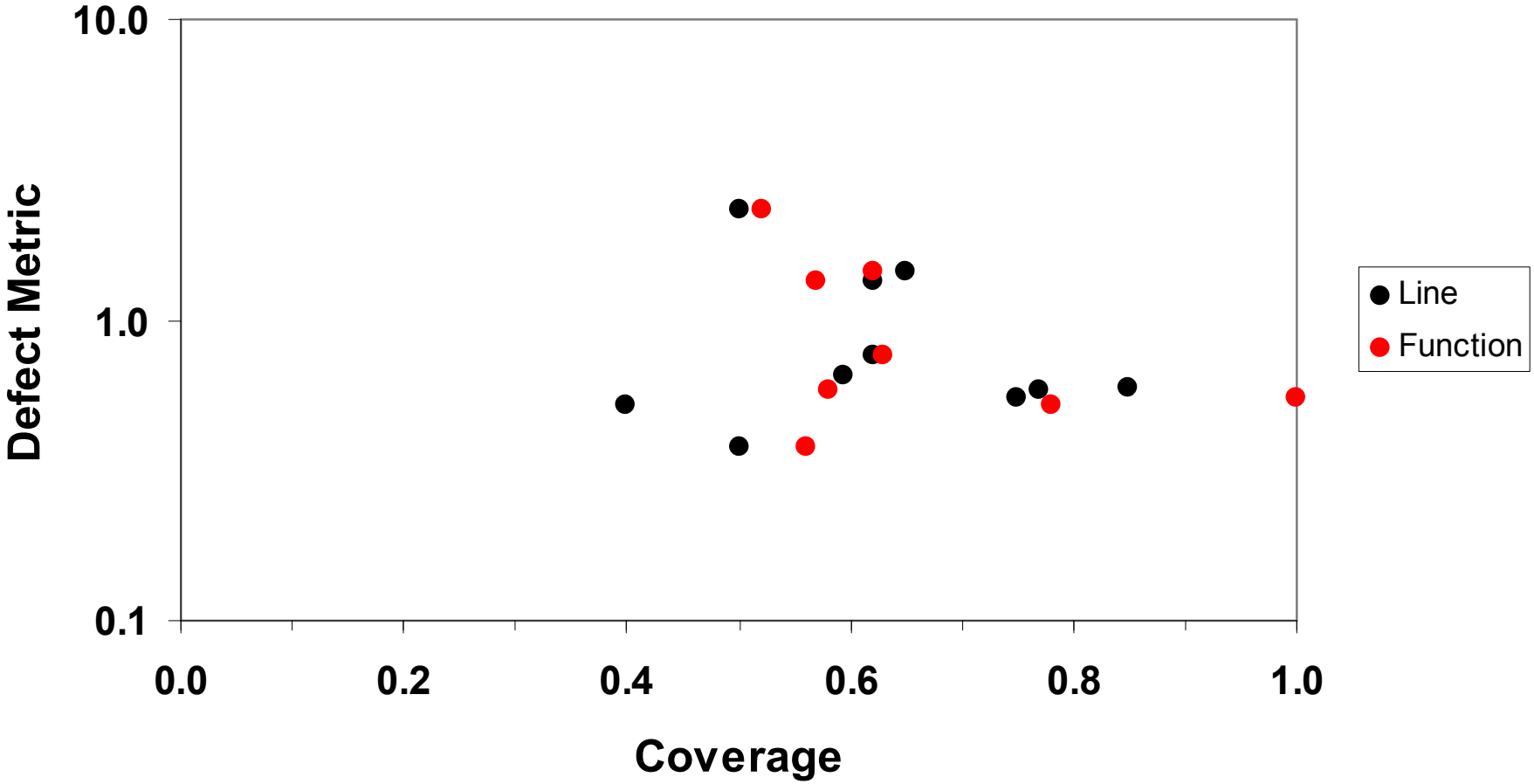
- Could be that Code 1 models are different from Code 2 models
- Could be a bug in Code 1 or Code 2
- Could be an algorithm flaw in Code 1 or Code 2
- Could be that Code 1 or Code 2 model is not converged

Points to path for better code-to-code comparisons; but if Code 2 is formally verified, why not verify Code 1 to the same verification test suite? And if not, why bother with the code-to-code comparison?

Good SQE Practices Reduces Defects



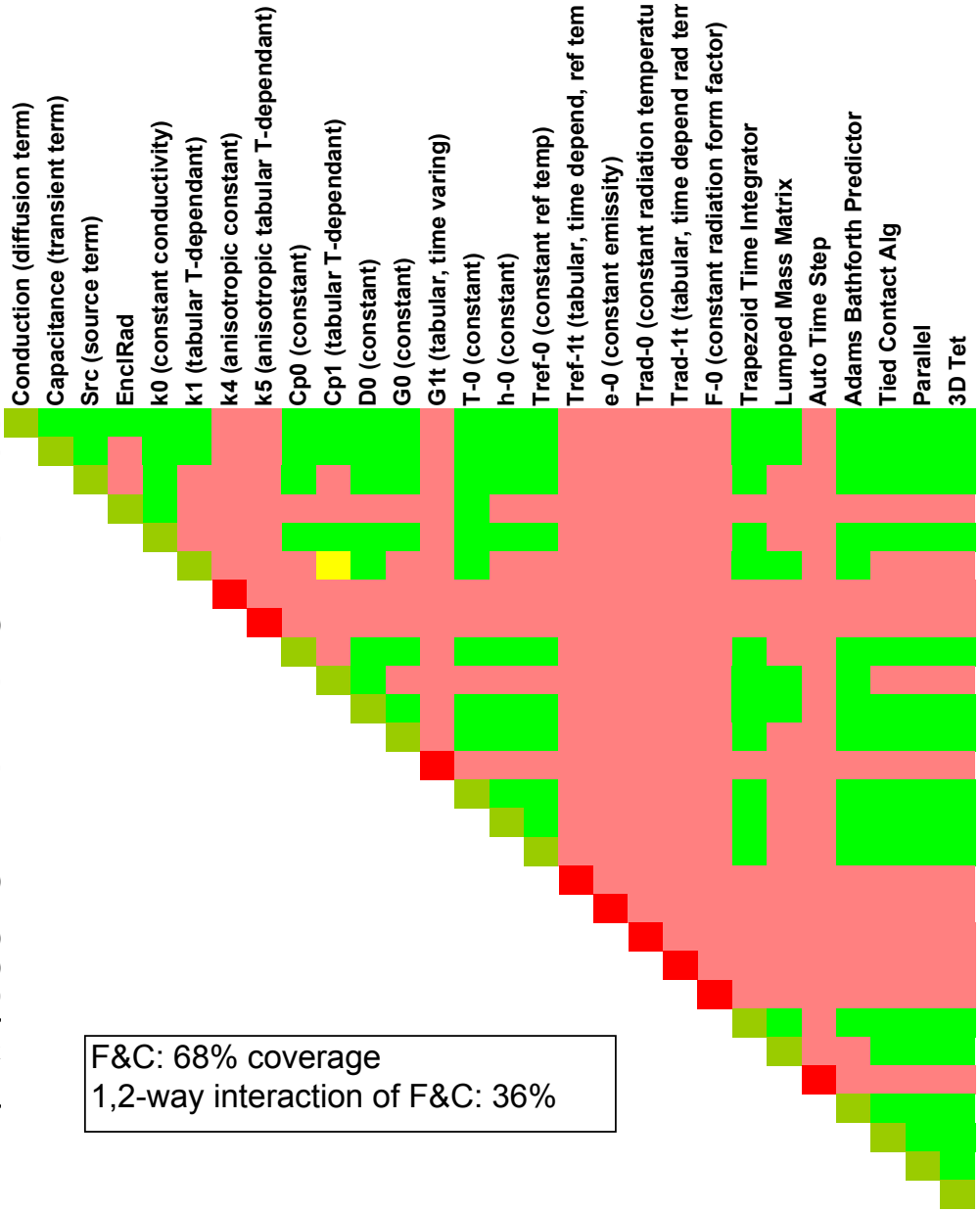
Why Doesn't Code Testing Have a More Decisive Impact on Defects?



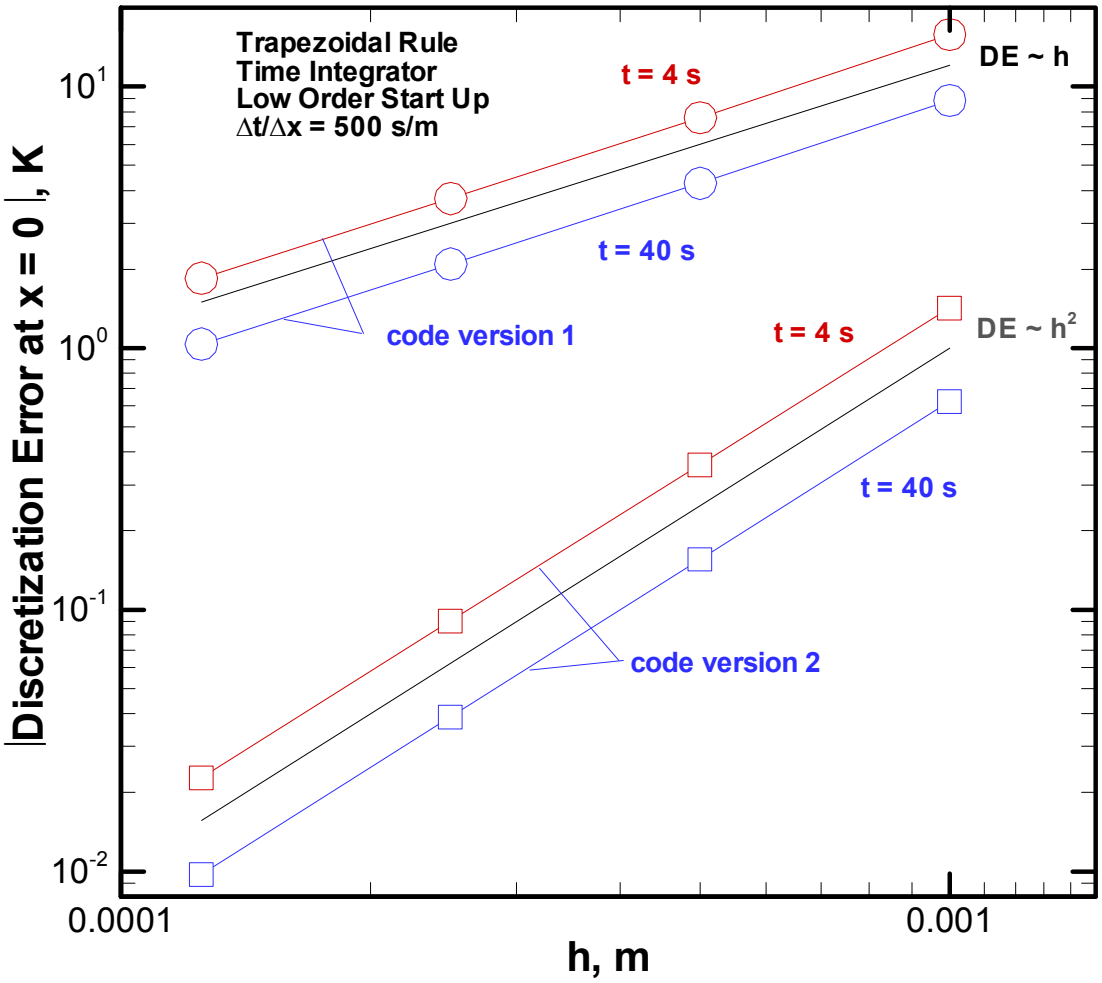
Speculation: Simple line coverage says nothing about coverage for a particular application, usually says nothing about algorithm deficiencies, and say nothing about features and capabilities and their interactions for a specific application

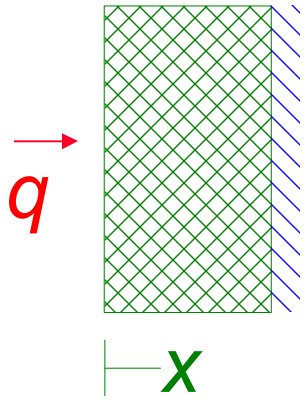
We Are Shifting Our Focus to Verification of Features and Capabilities and Their Interactions

Thermal analysis of a weapon in a fuel fire



"Order of Convergence" is a Sensitive Metric for Detecting Algorithm Deficiencies



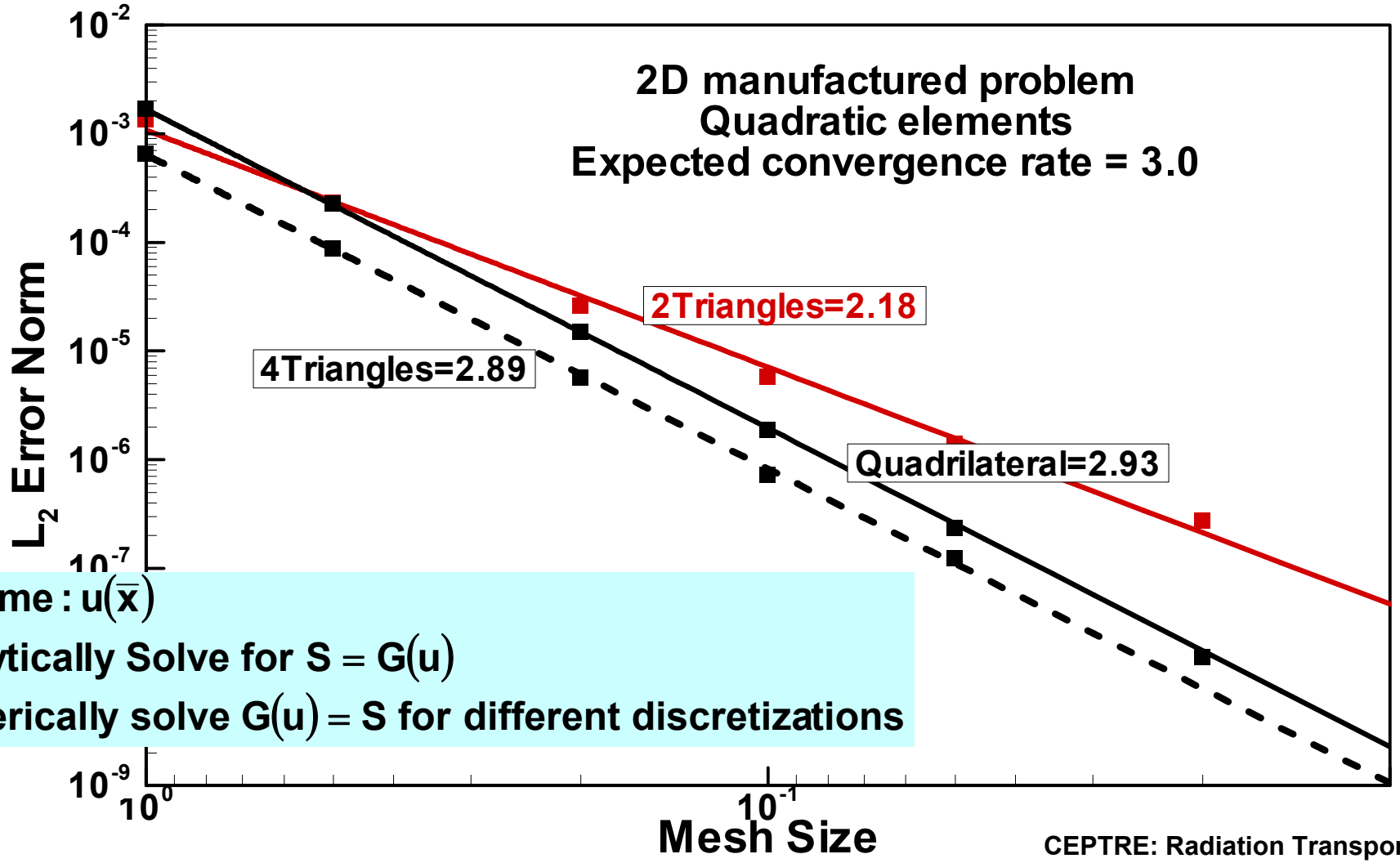


- Transient response of planar 1-D slab to constant flux with analytic solution as the benchmark
- Code bug discovered and fixed based on priority and resource availability. Status tracked in code issue log, which can be accessed by analysts

Modeled as full 3-D object

No Exact Analytic Solution?

Verification with a Manufactured Solution



Assume : $u(\bar{x})$

Analytically Solve for $S = G(u)$

Numerically solve $G(u) = S$ for different discretizations

Attributes of Code and Solution Verification

Demonstrating Convergence to Correct Answer for the Intended Application

Solution Verification: Convergence for intended application, *but is it the right answer?*

- *Address adequacy of spatial AND temporal AND other discretizations AND numerical knobs*

Inference → Application

Coupled Multi-Physics Across Codes

Coupled Multi-Physics Within Code

Separate Physics

Issues: non-smooth solutions, contact, constitutive laws, internal constraints, multiscale physics, global/local norms, etc.

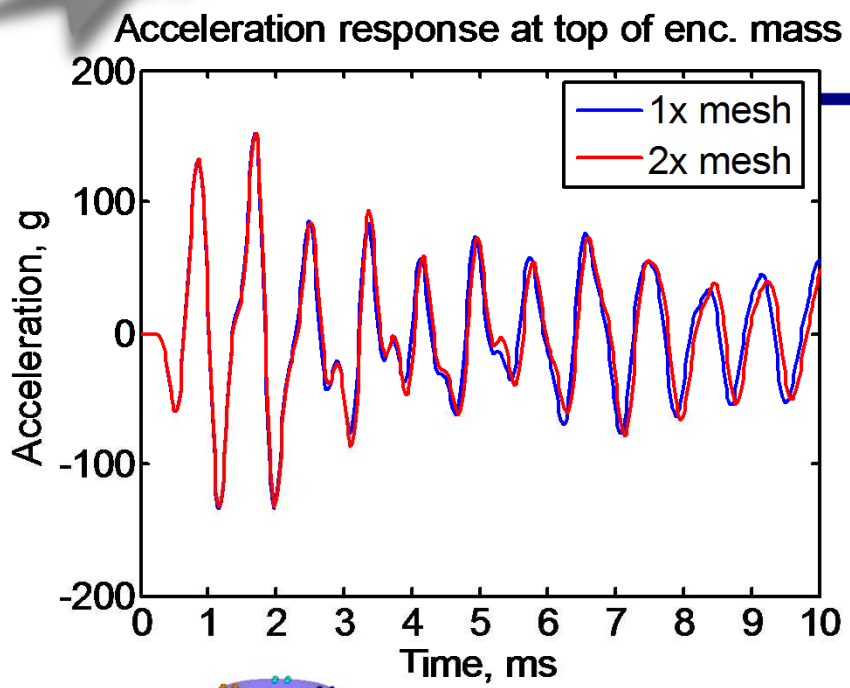
Inference

Regression Testing

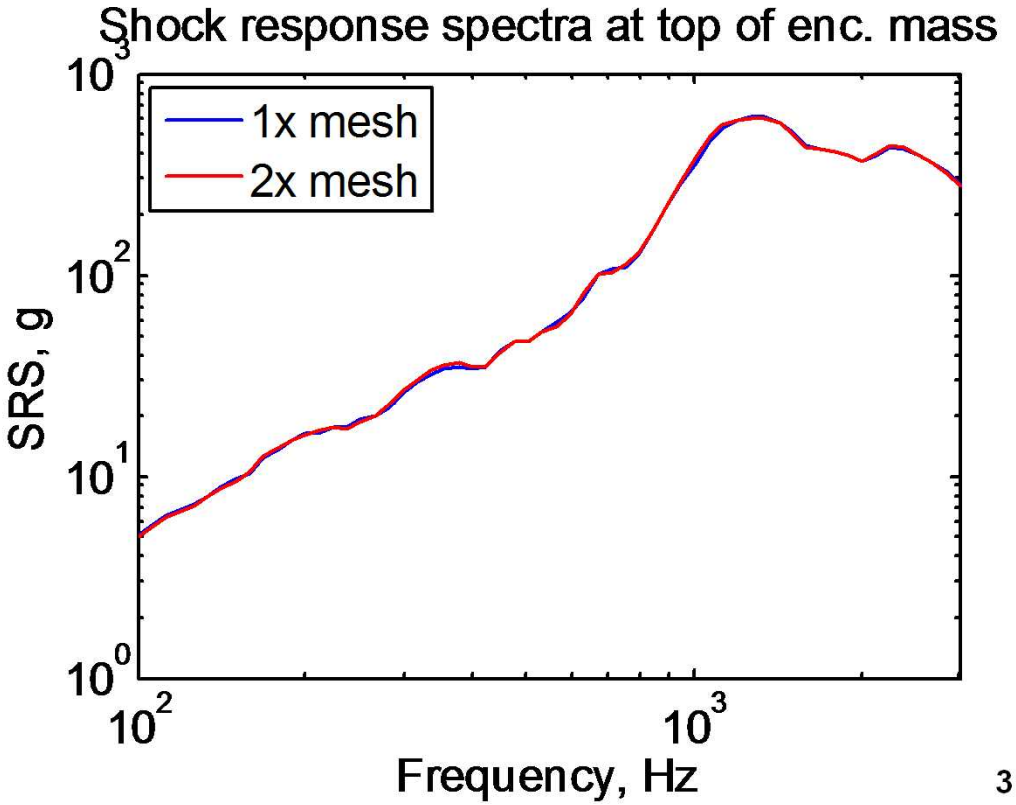
Code Verification: Convergence to correct answer, *wrong application*

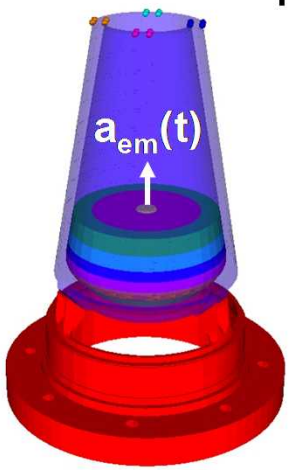
- Eliminate code bugs *AND* inadequate algorithms

It's Common to Explore Sensitivity to Mesh Parameters



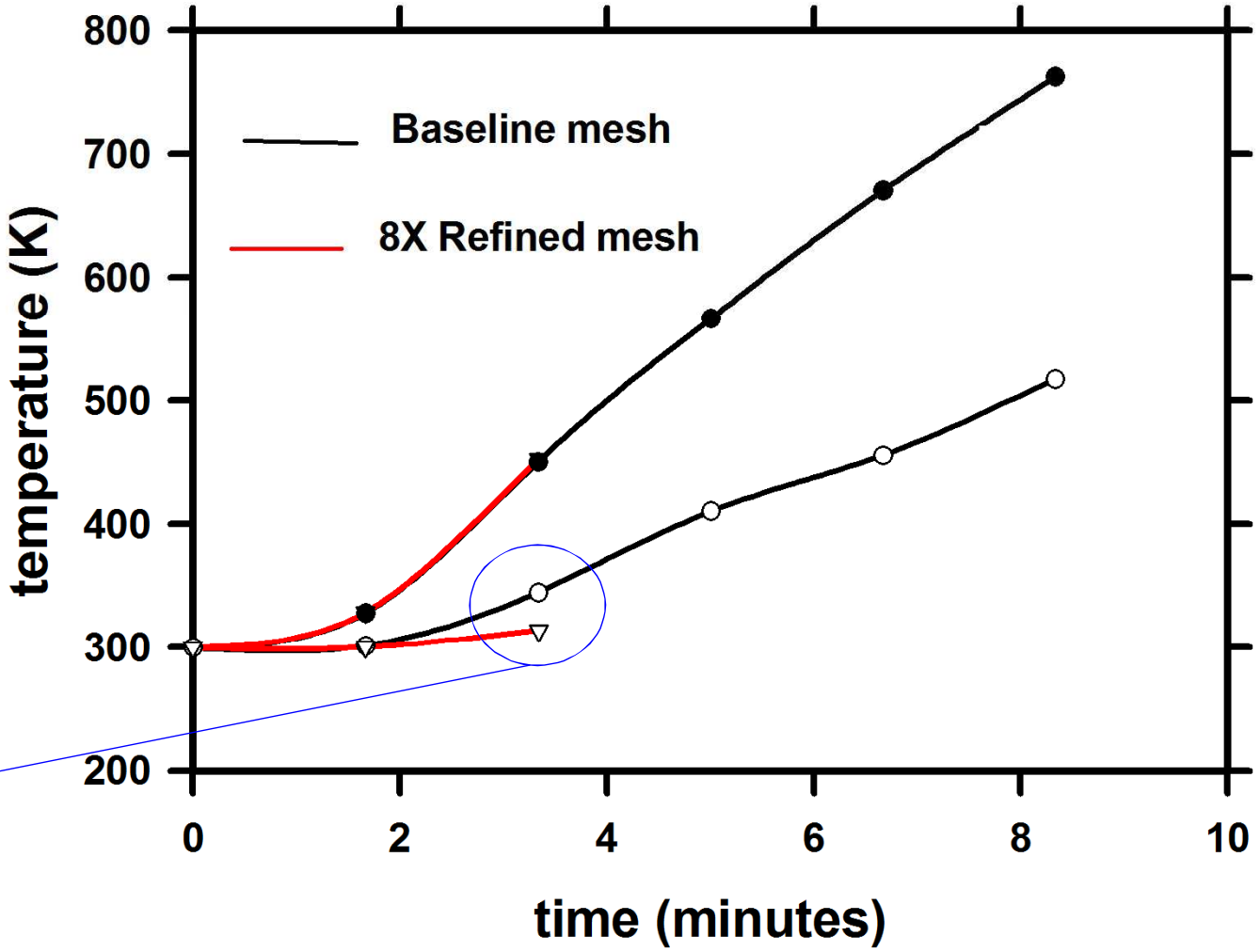
Max. relative error between SRS: +/- 5%





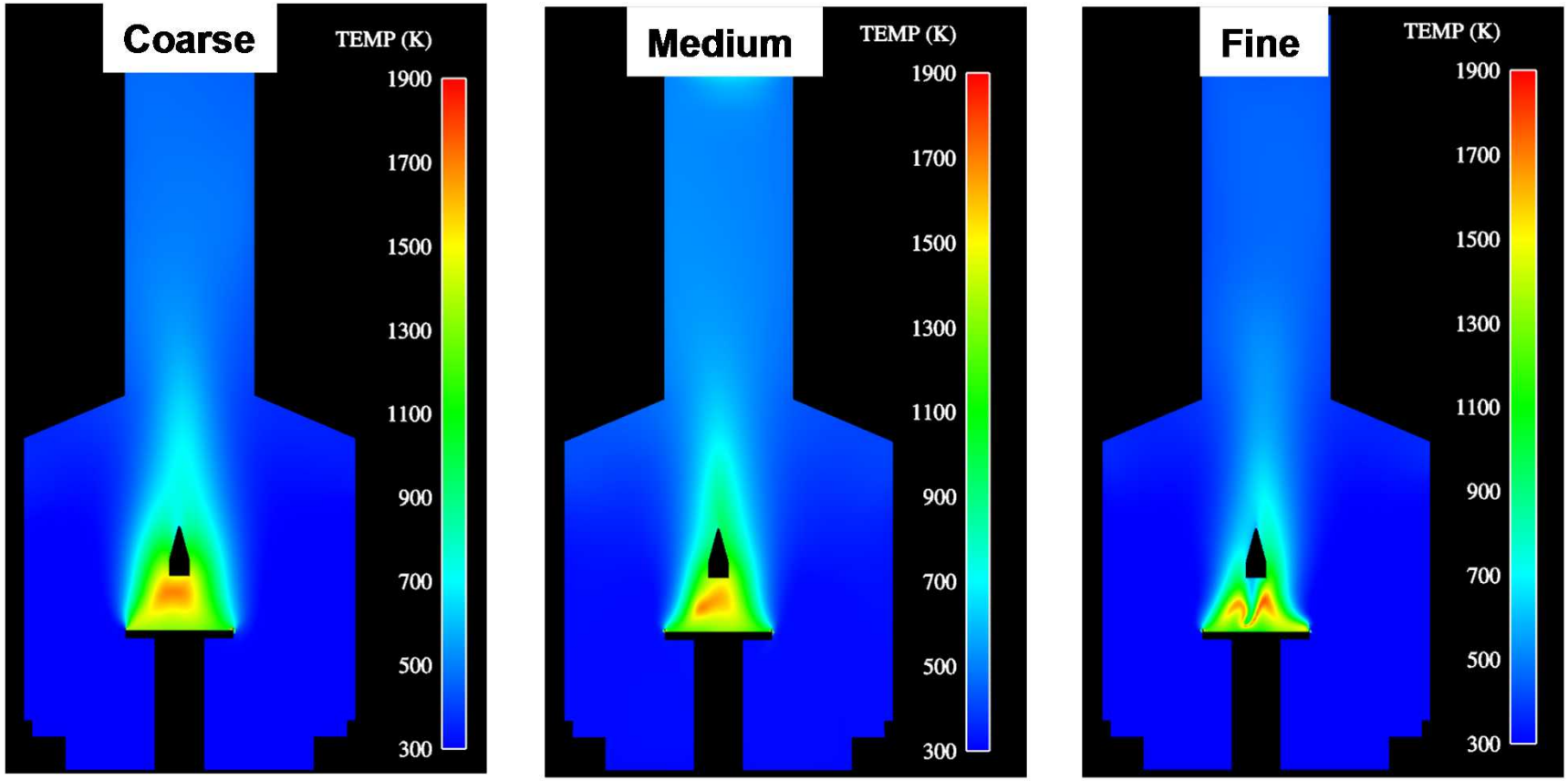
Structural Dynamics

Solution Verification on High Fidelity Models is Hard



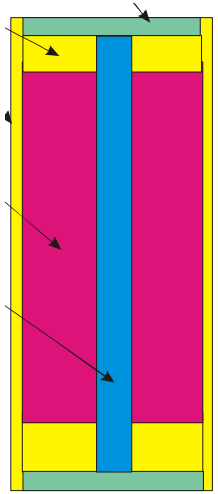
Critical heat transfer path under-resolved because of large discontinuity in material properties

Discretization Study Revealed Bifurcation of Solution Space

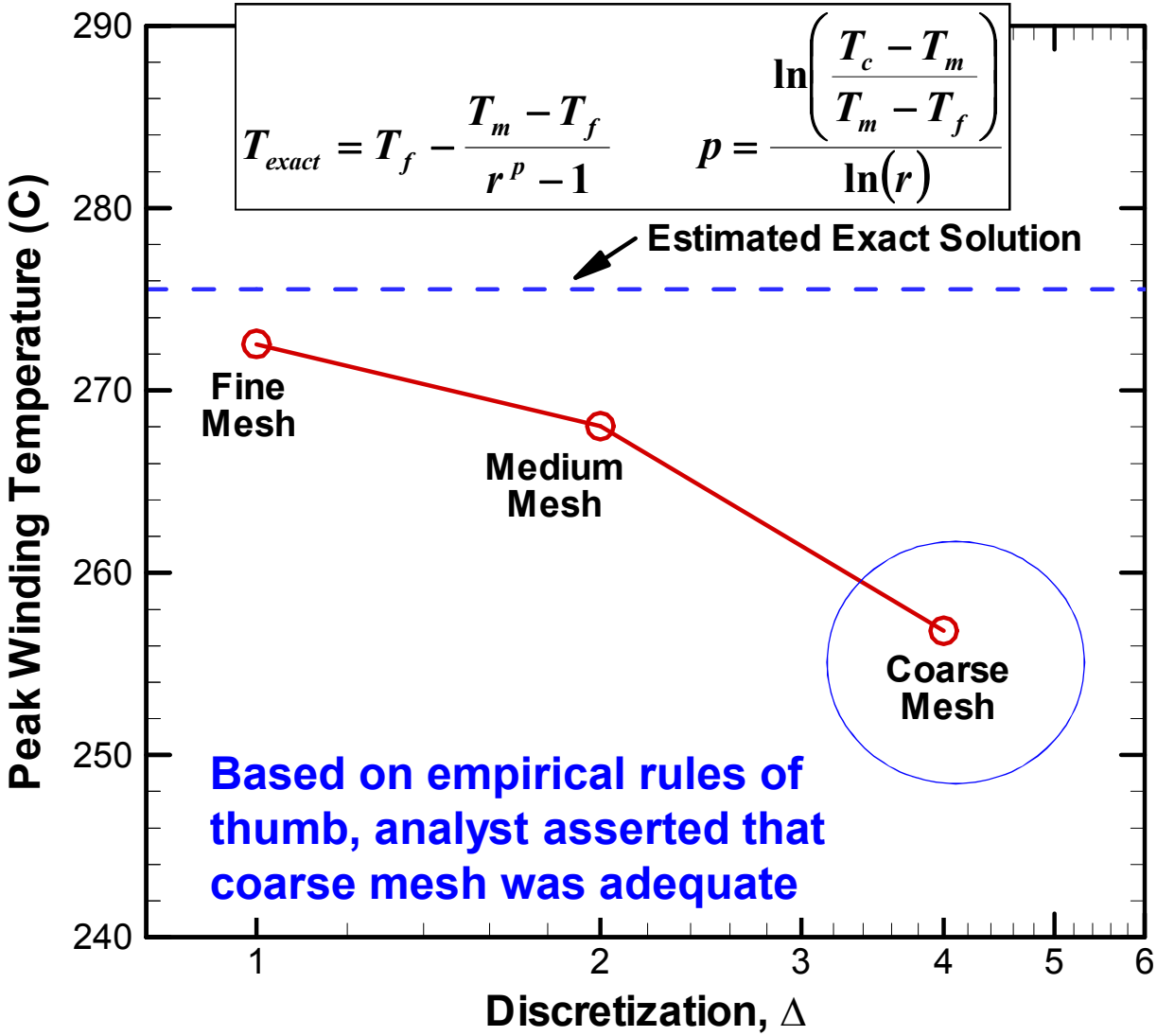


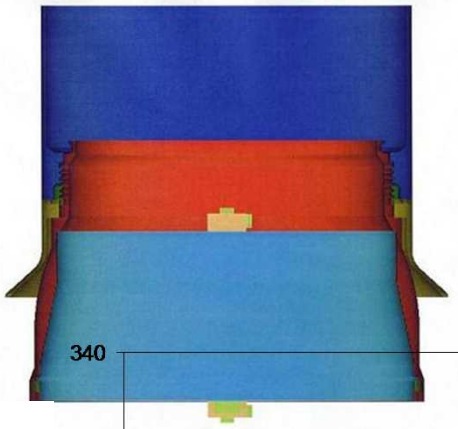
Calorimeter Fire: BVG Solutions

Numerical Errors
Pollute Validation Assessments



$$T_{exact} = T_f - \frac{T_m - T_f}{r^p - 1} \qquad p = \frac{\ln\left(\frac{T_c - T_m}{T_m - T_f}\right)}{\ln(r)}$$

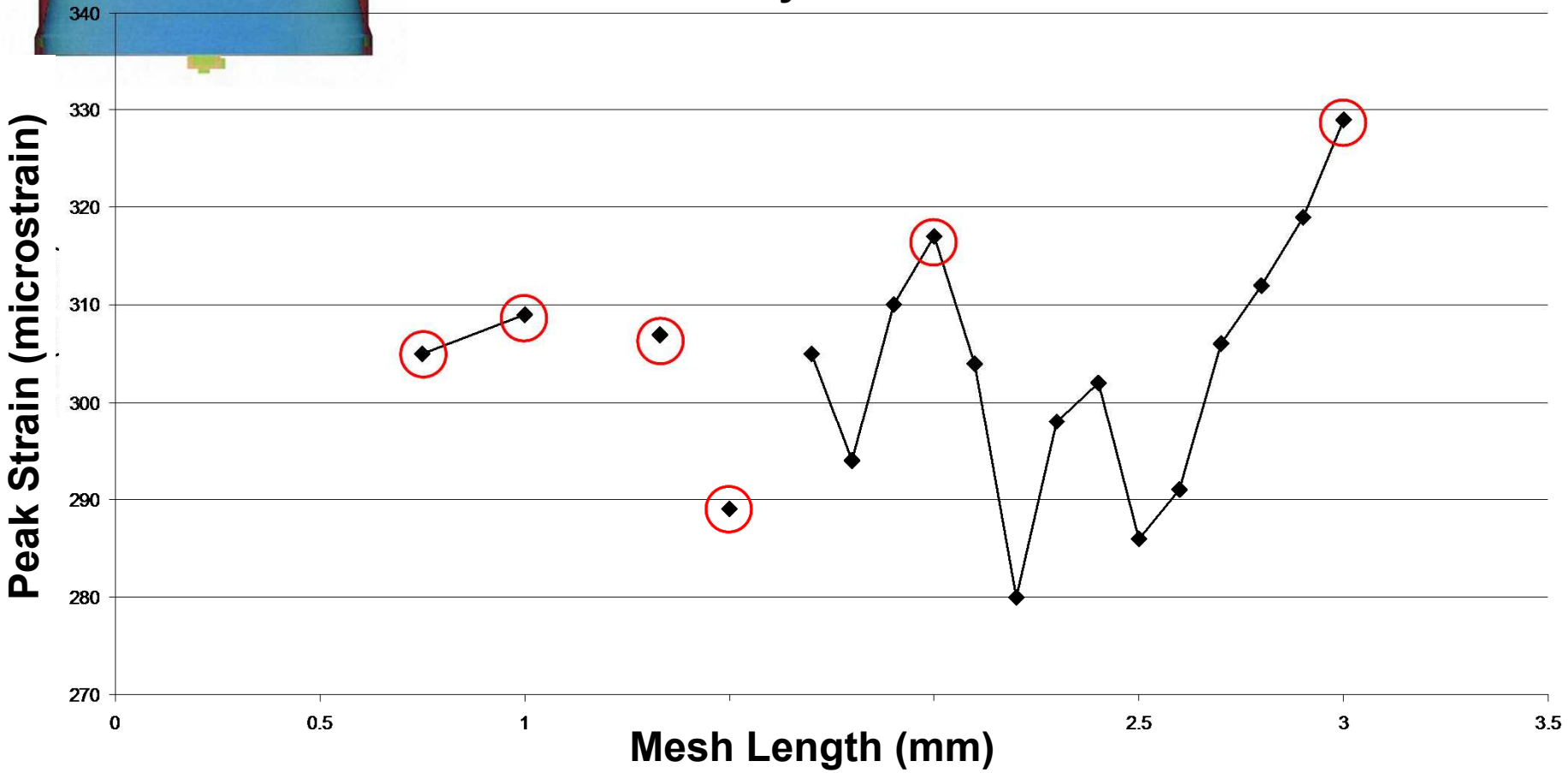




Solutions Don't Always Converge

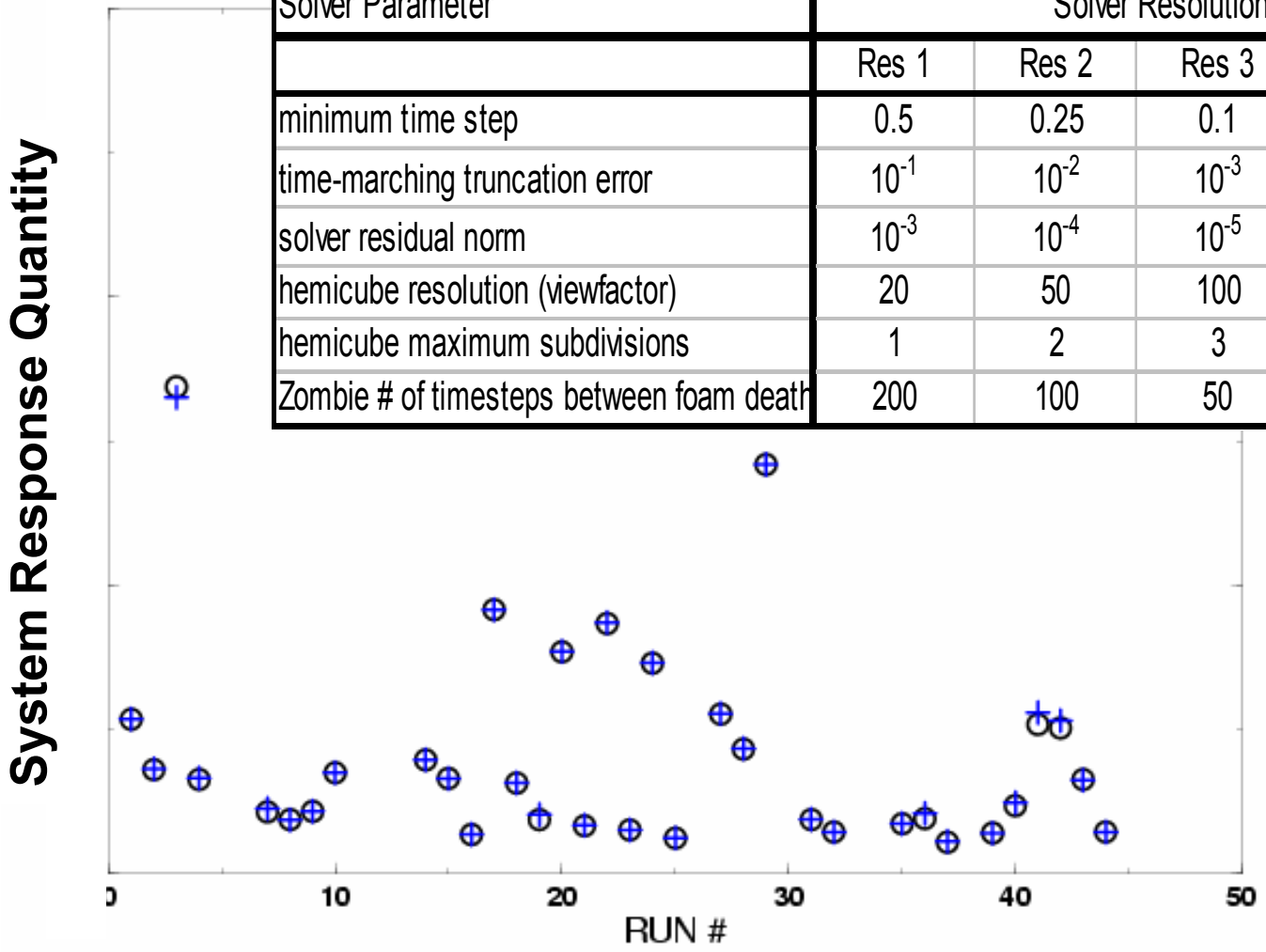
Ryan Maupin, ESA-WR, LANL: IMAC-XXIV 1/31/06

Threaded assembly



Solution Verification Must Address Solver Settings as Well as Discretization Parameters

Solver Parameter	Solver Resolution Settings				
	Res 1	Res 2	Res 3	Res 4	Res 5
minimum time step	0.5	0.25	0.1	0.05	0.01
time-marching truncation error	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
solver residual norm	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
hemicube resolution (viewfactor)	20	50	100	200	300
hemicube maximum subdivisions	1	2	3	4	5
Zombie # of timesteps between foam death	200	100	50	20	1



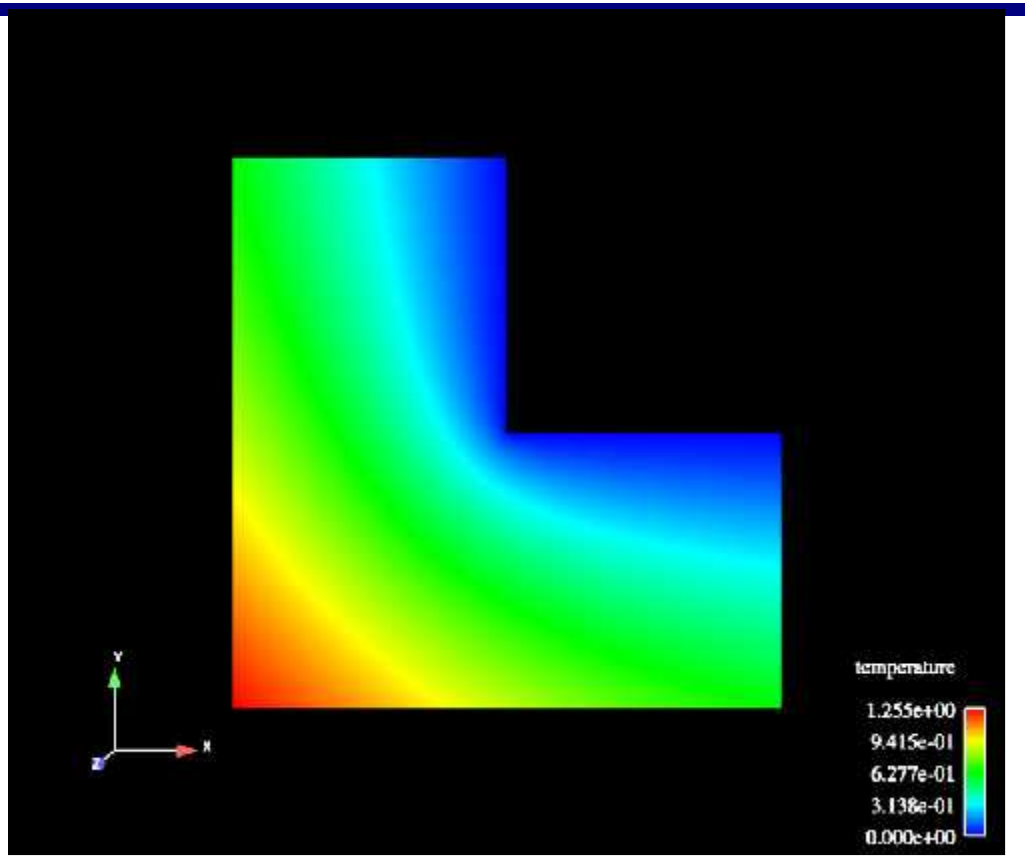
Verification of Error Estimator and Adaptive Algorithm

- 2D Exact Solution:

$$u = r^{2/3} \sin\left(\frac{2}{3}\theta\right)$$

- Linear elements
- ZZ error estimator

- Feedback adaptive algorithm:

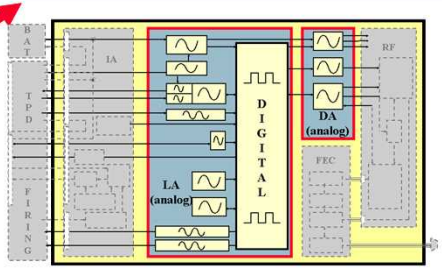


$$\text{if } \left(||e^*||_{H_1(\omega_j)} > 0.995 \max_{1 \le i \le N_\omega} ||e^*||_{H_1(\omega_i)} \right) \text{ then refine } \omega_j$$

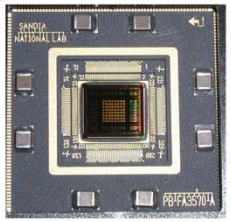
A Hierarchy of Science-Based Validation Experiments Ensures Models Get the Right Answer for the Right Reasons

Hierarchal Validation: Right answer for the right reason

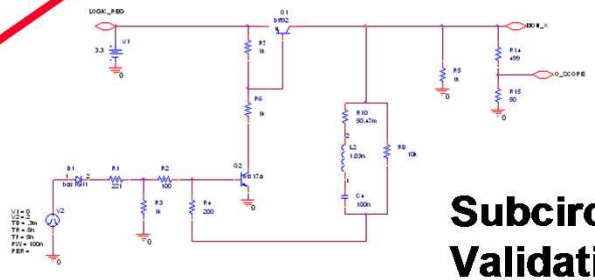
Increasing complexity, Decreasing number of tests



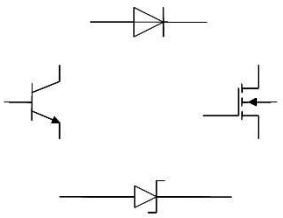
System-Level Circuit Validation



Single ASIC Validation



Subcircuit Validation



Single Device Characterization and Validation

- **Application relevant parameter space**
- **Formal DOE and replicate tests**
- **Attention to diagnostic bias and precision**

A Hierarchy of Science-Based Validation Experiments Ensures Models Get the Right Answer for the Right Reasons

Hierarchal Validation: Right answer for the right reason

Increasing complexity, Decreasing number of tests

Full System Test

Mockup with jointed structure and foam embedded object

Jointed structure validation

Single joint validation

Joint parameter characterization

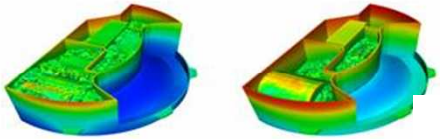
- Application relevant parameter space
- Formal DOE and replicate tests
- Attention to diagnostic bias and precision

A Hierarchy of Science-Based Validation Experiments Ensures Models Get the Right Answer for the Right Reasons

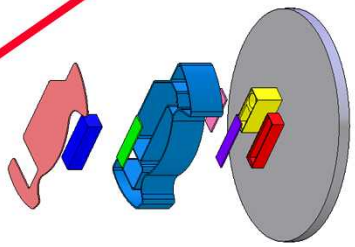
Hierarchal Validation: Right answer for the right reason

Full System Test

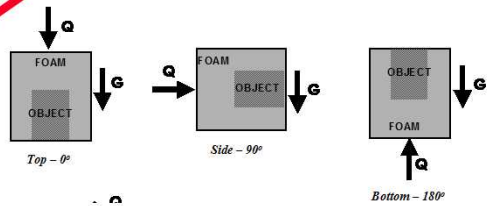
Increasing complexity, Decreasing number of tests



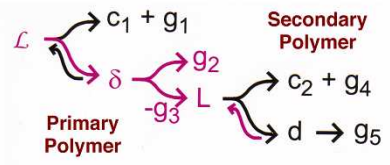
Validation Real Sub-systems



Validation with mockups



Foam recession

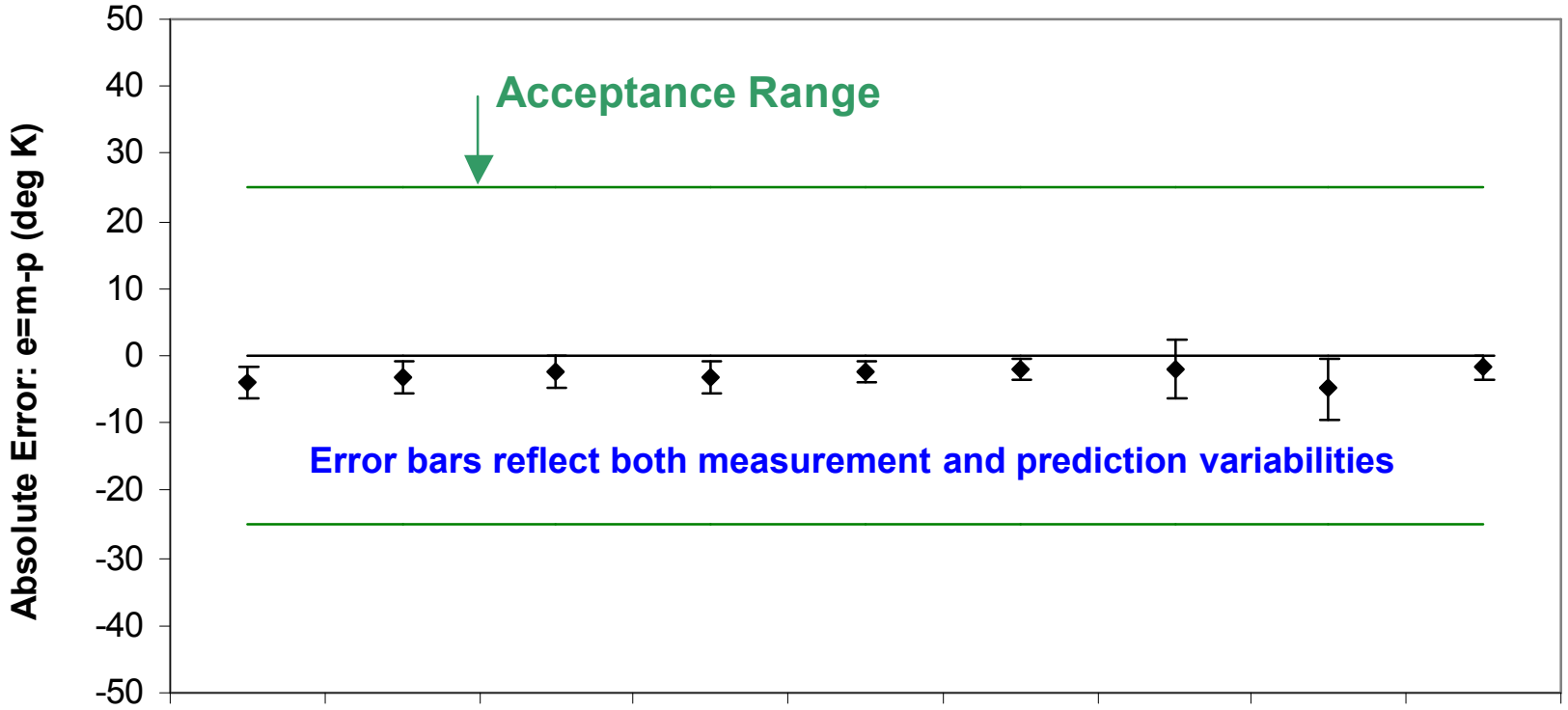


Chemistry characterization/validation

- **Application relevant parameter space**
- **Formal DOE and replicate tests**
- **Attention to diagnostic bias and precision**

Well Established Physics Fidelity

e~2K for conduction

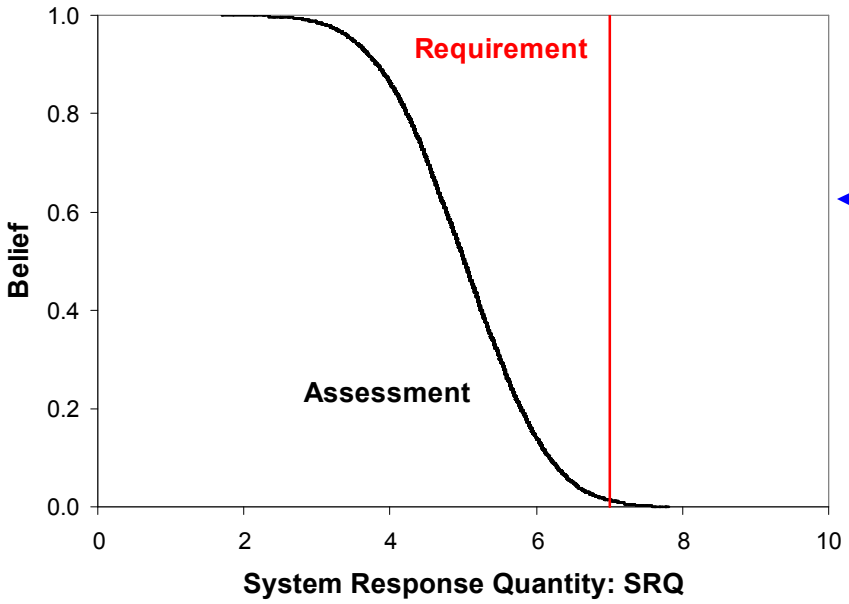


Distinguish Between Aleatory and Epistemic Uncertainties

- **Aleatory uncertainty**: Inherent randomness in behavior of system under study **(frequency interpretation)**
  - Alternatives: Variability, stochastic uncertainty, irreducible uncertainty, type A uncertainty
  - Examples: component failures or material properties derived from statistically significant testing under conditions relevant to intended application
- **Epistemic uncertainty**: Lack of knowledge about appropriate value to use for a quantity that is assumed to have a fixed value in the context of a specific analysis **(confidence or belief interpretation)**
  - Alternatives: state of knowledge uncertainty, subjective uncertainty, reducible uncertainty, type B uncertainty
  - Examples: representative scenarios, unknown parameters in frequency distributions, parameters or models with defensible bounds but no sense of frequency

Infer From Epistemic Results Only What Is Justified

$$SRQ = \sum_{i=1}^{10} X_i \qquad X_i = [0,1] \qquad \text{Requirement}: SRQ \leq 7$$



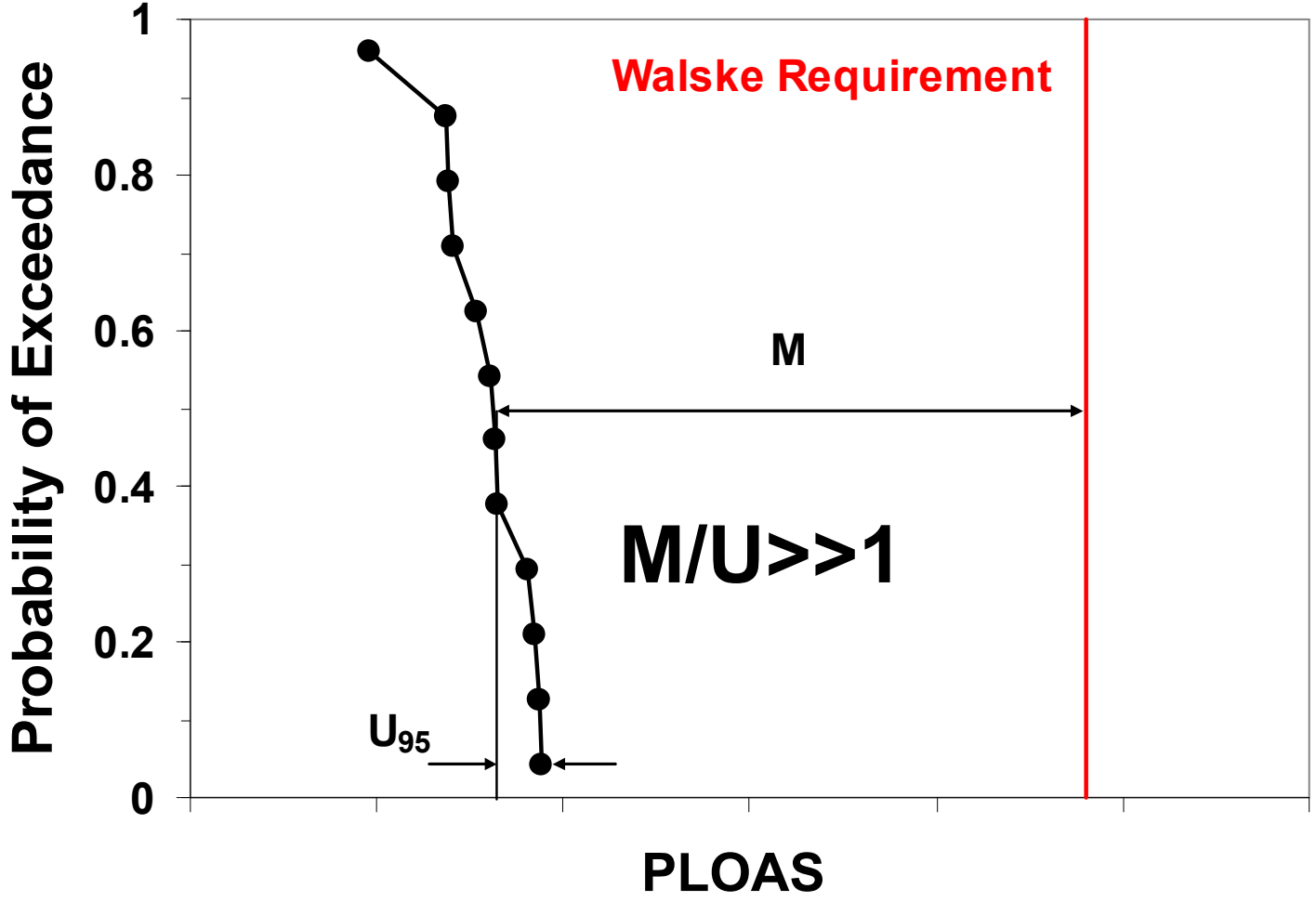
?

$$SRQ \equiv [0,10]$$

$$\text{Belief}(SRQ > 7) = 0.014 \qquad \text{Belief}(SRQ > 7) = 1$$

Uncertainty propagation should not make something from nothing

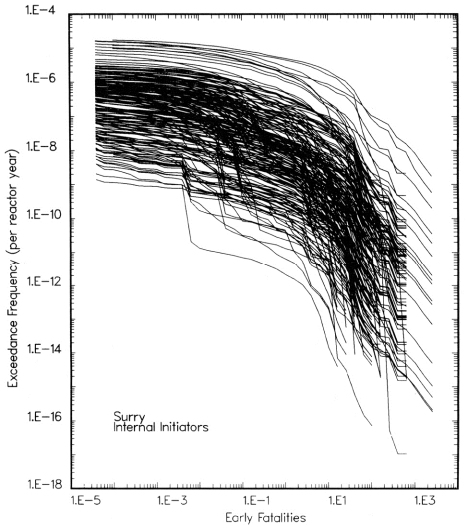
Quantifying Margins and Uncertainties (QMU aka QRA) Supports Risk-Informed Decisions



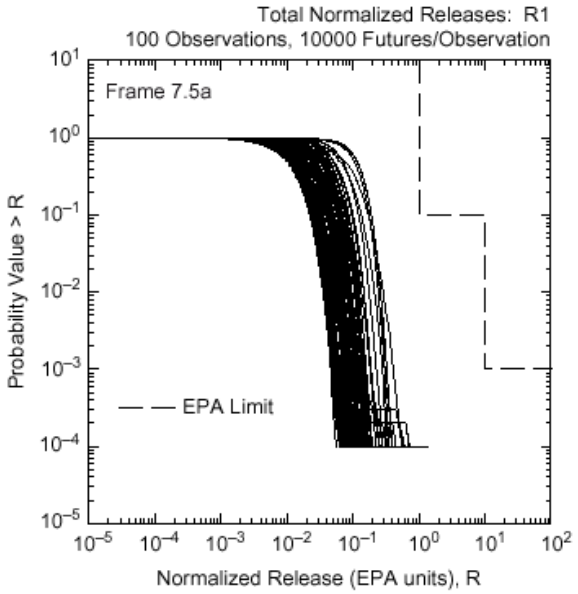
QRA = Quantitative Risk Assessment

Sandia and the Nation Has Significant Experience In Quantitative Risk Assessment (QRA)

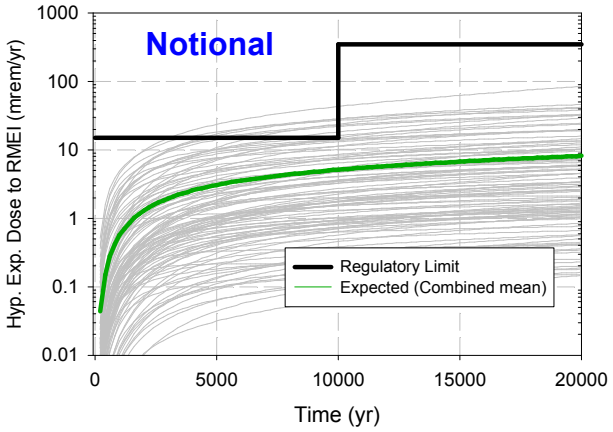
Reactor Safety NUREG-1150: 1990



Waste Isolation Pilot Plan (WIPP) 1999



Yucca Mountain Project (YMP): Present



- QRA is the scientific methodology for addressing these high-consequence M&S-centric issues of national interest