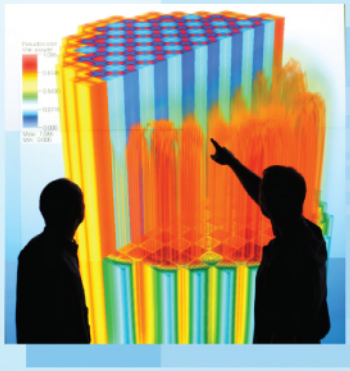
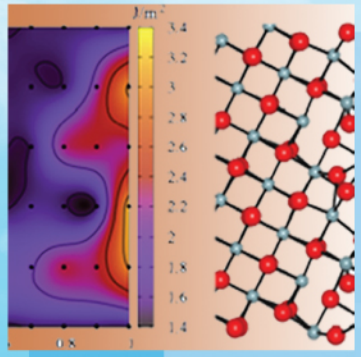
CASL-U-2010-0047-000

Power uprates
and plant life extension



Engineering design
and analysis

Science-enabling
high performance
computing



Fundamental science

Plant operational data

Consortium for Advanced Simulation of LWRs

Gaining Confidence in Computational Simulation

Measured Credibility, on Demand, for Stockpile Applications - CASL Project Powers Nuclear Energy Resurgence

Martin Pilch

Sandia National Laboratory

July 15, 2009

CASL-U-2010-0047-000

U.S. DEPARTMENT OF ENERGY | Nuclear Energy

Unclassified, Unlimited Release

Gaining Confidence in Computational Simulation

Measured Credibility, on Demand, for Stockpile Applications

Martin Pilch PhD, PMP

QMU and Management Support, 1551

Advanced Simulation and Computing Program

mpilch@sandia.gov

Presented to:

ASC Management Team

Sandia National Laboratories

July 15, 2009

CASL-U-2010-0047-000

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.

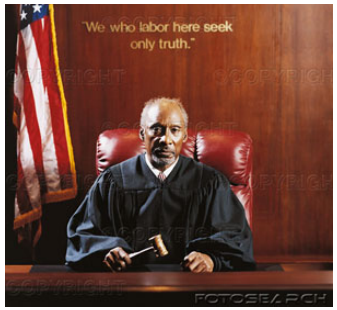
Predictive Capability Maturity Model (PCMM)

- Decision support
- Brief history of PCMM
- PCMM in a nutshell
  - Why I care
  - What PCMM is and is not
- What's “good enough”?
- Discussion

Begin With The End In Mind

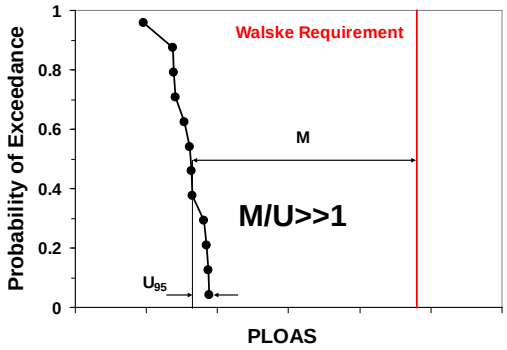
QMU-Informed Decisions Enabled by Advanced Simulation and Computing (ASC)

Testing weapons safety in a Fuel Fire



M&S supports risk-informed decisions:

PLOAS<10^{-6}?

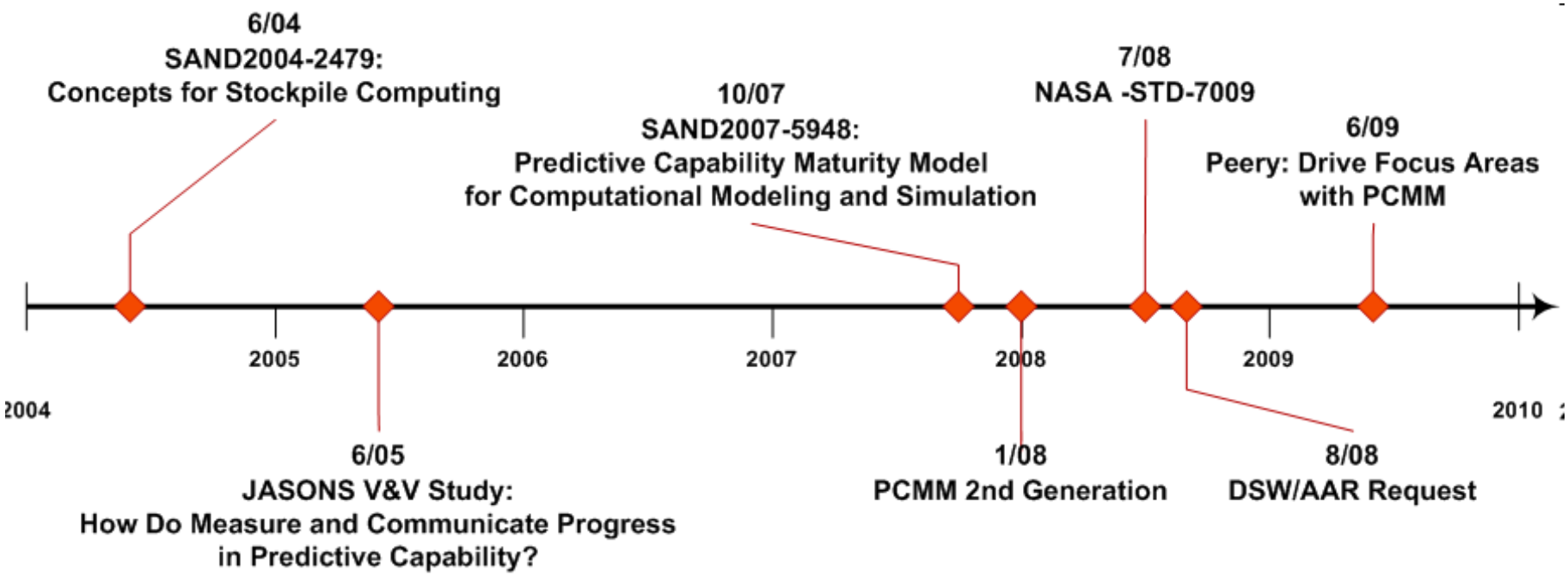


Quantified Margins and Uncertainties

Credibility That is Measured and Communicated

CASL-U-2010-0047-000

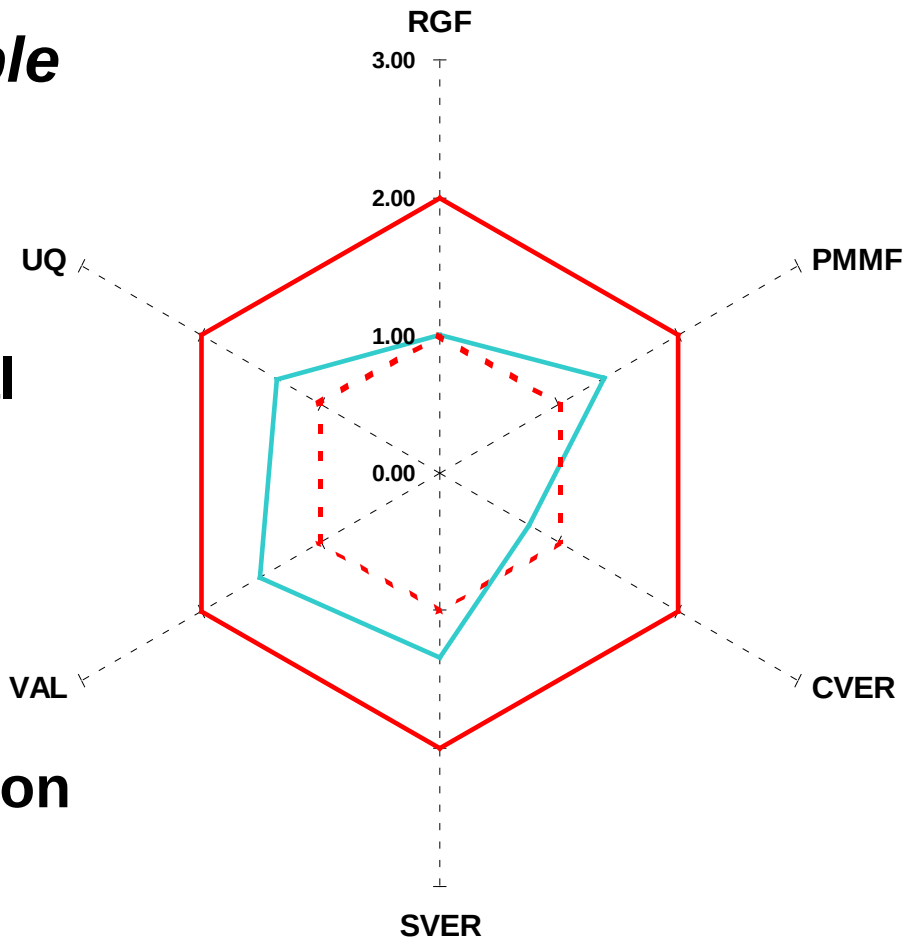
The History of PCMM is Not Fully Written



What Makes M&S Results Worthy of Confidence?

Processes that support *Credible Predictive Capability*

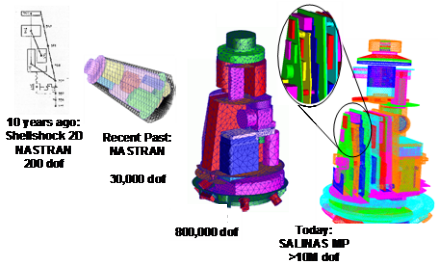
1. **RGF: Representation or geometric fidelity**
2. **PMMF: Physics and material model fidelity**
3. **CVER: Code verification**
4. **SVER: Solution verification**
5. **VAL: Validation**
6. **UQ: Uncertainty quantification**

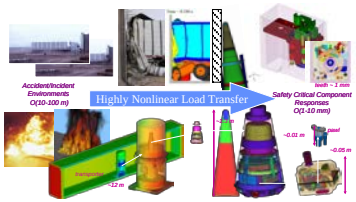


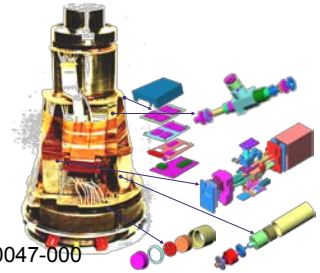
Representation or Geometric Fidelity

Are representation errors corrupting simulation results?

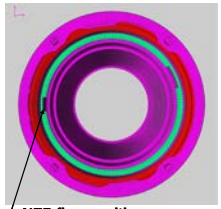
- **Characterize RGF**
  - **How close to "as built" you are representing the system?**
- **Quantify computation errors**
  - **What impact does imperfect RGF have on simulation results?**
- **Verify representation or geometry**
  - **Is what you represented really what was built?**







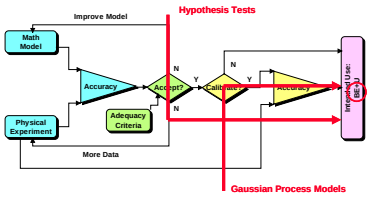
CASL-U-2010-0047-000



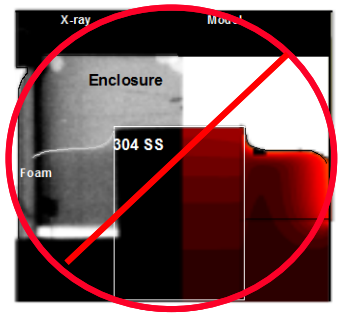
NEP flange with cable runs included

Physics and Material Model Fidelity

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

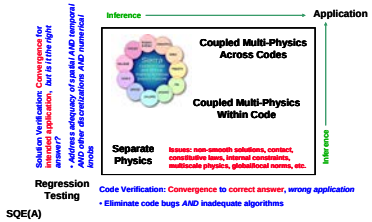


- **Characterize science basis for the models**
  - **Are the "models" best described as "knobs", empirical correlations, physics-informed, or fundamental physics?**
- **Quantify model accuracy**
  - **How accurate are the models?**
- **Assess the degree of interpolation or extrapolation**
  - **What 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**



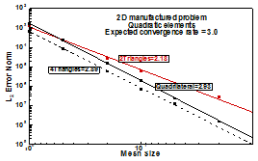
Code Verification

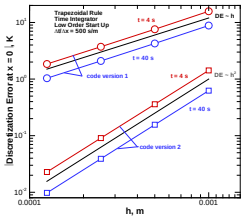
Are software errors or algorithm deficiencies corrupting simulation results?

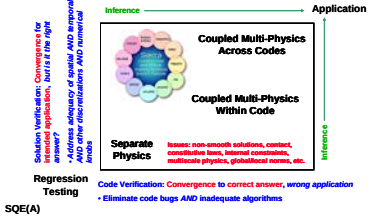


- **Apply good SQE processes**
  - **Do you have a mature code development process?**
- **Assess SQE processes**
  - **Verify that codes are developed with an appropriate level SQE maturity?**
- **Provide adequate test coverage**
  - **Can the user be confident that the code is adequately tested for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

Code1:Code2 Comparisons



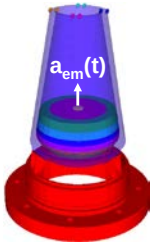


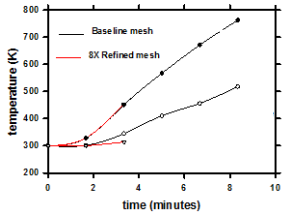


Solution Verification

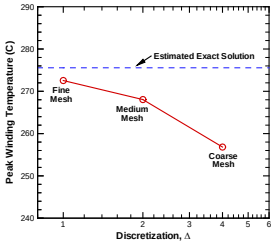
Are human procedural errors or numerical solution errors corrupting simulation results?

- **Quantify numerical solution errors**
  - **What is the impact of numerical solution errors on relevant system response quantities (SRQs)**

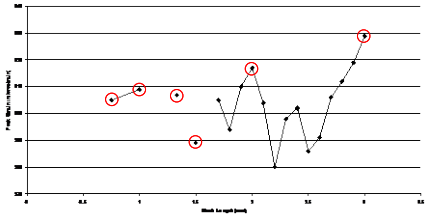


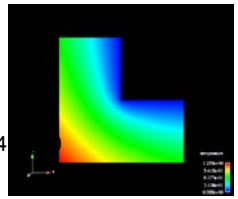


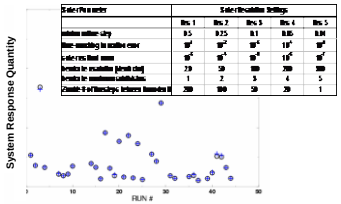
- **Verify all simulation inputs and outputs**
  - **Have we corrupted simulation results with incorrect inputs or post processing errors?**

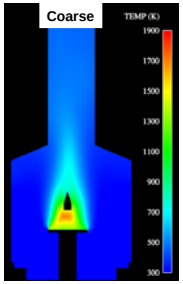


- **Perform technical review**
  - **Verify that the solution verification activities are relevant, adequate, and executed in a technically sound manner**





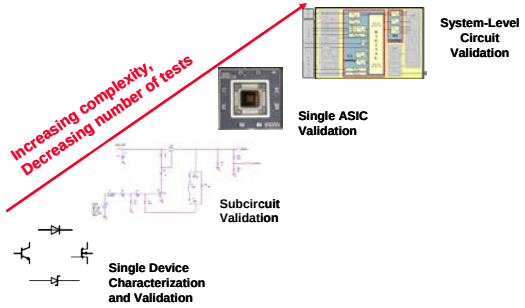




Validation

How accurate are the integrated physics and material models?

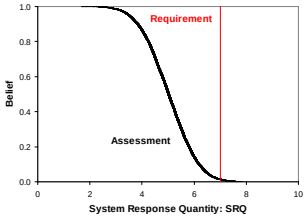
- Apply a validation hierarchy
  - Are you getting the right answers for the right reasons?
- Quantify model accuracy
  - How accurate are the models?
- Assess the degree of interpolation or extrapolation
  - What is the relevance of the underlying databases?
- Perform technical review
  - Verify that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

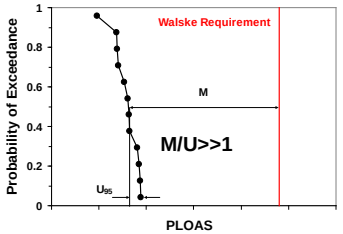


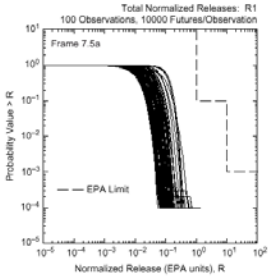
Uncertainty Quantification

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?

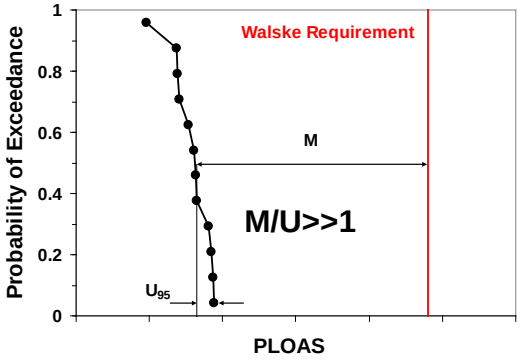
- **Assess completeness**
  - **Do we cast a broad enough net that all potentially significant sources of uncertainty or error are quantified or otherwise dealt with?**
- **Avoid strong assumptions**
  - **Do strong assumptions corrupt the accuracy of UQ/SA results?**
- **Perform technical review**
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

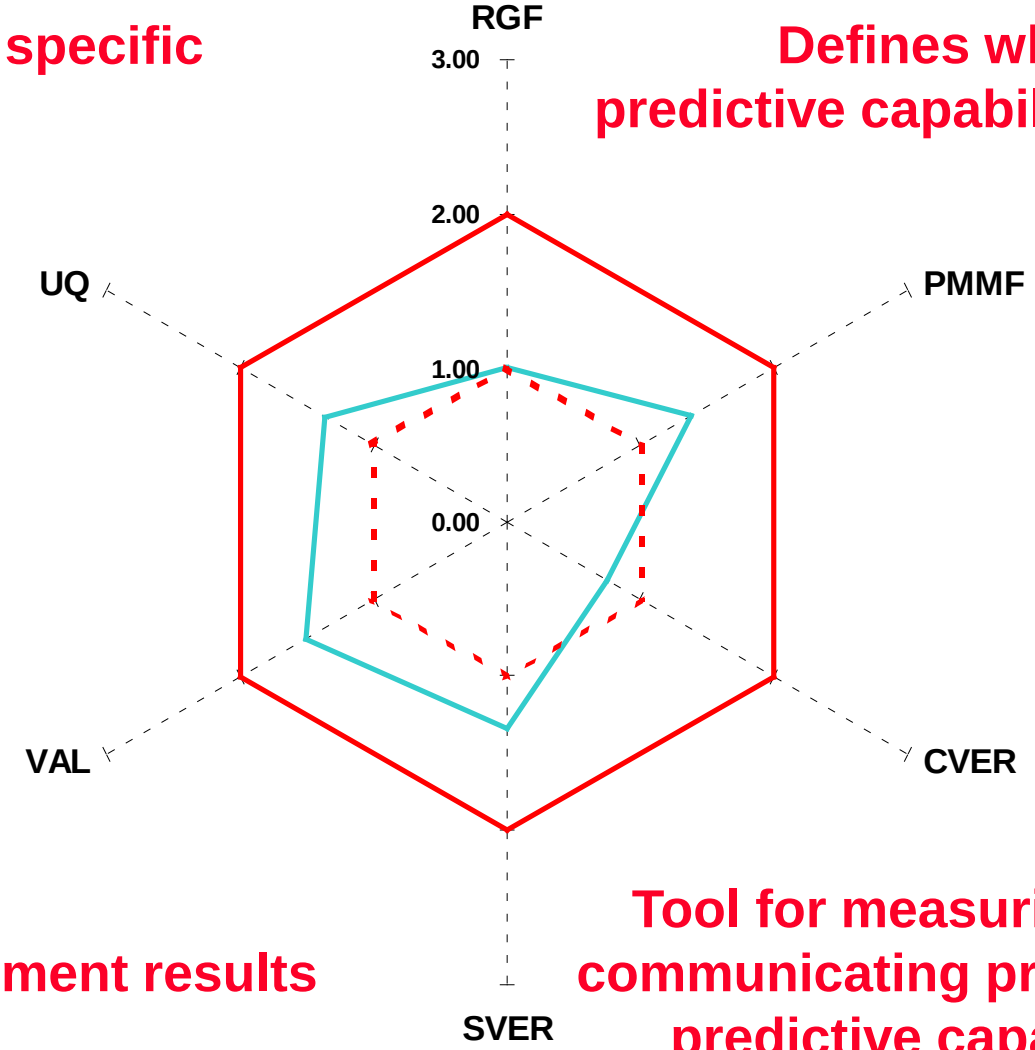
PCMM is not a Number or a Score

PCMM *is* a Communication Tool…

PCMM is application specific

PLOAS<10^{-6}?



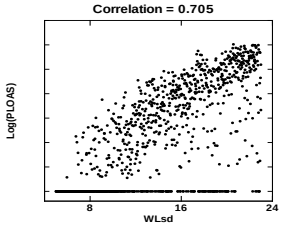


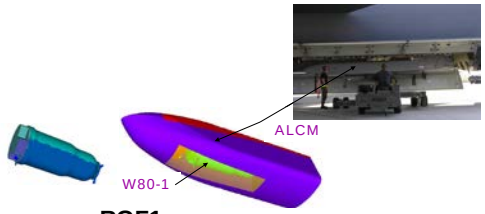
Defines what predictive capability means

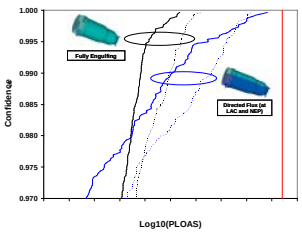
Compliments assessment results

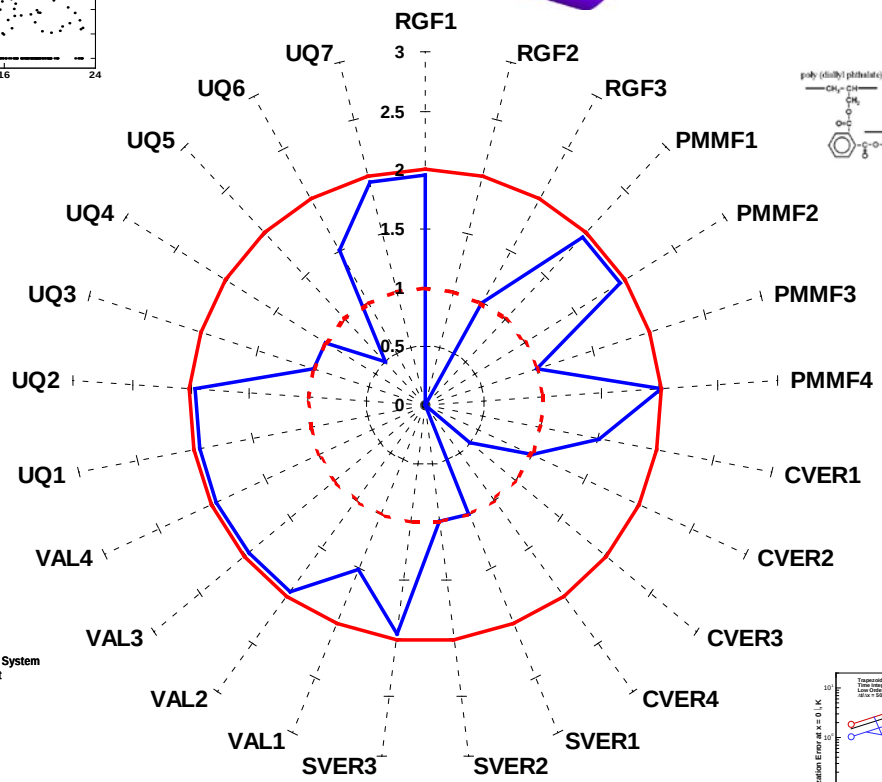
Tool for measuring and communicating progress in predictive capability

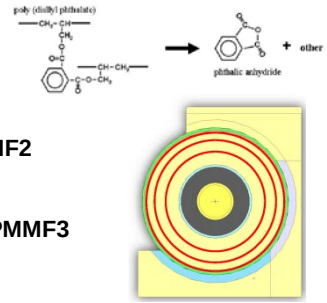
…That *Must* Include a Discussion of the Supporting Evidence

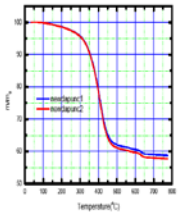


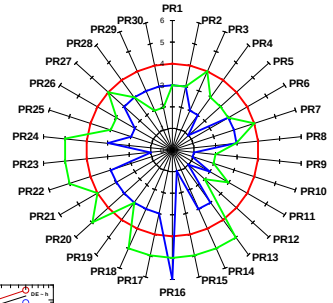


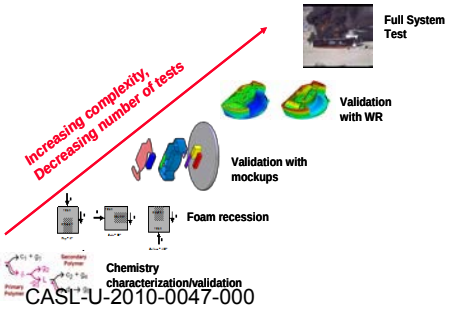


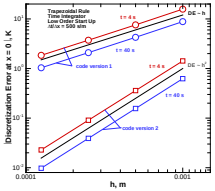


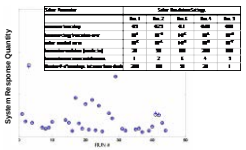


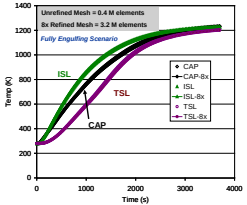


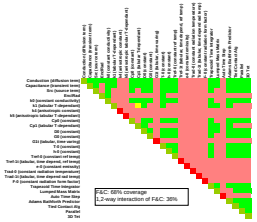








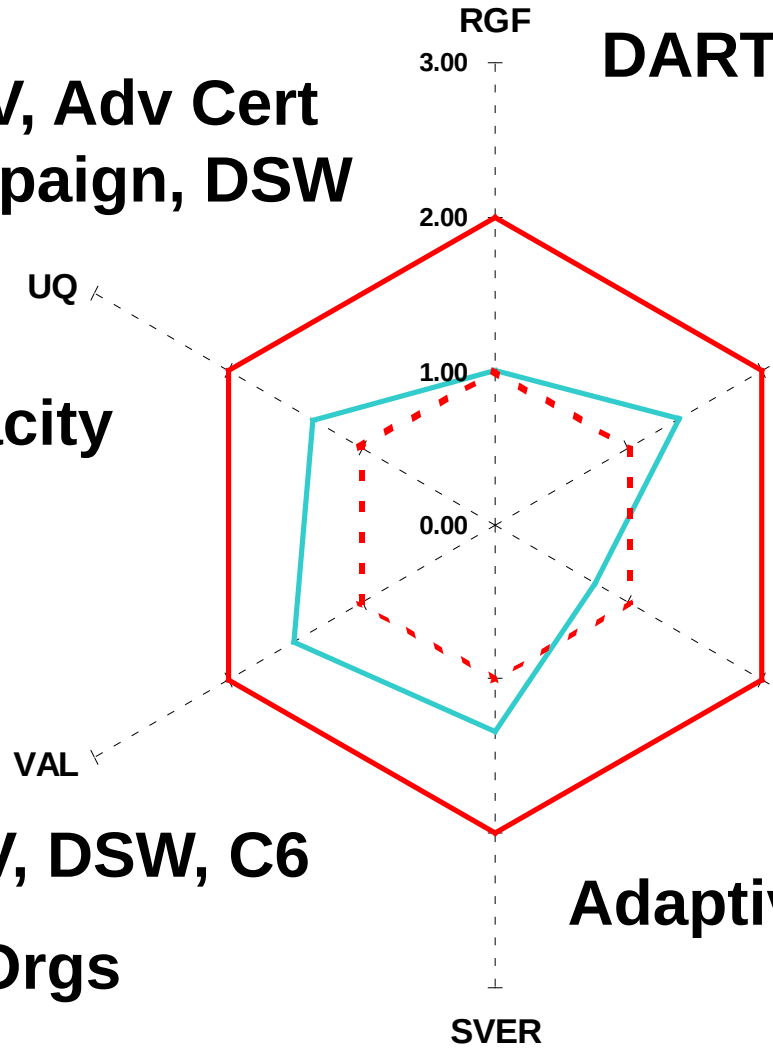




PCMM is a Tool for Managing Risk in the Use of Modeling and Simulation

- PCMM helps us avoid four types of errors in our computational modeling
  - Type 1 error: Believing that a correct model is incorrect
  - Type 2 error: Believing an incorrect model is correct
  - Type 3 error (Balci): Solving the wrong problem
  - Type 4 error (Trucano): Using computational information incorrectly

PCMM Must be an Organizational Value in Order to Coordinate Capability Development with Stockpile Delivery



DART

V&V, Adv Cert Campaign, DSW

PEM, C6, DSW

Capability/Capacity Computing

IC

V&V, DSW, C6

Adaptivity

Analysis Orgs

We all Avoid This Question

How Much is Enough?

How Much is Enough?

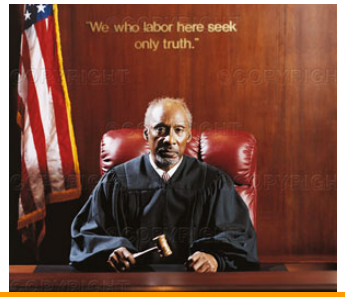
Can Only be Answered in a Decision Context

Greenaugh: DSW schedules are not based on delivery of M&S capabilities. We will do stockpile work with the best tools available at the time.

Testing weapons safety

What other sources of information do we have?

What does it cost and how long will it take to get



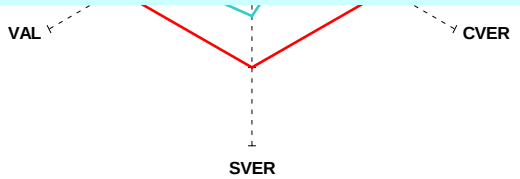
DSW at Sandia is not a sophisticated customer of M&S

PLOAS<10^{-6}?

The burden rests with ASC to show relevance, demonstrate impact, and demonstrate the necessary rigor

take to get more information?

Should we change the design or get more information?



PLOAS

Quantified Margins and Uncertainties

Credibility That is Measured and Communicated

How Much is Enough?
From the Perspective of CompSim Stewardship

Increasing completeness and rigor

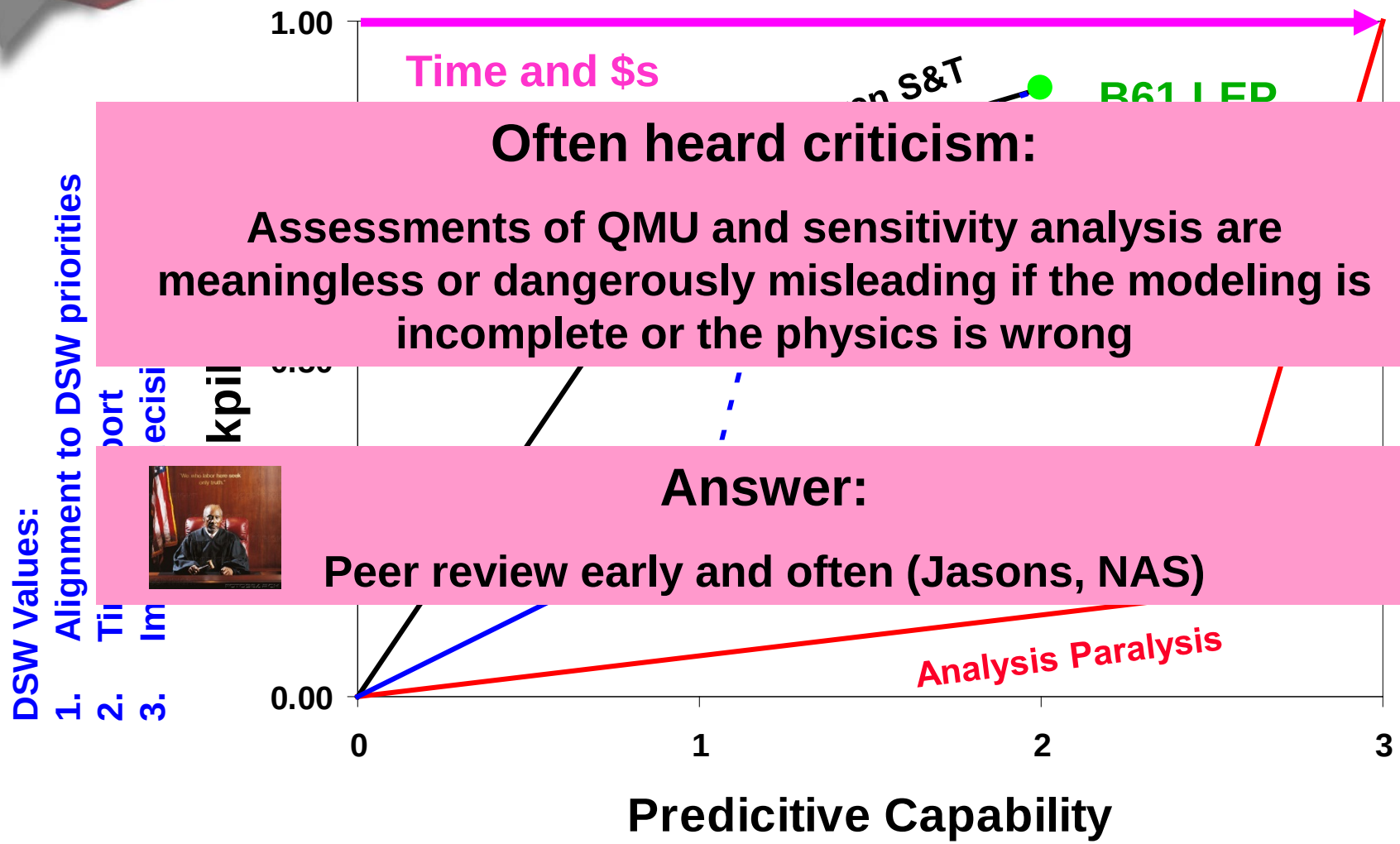
Decreasing risk

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

- **Level 0: Low consequence; minimal M&S impact (e.g., scoping studies)**
- **Level 1: Moderate consequence; some M&S impact (e.g., design support or qualification test support)**
- **Level 2: High consequence, high M&S impact (qualification decision support)**
- **Level 3: High consequence; decision making based predominately on M&S (dominant basis for qualification or certification)**

Set targets and schedules, measure and communicate progress regularly

QMU-Driven S&T is in the Best Interest of ASC and DSW



Often heard criticism:

Assessments of QMU and sensitivity analysis are meaningless or dangerously misleading if the modeling is incomplete or the physics is wrong

Answer:

Peer review early and often (Jasons, NAS)

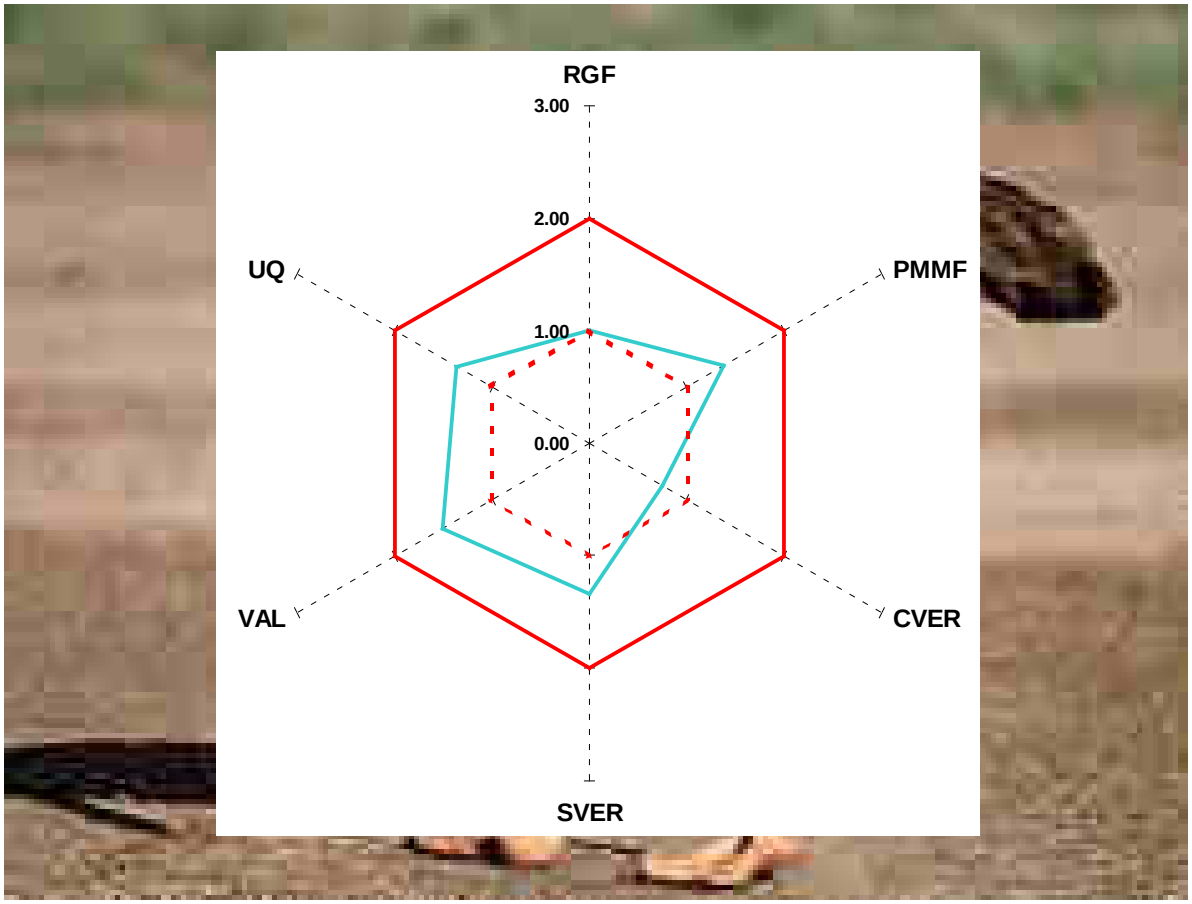
ASC Values:

1. **Development and use of HPC**
2. **Recognized leadership in Science and Engineering capabilities**
3. **Recognized leadership in V&V/QMU methods**

Some Practical Recommendations

- Align ASC capabilities to DSW priorities, provided they are ASC appropriate
  - We strive to do this through the focus areas
- Demonstrate “system-level” assessments ASAP
  - You can't assess QMU and sensitivities until you have an integrated assessment capability
  - Stanford was challenged with this and they responded
  - We need to move faster and more broadly on this
- Adopt PCMM as an organizational value
  - Make expectations clear, drive focus area planning with PCMM
- Establish peer review early and often for the focus areas
  - This needs to play out on several levels: PESP, L2’s, AAR contributions
- Measure and communicate progress with PCMM on a yearly basis
  - Establish targets, be neither optimistic nor pessimistic in assessments, evolve in a balanced manner

You Can't Know How Good or How Bad You Are If You Don't Ask!



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

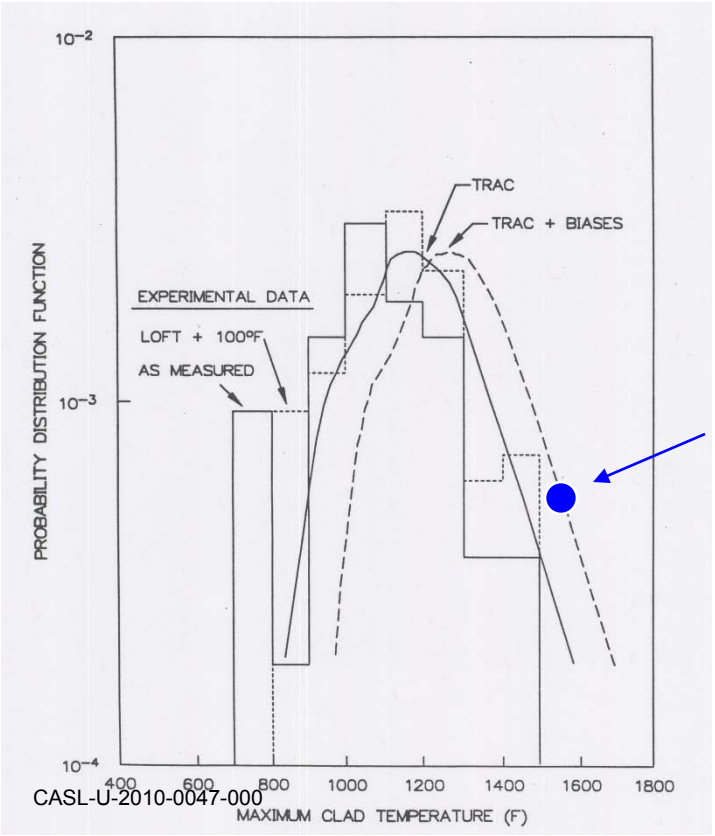
Hyperlinked Pages

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

Quantify Reactor Safety Margins 1988

Application of the Code Scaling, Applicability, and Uncertainty (CSAU) Evaluation Methodology to a LBLOCA

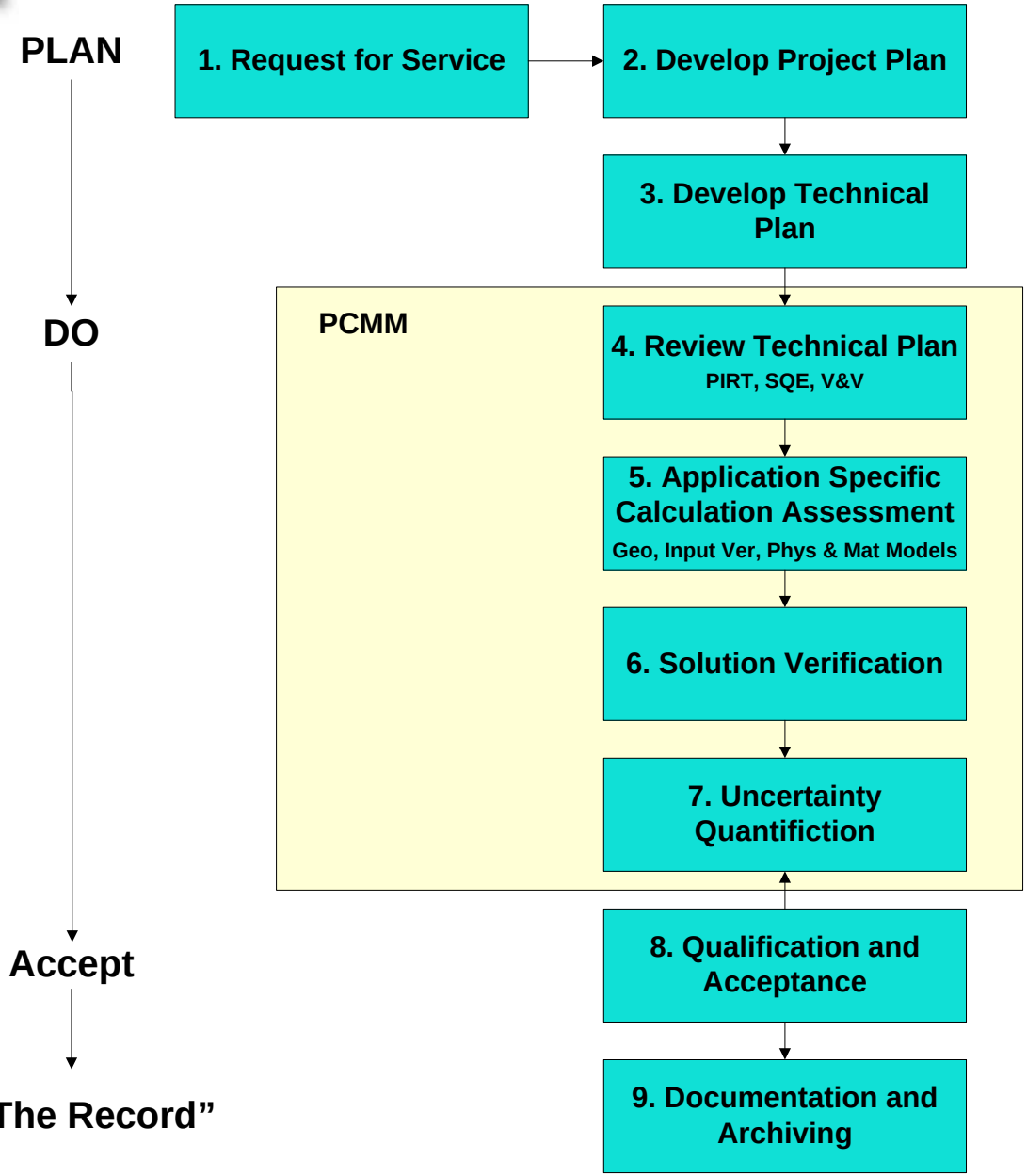
- First acceptance of BE+U for NRC regulatory purposes
- 10 CFR Part 50, Appendix K: requires PCT<2200 F



PCT(95/95) = 1572 F < 2200 F

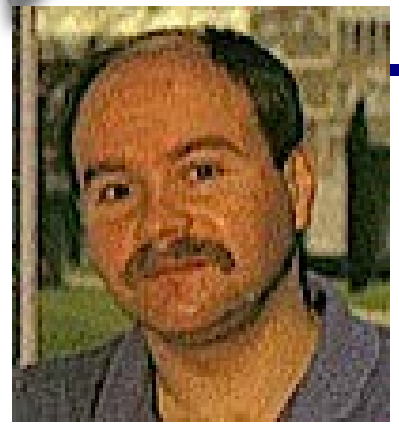
- CSAU Elements:
- Requirements and Code Capabilities
  1. Specify scenario
  2. Select NPP
  3. Identify and Rank Phenomena (PIRT)
  4. Select frozen code
  5. Provide complete documentation, code manual, user guide, developmental assessment, model and correlation QE
  6. Determine code applicability
- Assessment and ranging of parameters
  7. Establish assessment matrix
  8. Compare calculations against SET/IET and document
  9. Determine code and experiment accuracy
  10. Determine effect of scale
- Sensitivity and uncertainty analysis
  11. Determine effect of reactor input parameters and state
  12. Perform NPP sensitivity calculations
  13. Combine biases and uncertainties
  14. Total uncertainty to calculate specific scenario in a specific NPP

Key Attributes of PCMM Where First Discussed in *Concepts for Stockpile Computing* (SAND2004-2479)



2005 JASON V&V Study

How do you measure and communicate progress in predictive capability?



SAND2007-5948: Predictive Capability Maturity Model for Computational Modeling and Simulation

Increasing completeness and rigor →

Decreasing risk

Content ↓

MATURITY / ELEMENT	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 What features are neglected because of simplifications or stylizations?	• Judgment only • Little or no representational or geometric fidelity for the system and BCs	• Significant simplification or stylization of the system and BCs • Geometry or representation of major components is defined	• 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, physics-informed 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&C) 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 an 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 a validation hierarchy?	• Judgment only • Few, if any, comparisons with measurements from similar systems or applications	• Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest • Large or unknown exper-imental 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
Uncertainty Quantification and Sensitivity Analysis How thoroughly are uncertainties and sensitivities characterized and propagated?	• Judgment only • Only deterministic analyses are conducted • Uncertainties and sensitivities are not addressed	• Aleatory and epistemic (A&E) uncertainties propagated, but without distinction • Informal sensitivity studies conducted • Many strong UQ/SA assumptions made	• A&E uncertainties segregated, propagated and identified in SRQs • Quantitative sensitivity analyses conducted for most parameters • Numerical propagation errors are estimated and their effect known • Some strong assumptions made • Some peer review conducted	• A&E uncertainties comprehensively treated and properly interpreted • Comprehensive sensitivity analyses conducted for parameters and models • Numerical propagation errors are demonstrated to be small • No significant UQ/SA assumptions made • Independent peer review conducted

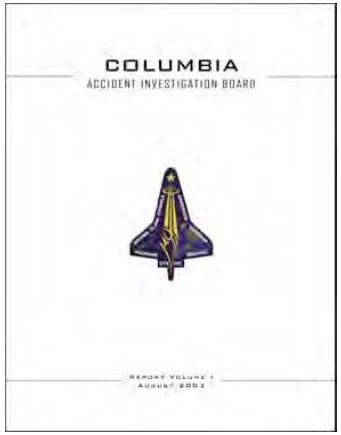
PCMM Second Generation 1/08

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)	• (unqualified) conceptual abstraction of the whole system	• Significant (unqualified) simplification or stylization of the system at the level of major elements	• Limited (unqualified) 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 calibrated to data	• Physics-informed models. Key dependencies derived from fundamental theory and modified or augmented to account for model discrepancies. The calibration of some model parameters may not be unique (i.e., the calibration of parameters in separate effects physics with data from integral tests)	• Physics-based models. Key dependencies derived from fundamental or well accepted theory without need for modification or augmentation. Model parameters derived from theory or uniquely calibrated to relevant data
	Model Accuracy (how accurate are the models) i.e., SET validation	• 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 (CVER) Are software errors or algorithm deficiencies corrupting the simulation results?	Software Quality Engineering practices (SQE: how mature are the SQE practices)	• Judgment only, codes informally managed to SQE practices or no documented SQE process requirements • Software process is characterized as ad hoc, and occasionally even chaotic	• Codes managed to repeatable and 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	• The SQE process is managed • Managed: Detailed measures of software process and product quality are collected. Both the software process and products are quantitatively understood and controlled.	• The SQE process is optimized • 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 • 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	• 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)
Solution Verification (SVER) Are human procedural errors or numerical solution errors corrupting simulation conclusions?	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
	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 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
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 Uncertainties	• PI aggregates evidence for the characterization of 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	• PI aggregates evidence for the characterization of important uncertainties • The supporting evidence, aggregations, and characterizations are formally documented	• Semi-formal elicitation process for key uncertainties that 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	• Formal elicitation process for key uncertainties that 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), RAG 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 undermine the rigor or accuracy of UQ/SA results) [illegible] dependencies, shape of distributions, consistency of measured property variability and underlying assumptions of physical model, BE+U	• Strong assumptions not addressed or acknowledged	• Many strong assumptions	• Some strong assumptions	• No strong assumptions
	Technical Review (has the UQ/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

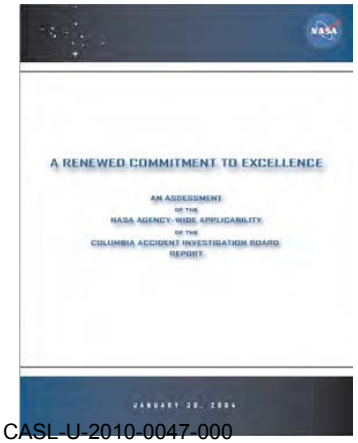
- The finer levels of resolution serve to better define the highest level element
- The finer levels of resolution also serve as evaluation principles for the PCMM
- Version 2 attempts to provide refined levels that are likely meaningful for many ASC (code, application) pairs in the Focus Areas
- Depending on the intended use, the refined levels could be increased for given high level elements, or a subset of them could be used

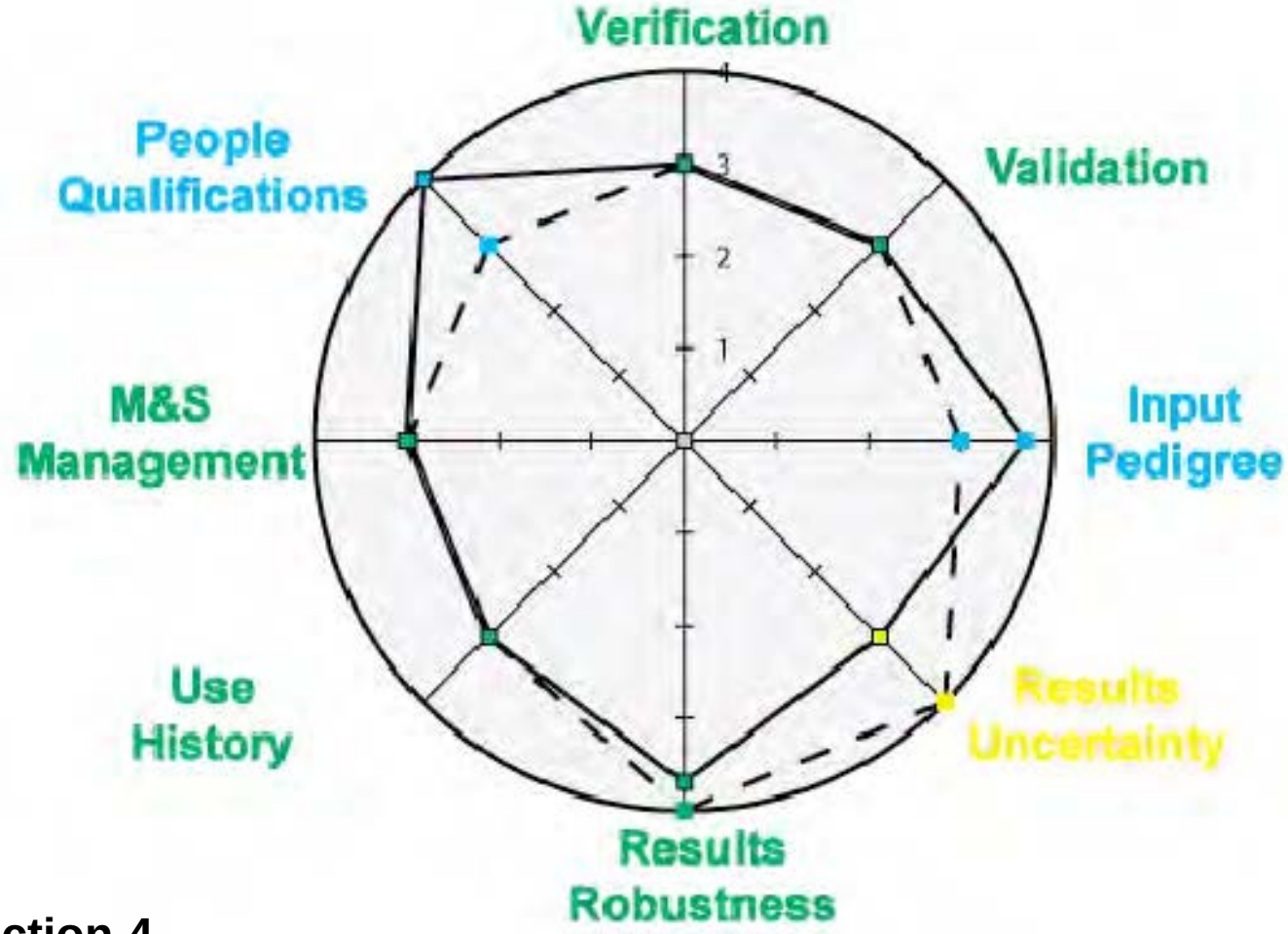
NASA-STD-7009
Standard for Models and Simulations

CAIB



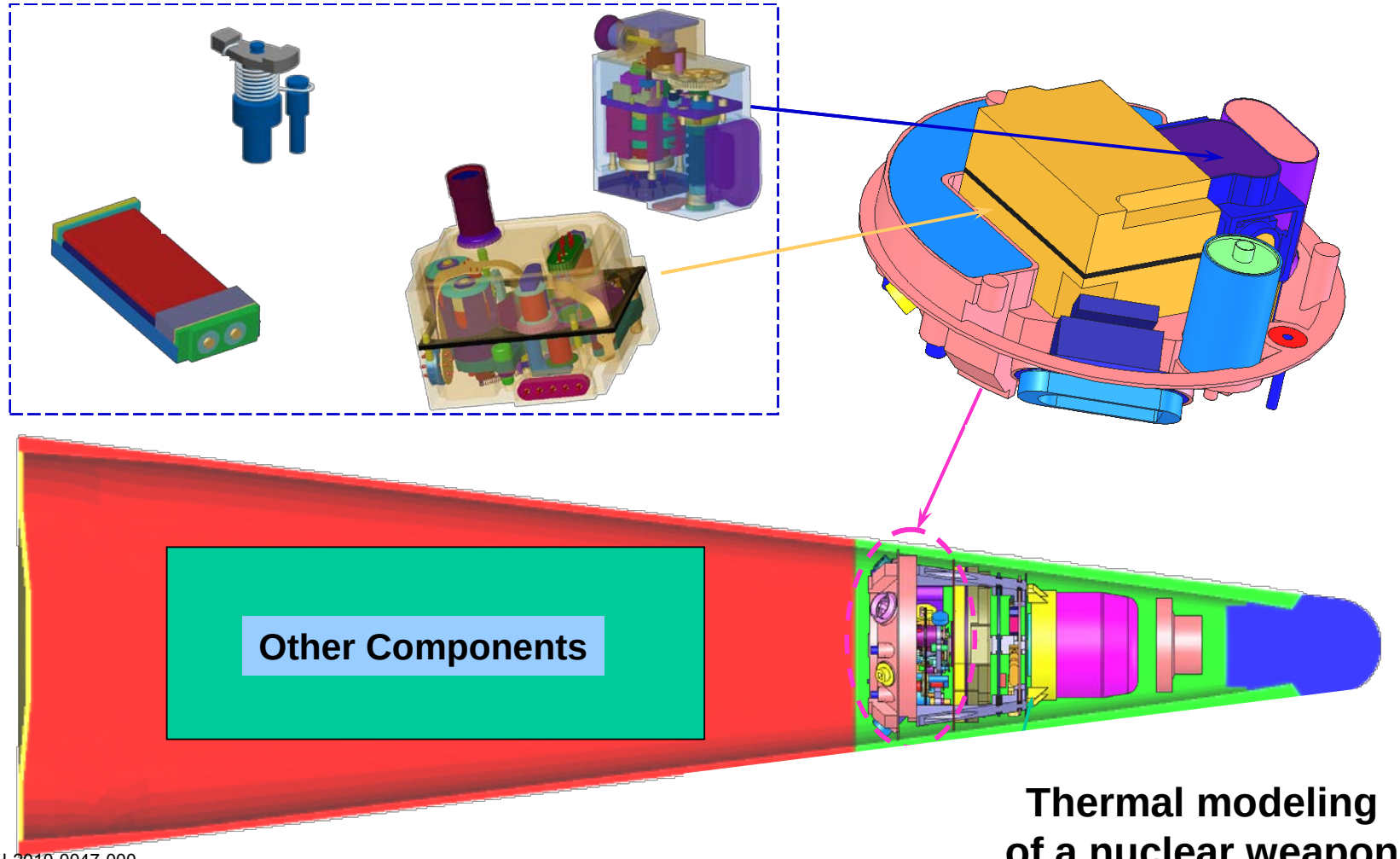
Diaz Report





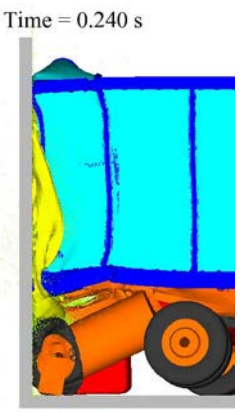
Action 4

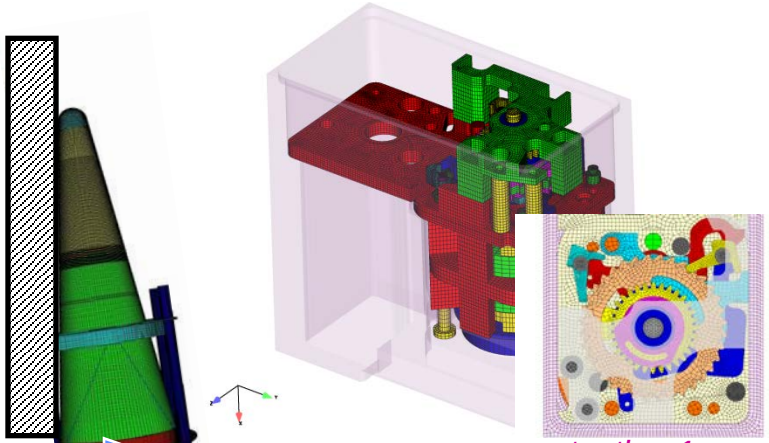
Modern Computing Platforms Enable "~As Built" Geometric Fidelity



Thermal modeling of a nuclear weapon

Modern Computing Platforms Enable (or Will Enable) "~As Built" Geometric Fidelity



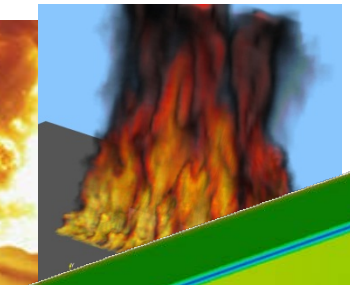


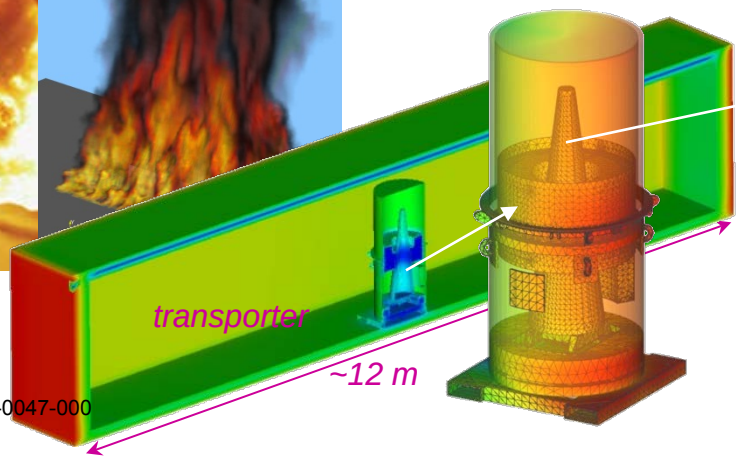
Accident/Incident Environments O(10-100 m)

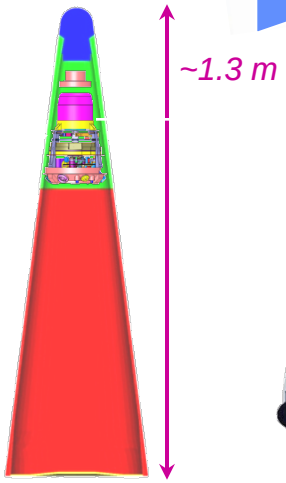
Highly Nonlinear Load Transfer

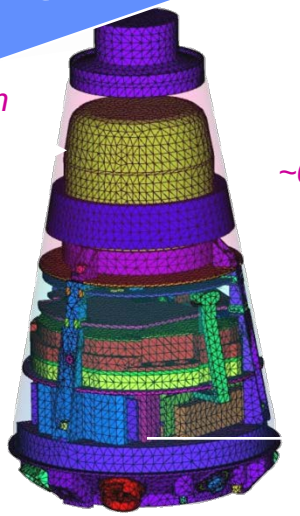
teeth ~ 1 mm

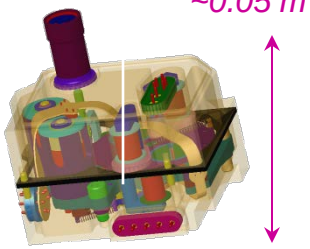
Safety Critical Component Responses O(1-10 mm)



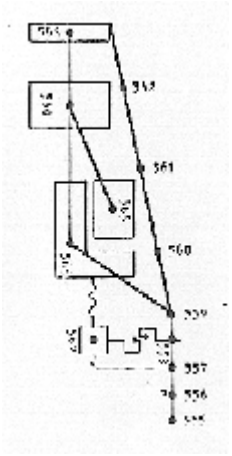




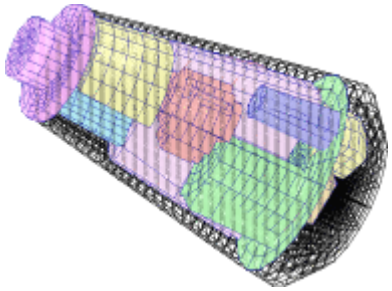




Modern Computing Platforms Enable "~As Built" Geometric Fidelity

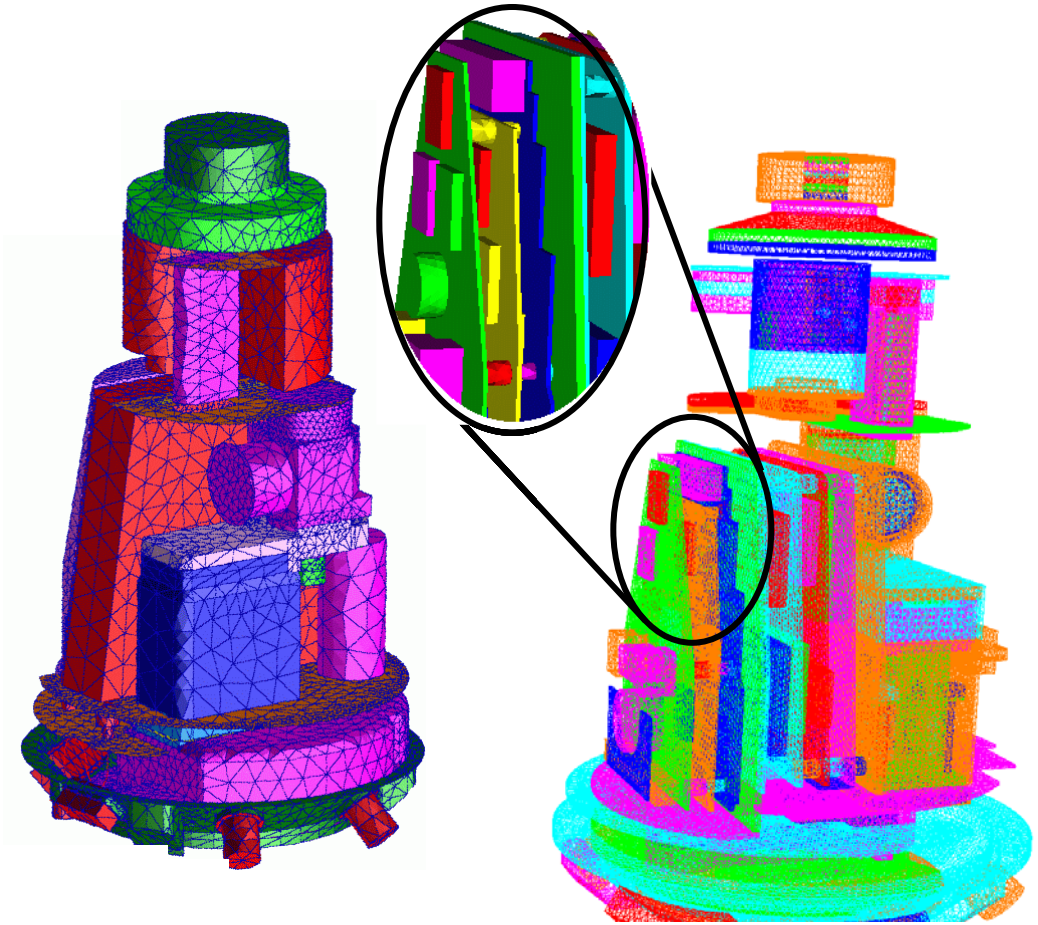


10 years ago:
Shellshock 2D
NASTRAN
200 dof



Recent Past:
NASTRAN

30,000 dof



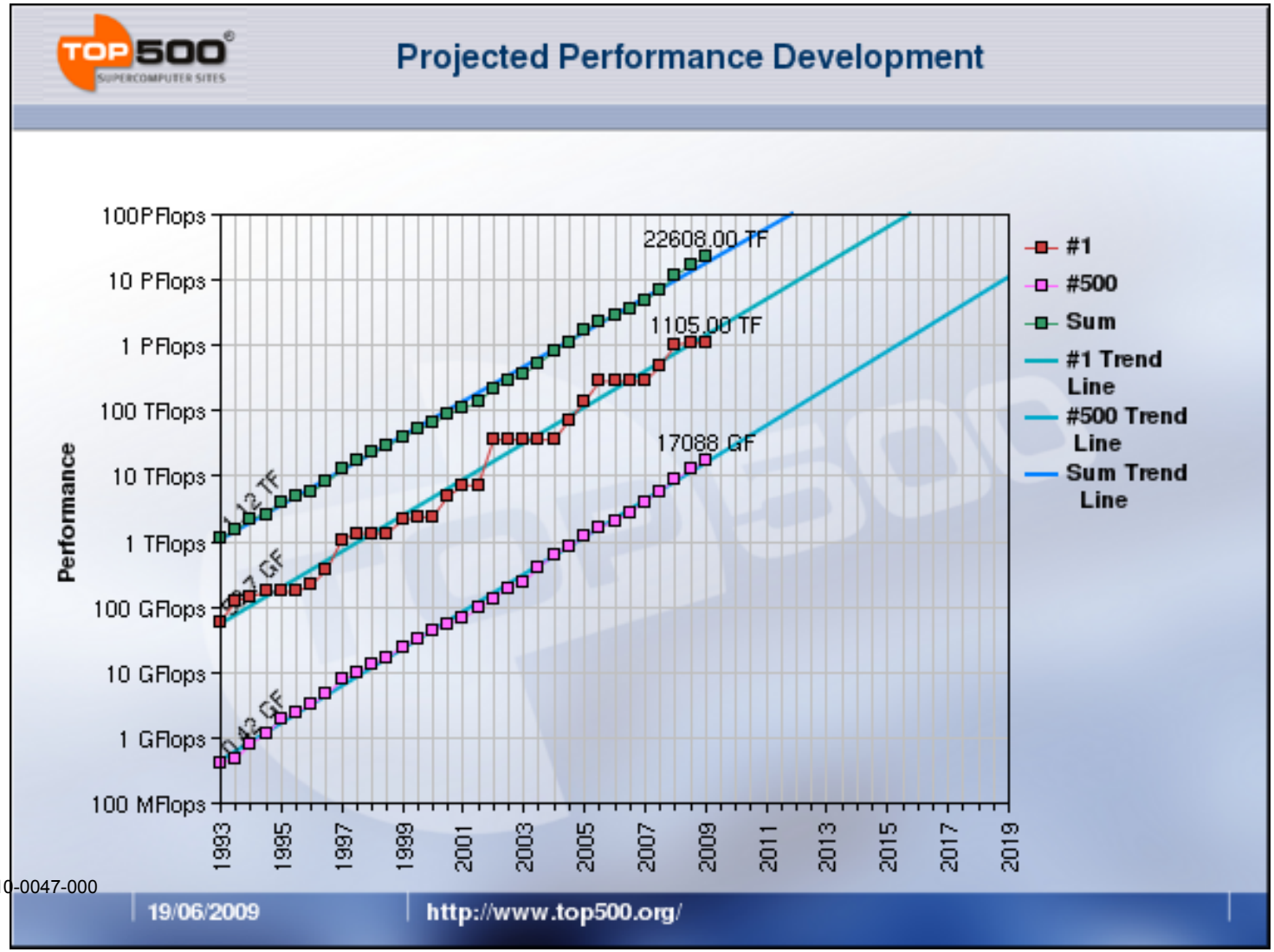
800,000 dof

Today:
SALINAS MP
>10M dof

Structural dynamics

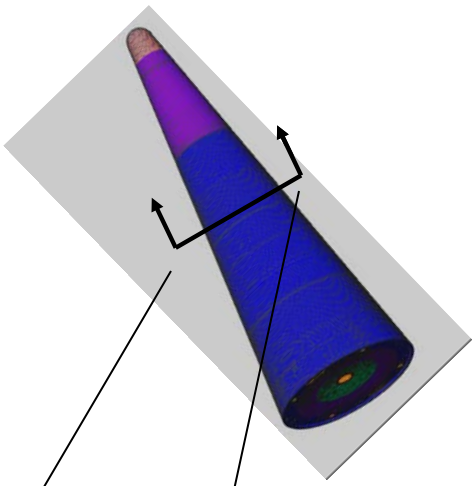
CASL-U-2010-0047-000

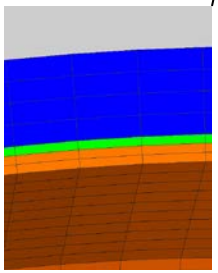
If Not Today, Then Tomorrow



Calibration Hides Errors Resulting From Inadequate Geometric Fidelity

Re-Entry Body

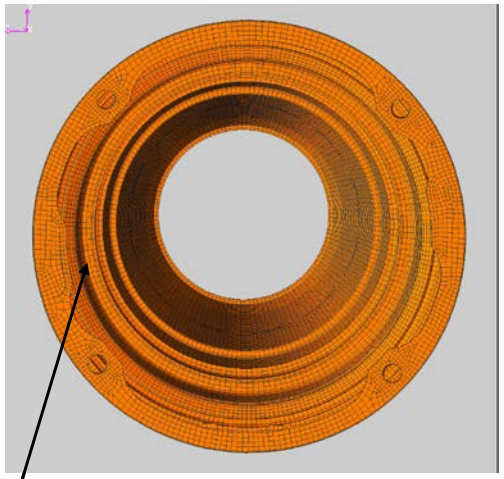




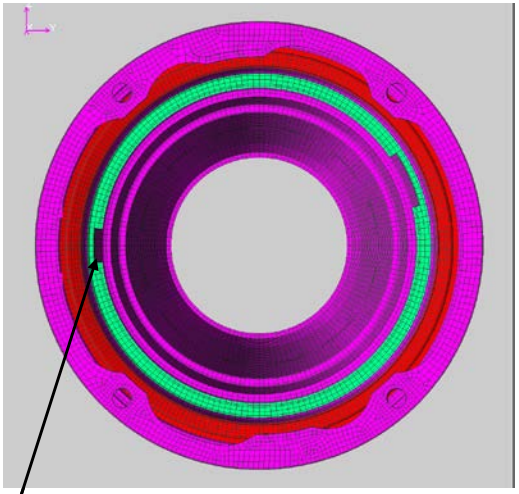
Heat Shield

Bond

Substructure

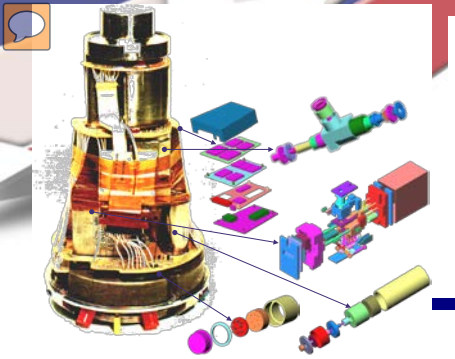


NEP flange with cable runs defeatued



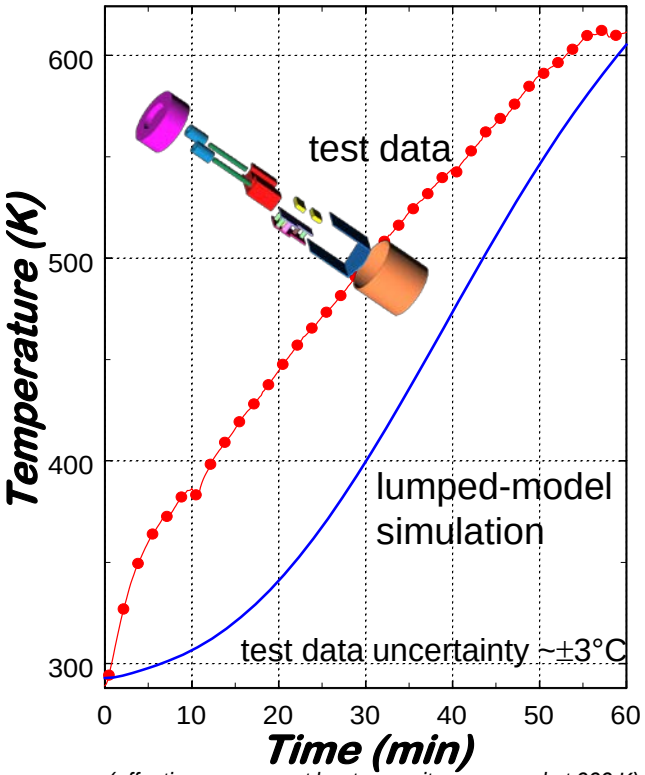
NEP flange with cable runs included

Mode	Meas. Freq. (Hz)	E(%)=(P-M)/M		
		Defeatured	Unphysical Calibraton	Full Feature
1st Ovaling	198	34.9	-4.5	-3.5
1st Orthog. Ovaling	208	38.0	-5.1	-3.8
1st Bending	978	6.4	-2.6	-0.7
1st Orthog Bending	1008	3.2	-4.9	-2.3



High Geometric Fidelity Allows Identification of Inadequate Physics Models

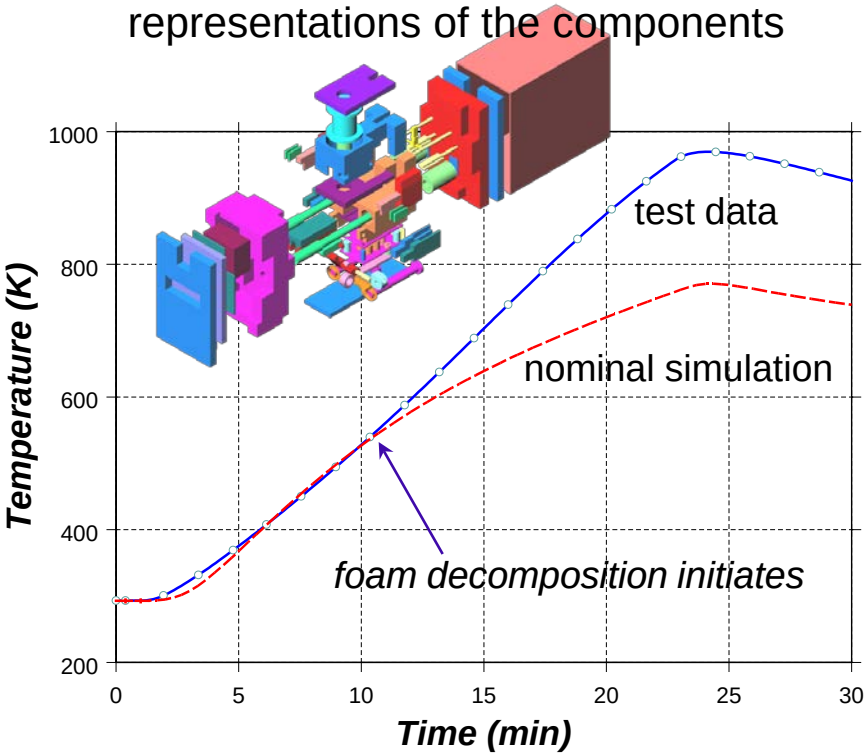
lumped models with measured or estimated effective properties



(effective component heat capacity measured at 333 K)

components are too complex to capture with lumped models

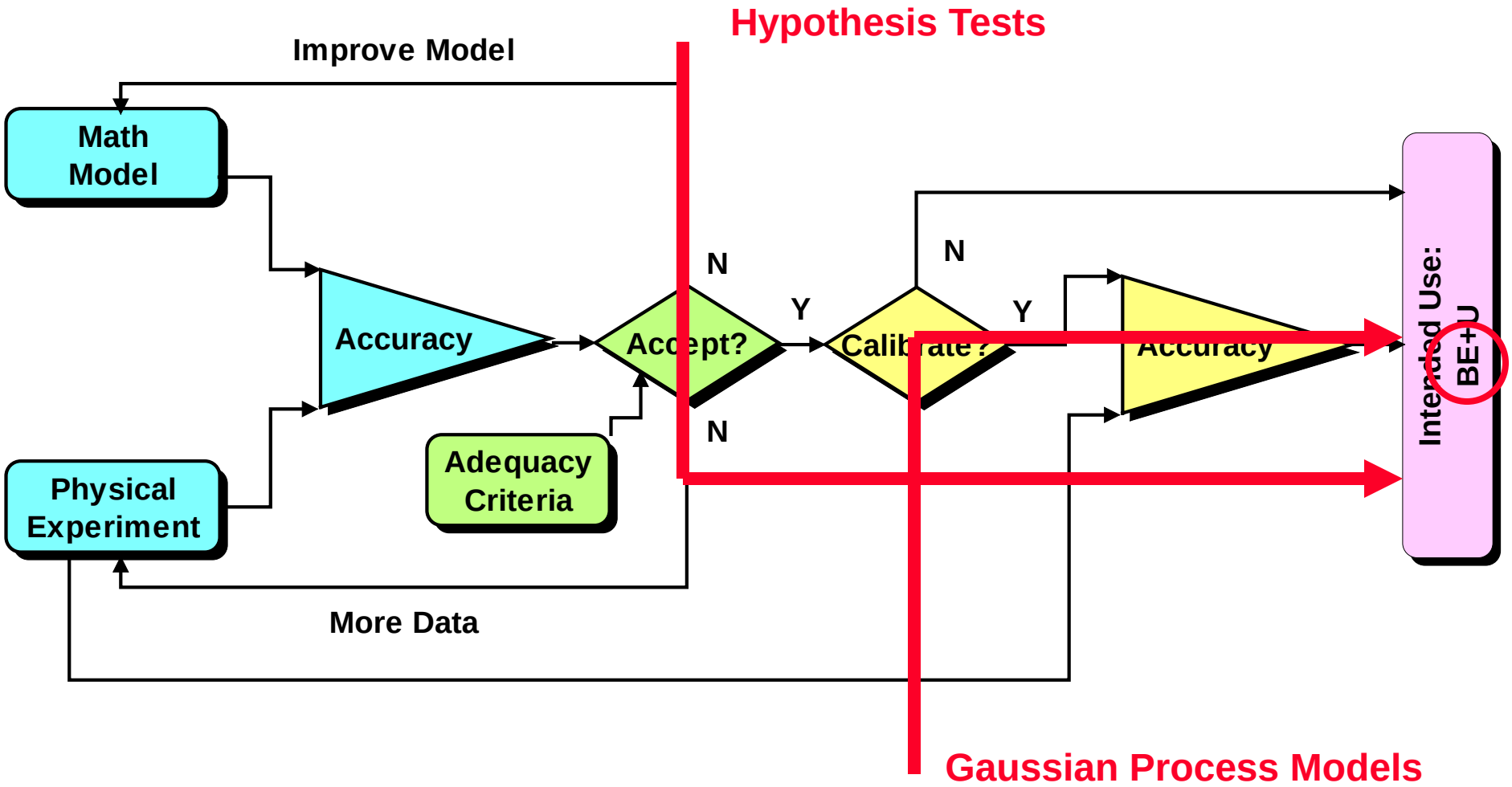
high-fidelity models with detailed representations of the components



Improved geometric fidelity only part of the problem—high fidelity physics models, such as a foam decomposition heat transfer model, are also needed for predictive capability

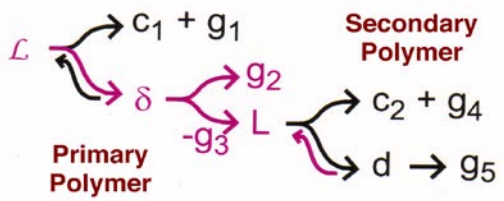
Validation is Assessment Supporting BE+U

Calibration is not Validation



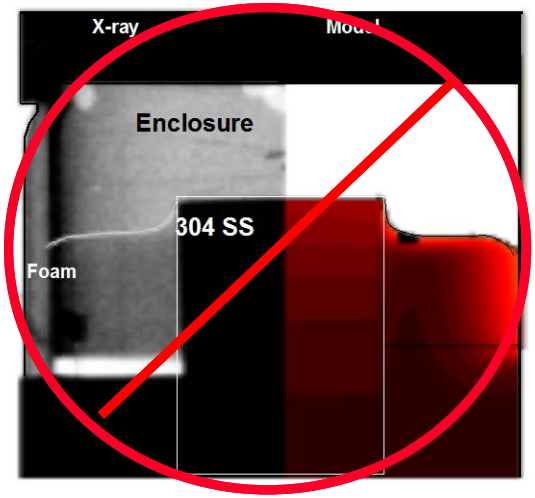
Validation is Statistical

Vugraph Norms Are Not Adequate

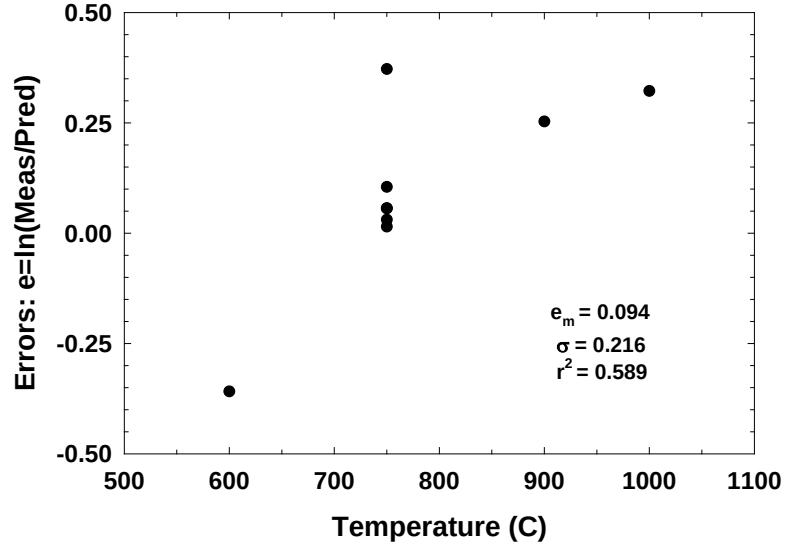


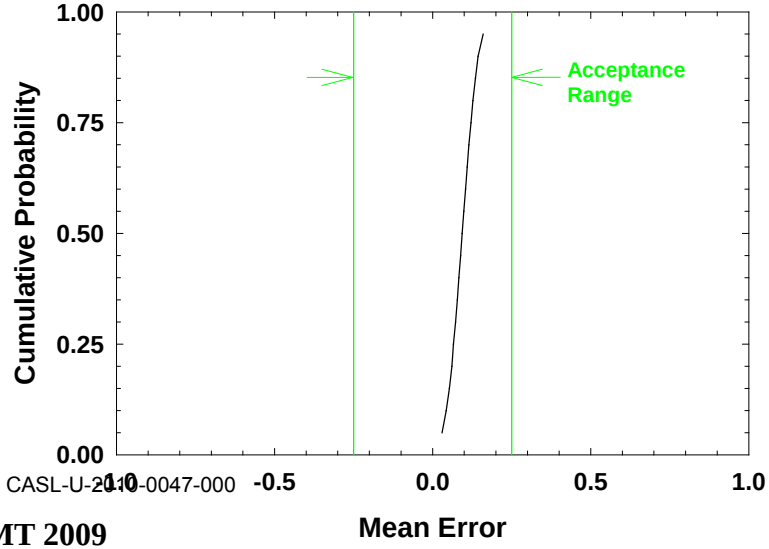
Physics-informed model

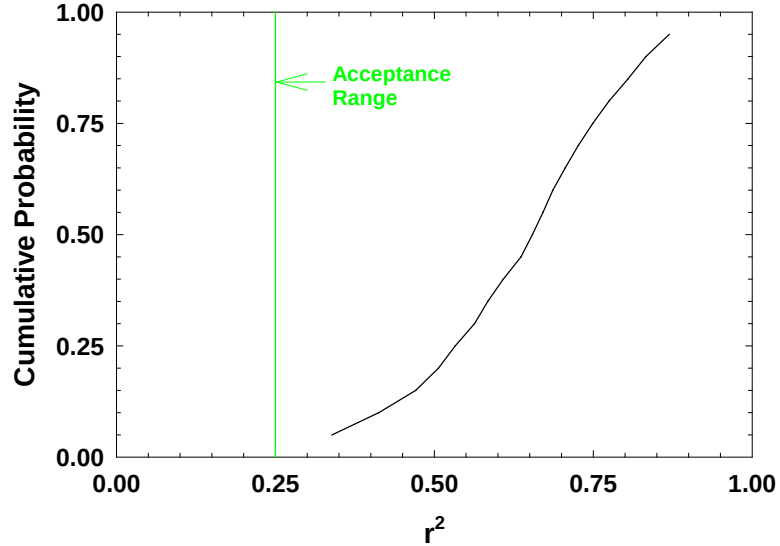
Vugraph Norm



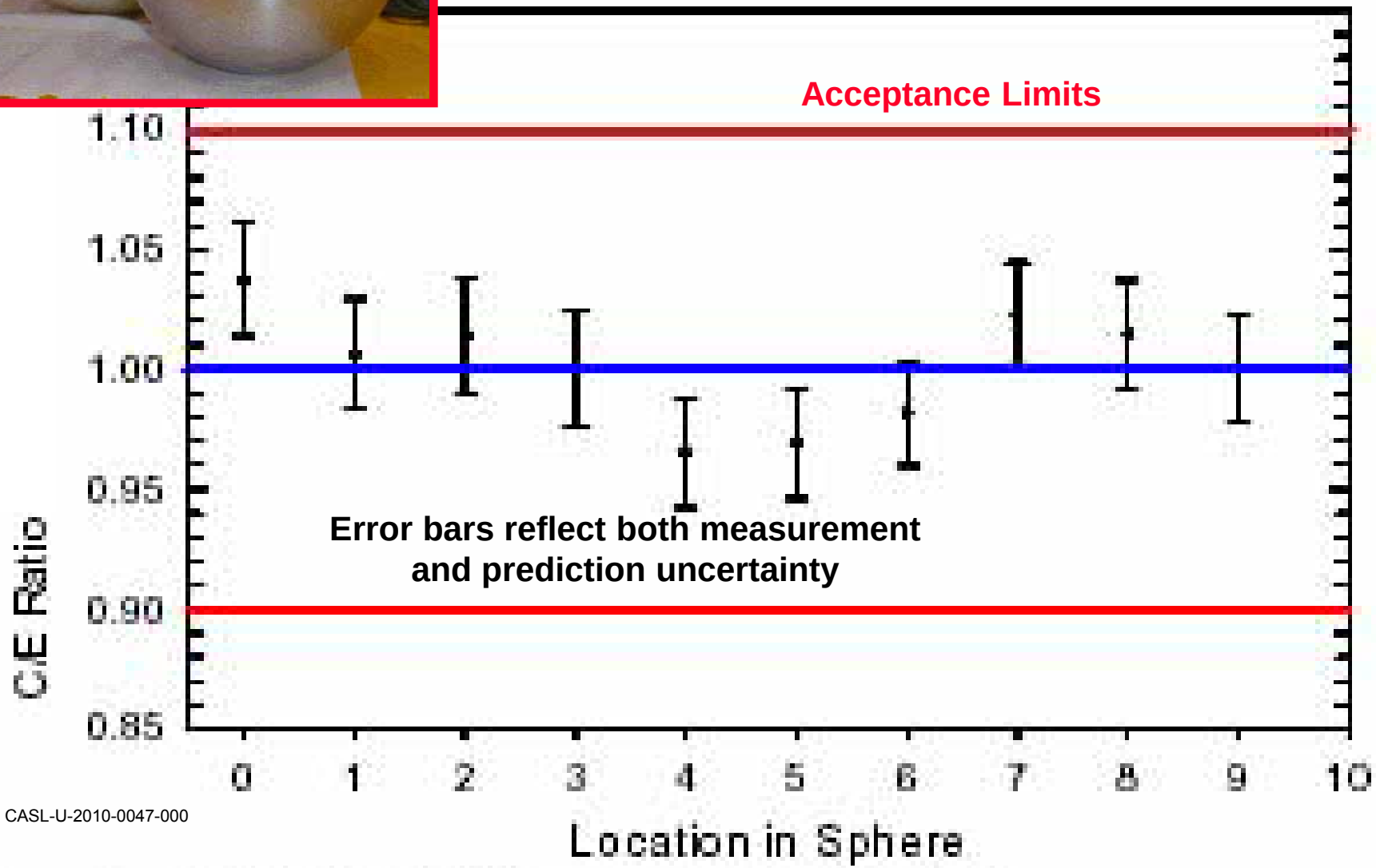
T-dot BC is more appropriate than as-tested T-fixed







Neutron Attenuation in Test Objects



CASL-U-2010-0047-000

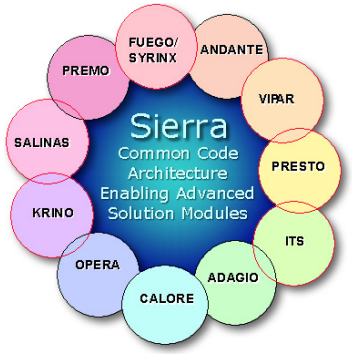
Attributes of Code and Solution Verification

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

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

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

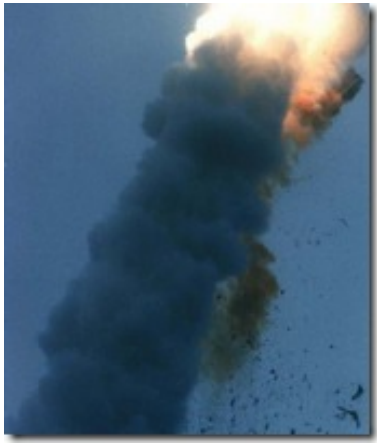
What if this term is not negligible?

- **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?

Software Errors Are Costly: Both in Dollars and Reputation

Hartford Coliseum ($70M)

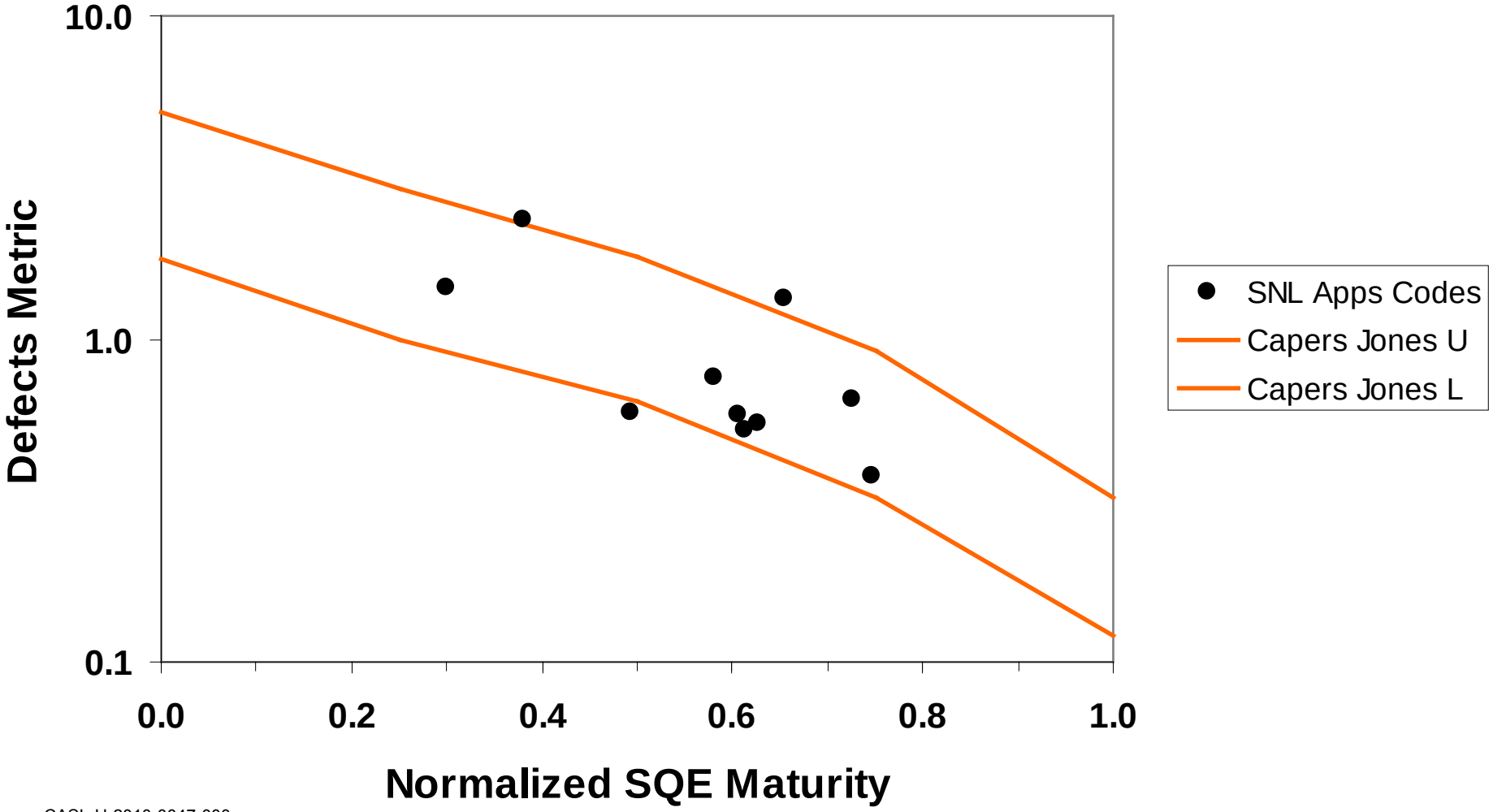


Ariane 5 ($500M)

Mars Climate Orbiter ($125M)

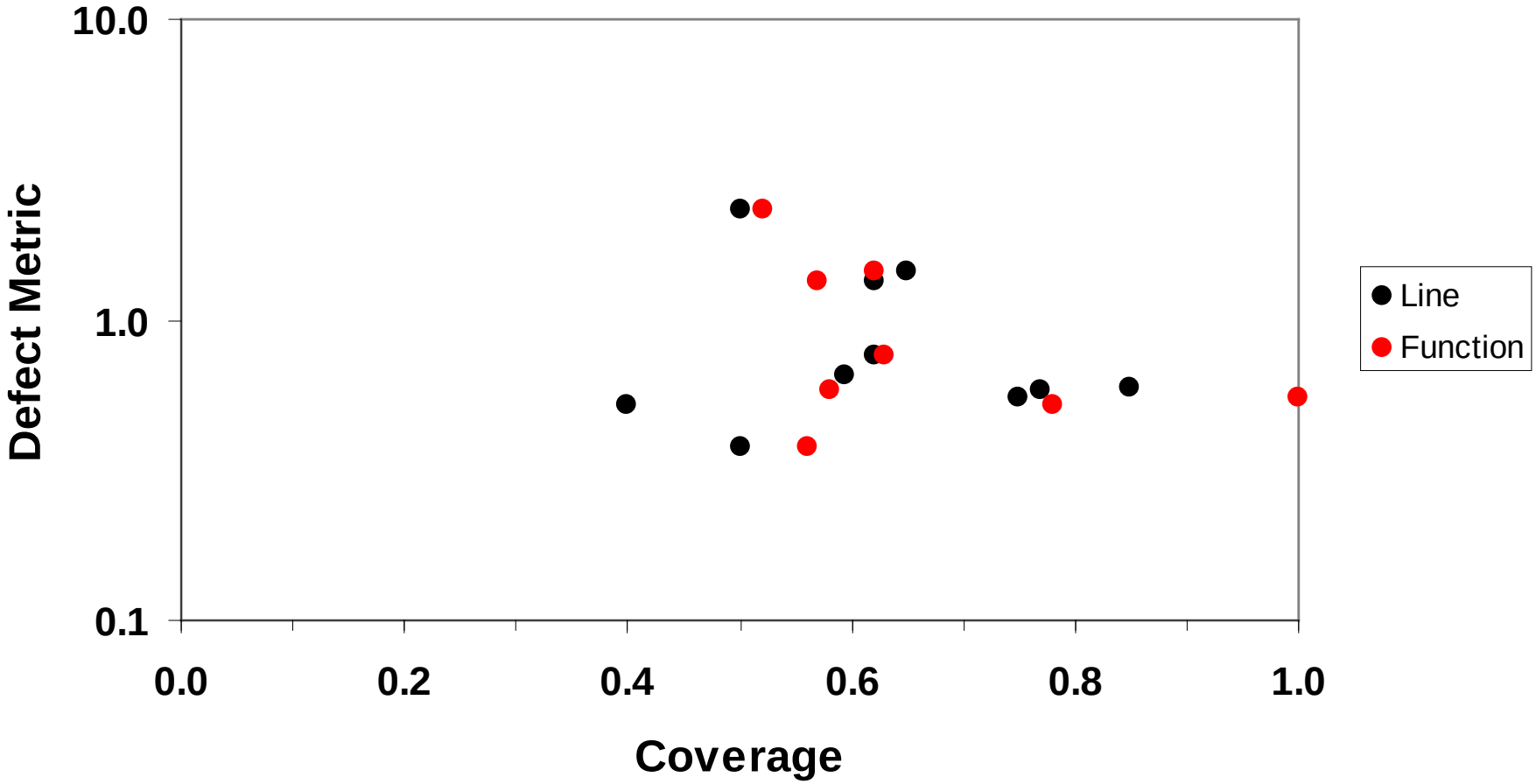
http://www.devtopics.com/20-famous-software-disasters/

Good SQE Practices Reduces Defects



CASL-U-2010-0047-000

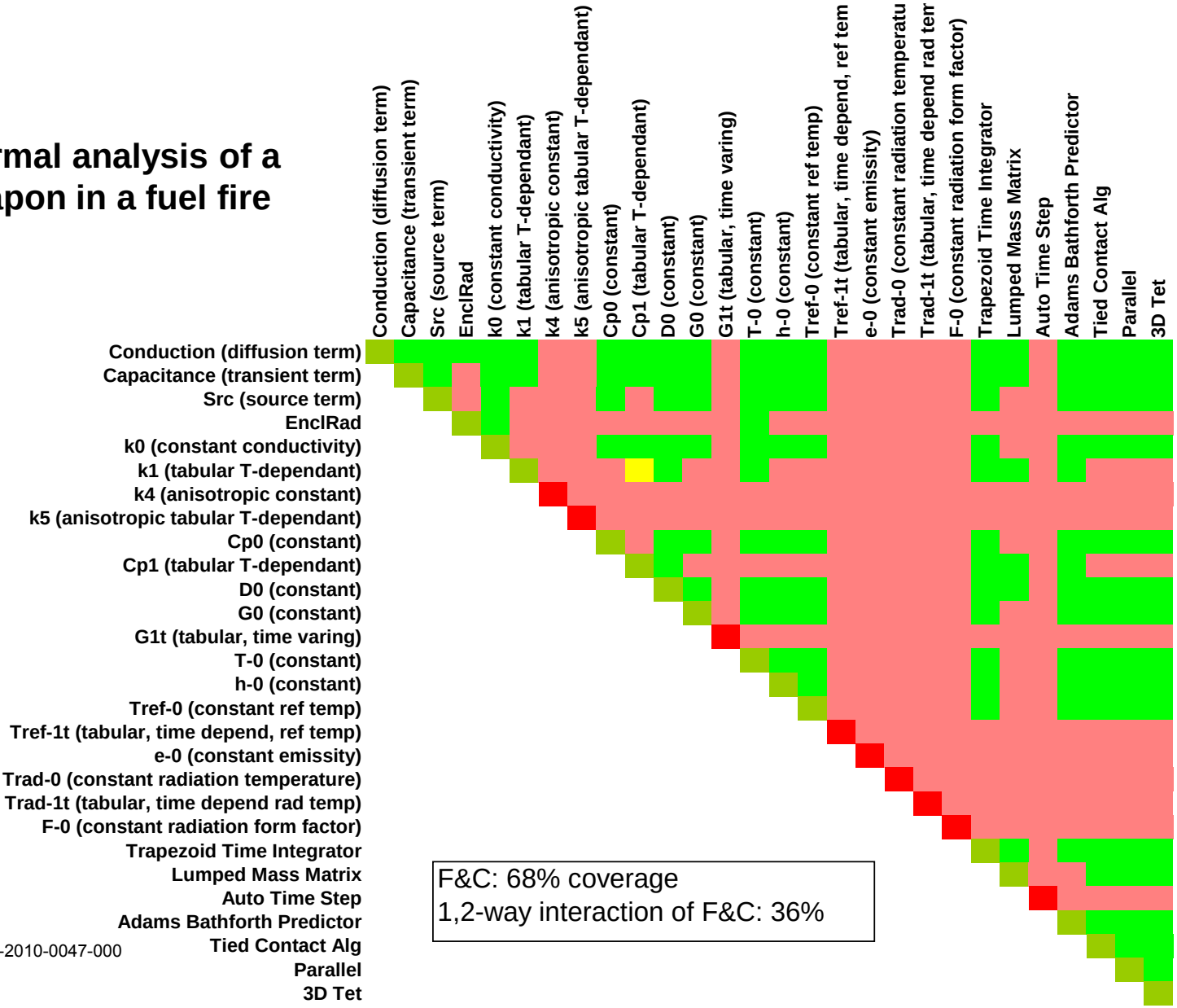
Why Doesn't Code Testing Have a More Decisive Impact on Defects?



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

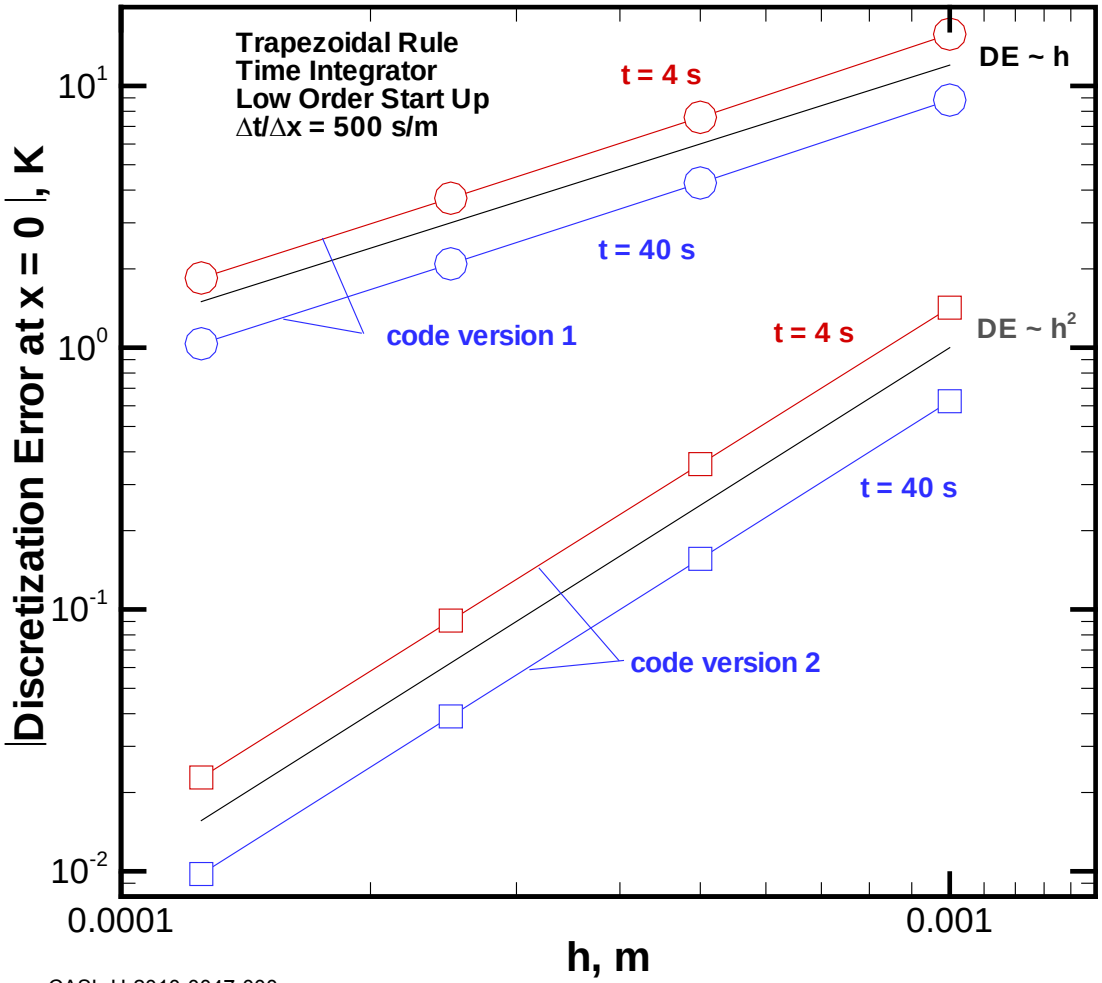
We Are Shifting Our Focus to Verification of Features and Capabilities and Their Interactions

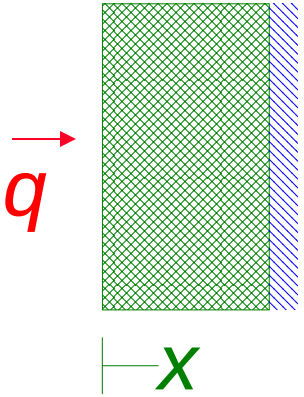
Thermal analysis of a weapon in a fuel fire



F&C: 68% coverage
1,2-way interaction of F&C: 36%

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



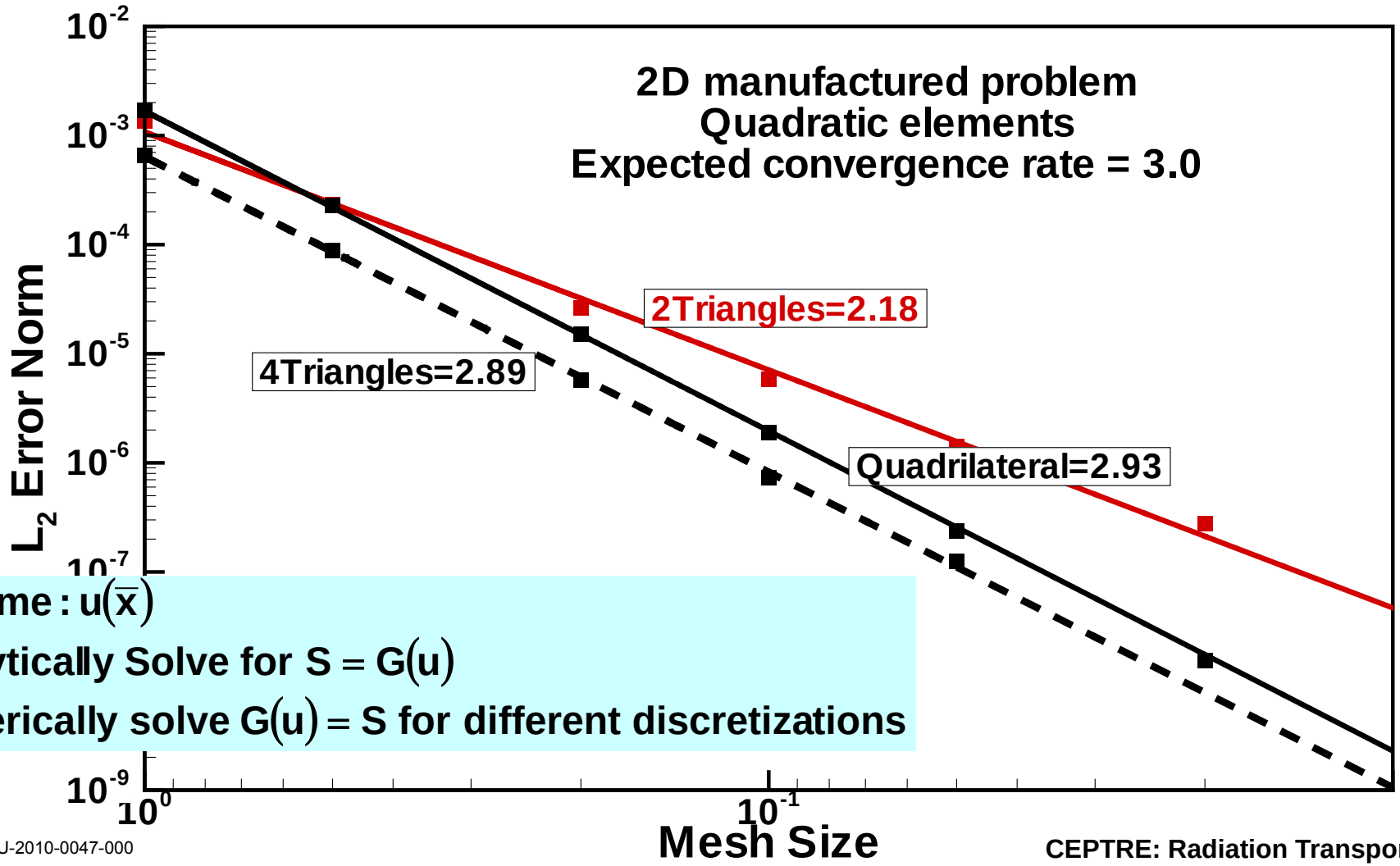


- Transient response of planar 1-D slab to constant flux with analytic solution as the benchmark
- Code bug discovered and fixed based on priority and resource availability. Status tracked in code issue log, which can be accessed by analysts

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

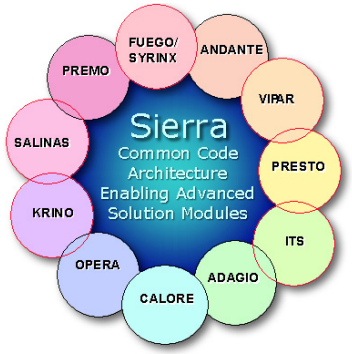
Attributes of Code and Solution Verification

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.

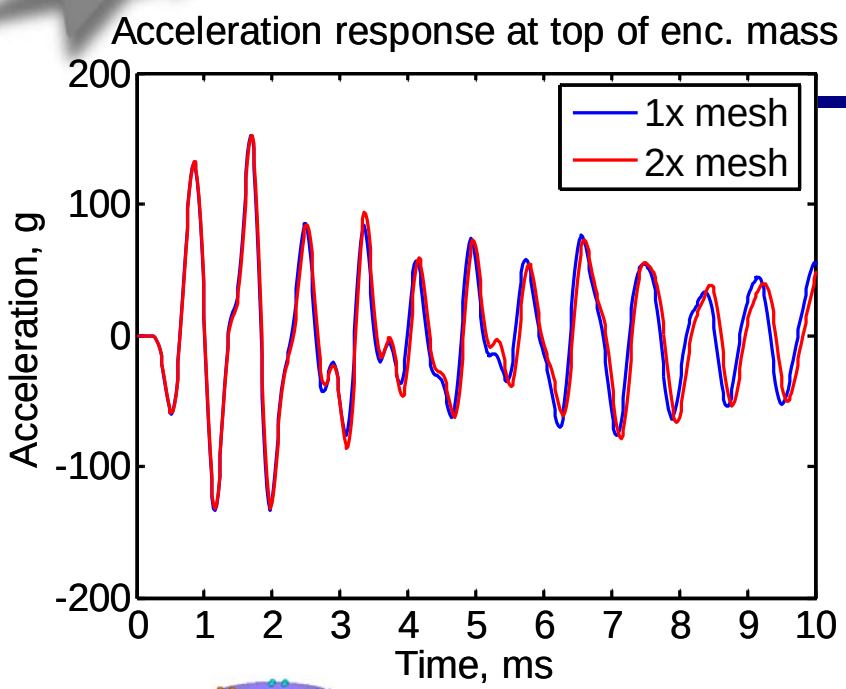
Inference

Regression Testing

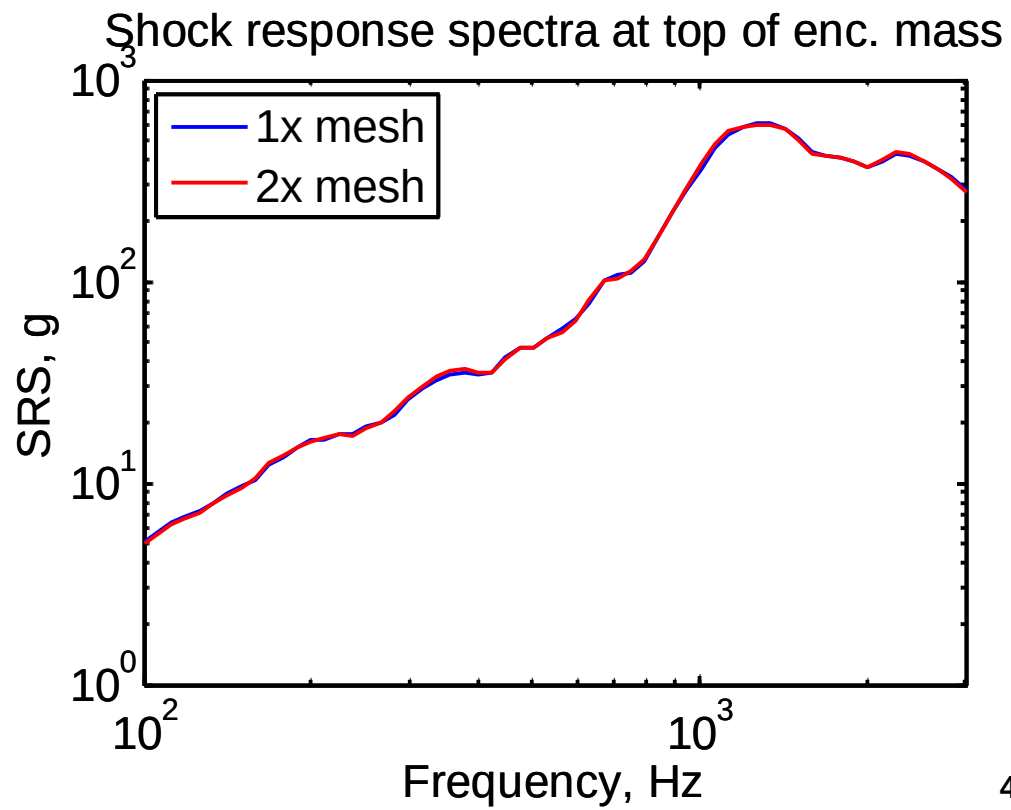
Code Verification: Convergence to correct answer, *wrong application*

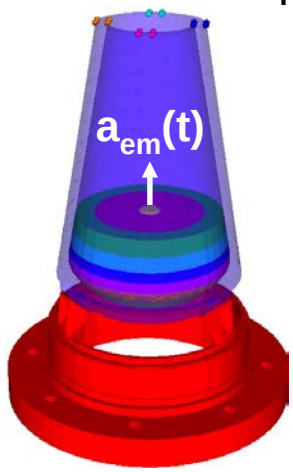
- Eliminate code bugs *AND* inadequate algorithms

It's Common to Explore Sensitivity to Mesh Parameters



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

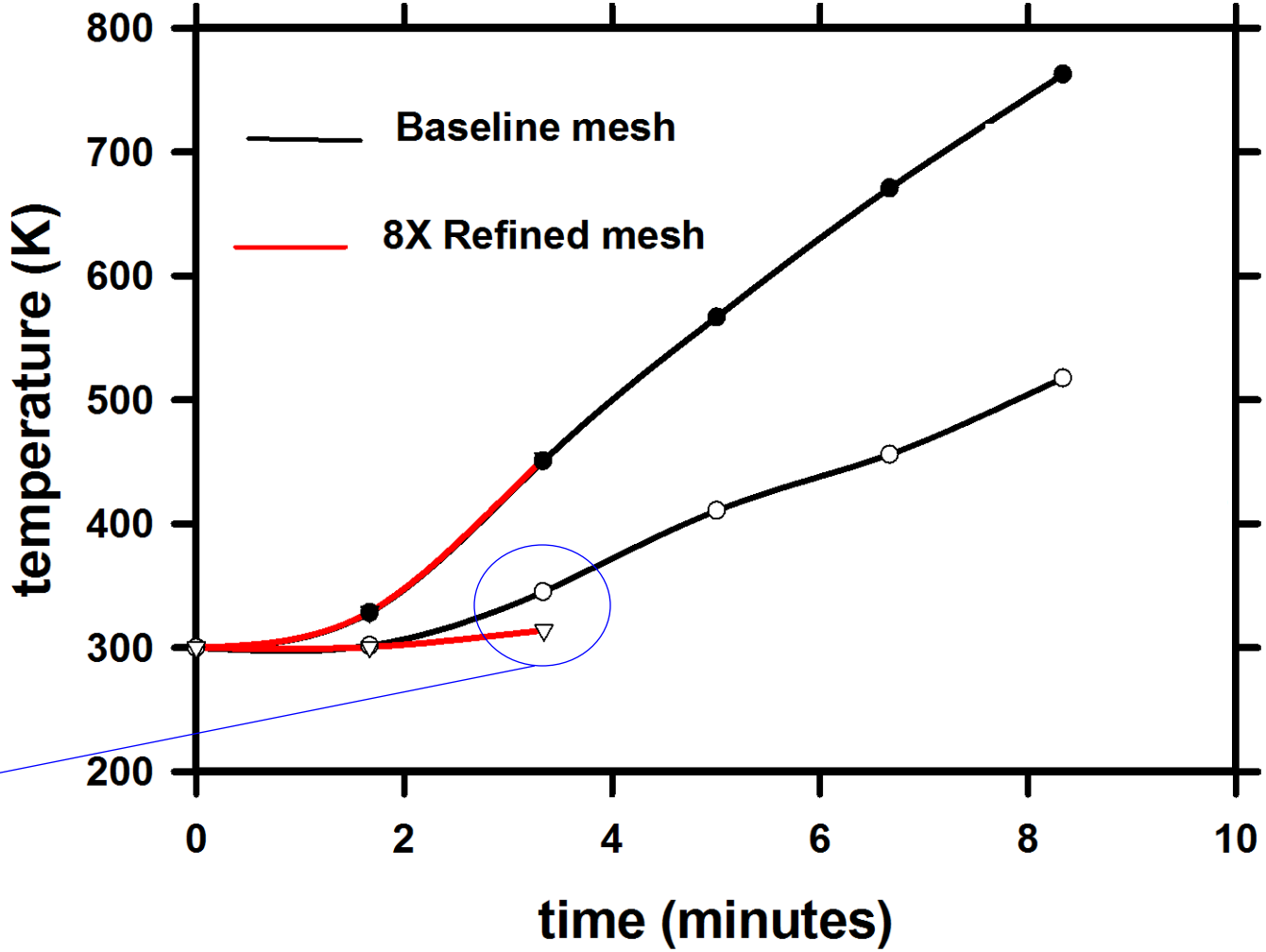




Structural Dynamics

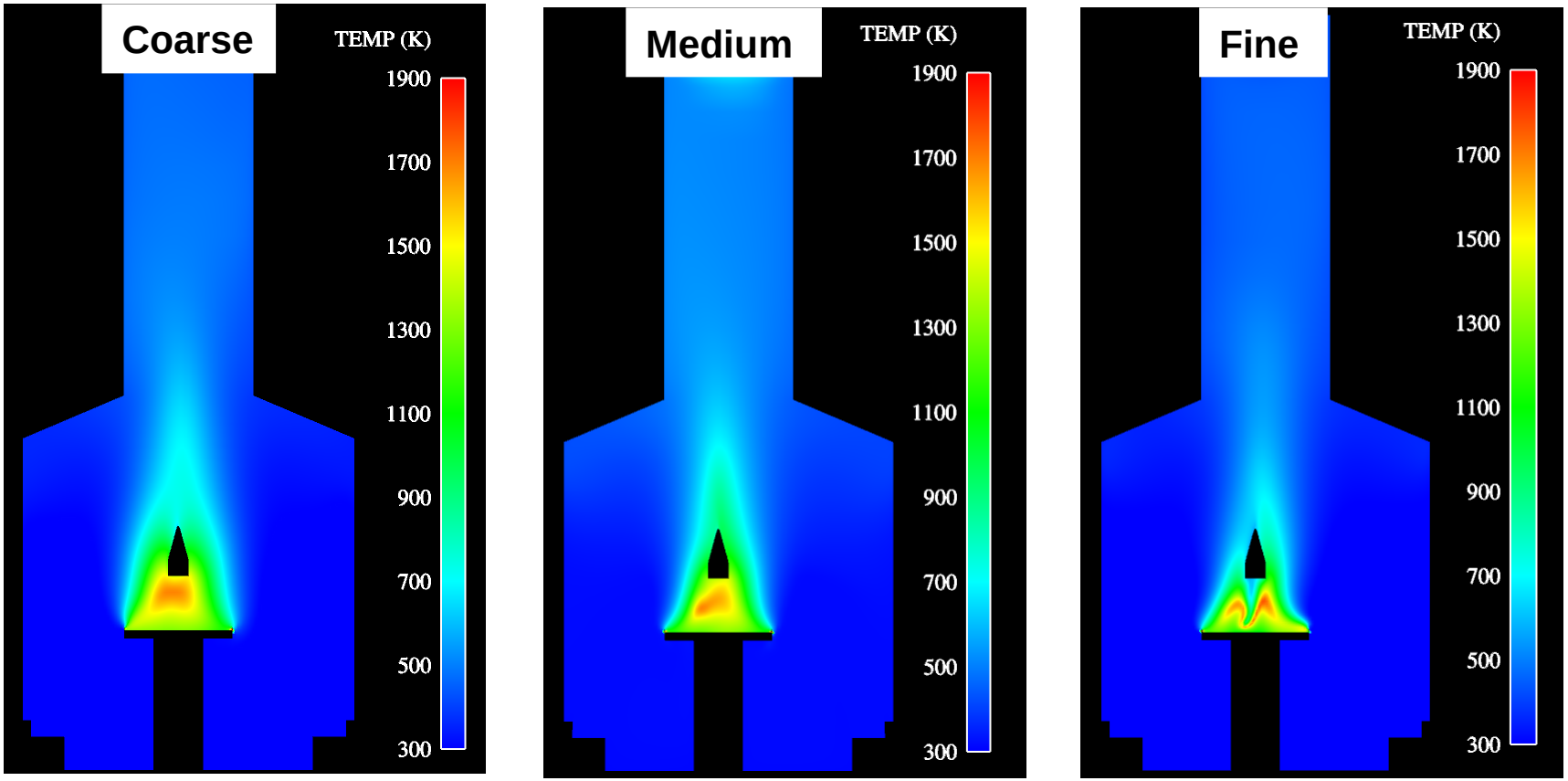
CASL-U-2010-0047-000

Solution Verification on High Fidelity Models is Hard



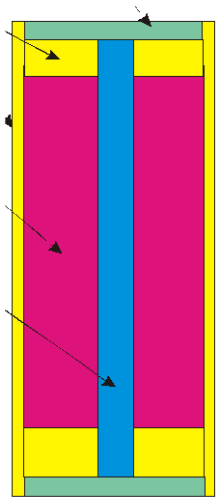
Critical heat transfer path under-resolved because of large discontinuity in material properties

Discretization Study Revealed Bifurcation of Solution Space

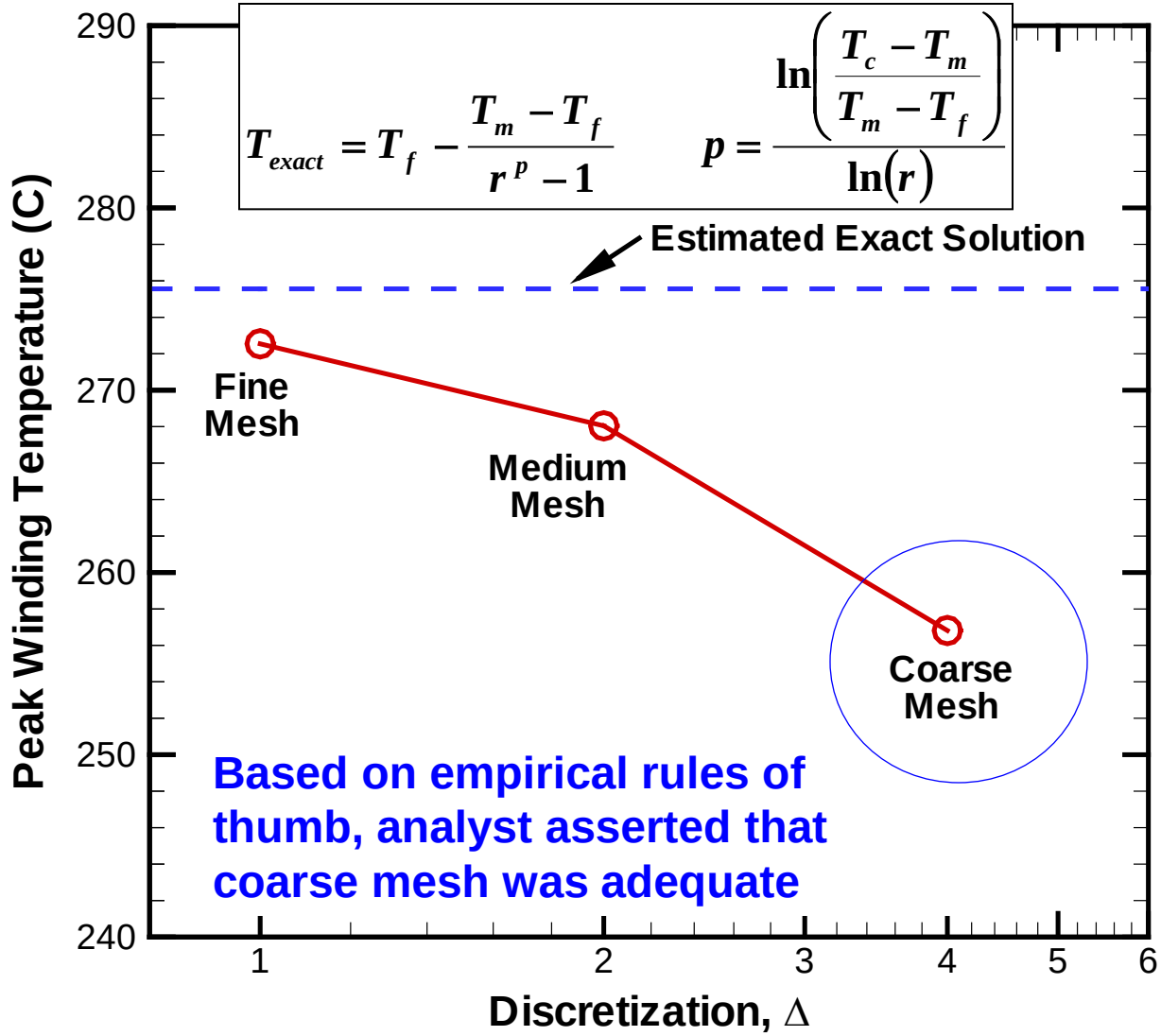


Calorimeter Fire: BVG Solutions

Numerical Errors Pollute Validation Assessments



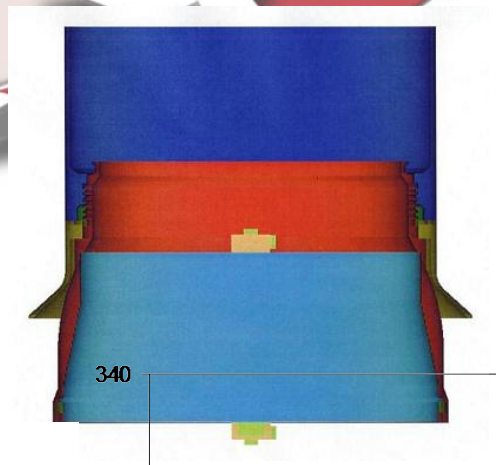
$$T_{exact} = T_f - \frac{T_m - T_f}{r^p - 1} \qquad p = \frac{\ln\left(\frac{T_c - T_m}{T_m - T_f}\right)}{\ln(r)}$$



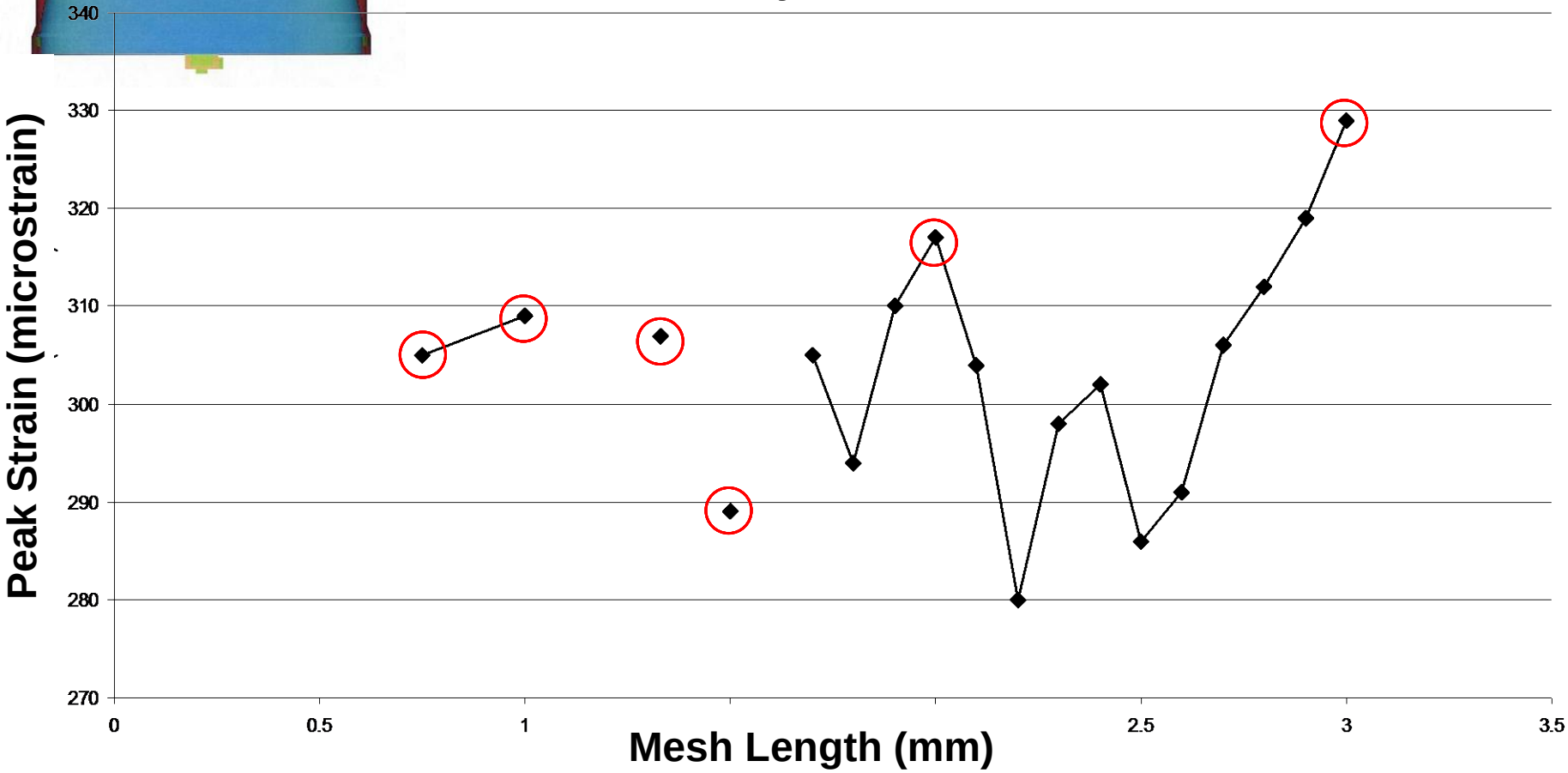
Based on empirical rules of thumb, analyst asserted that coarse mesh was adequate

Solutions Don't Always Converge

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



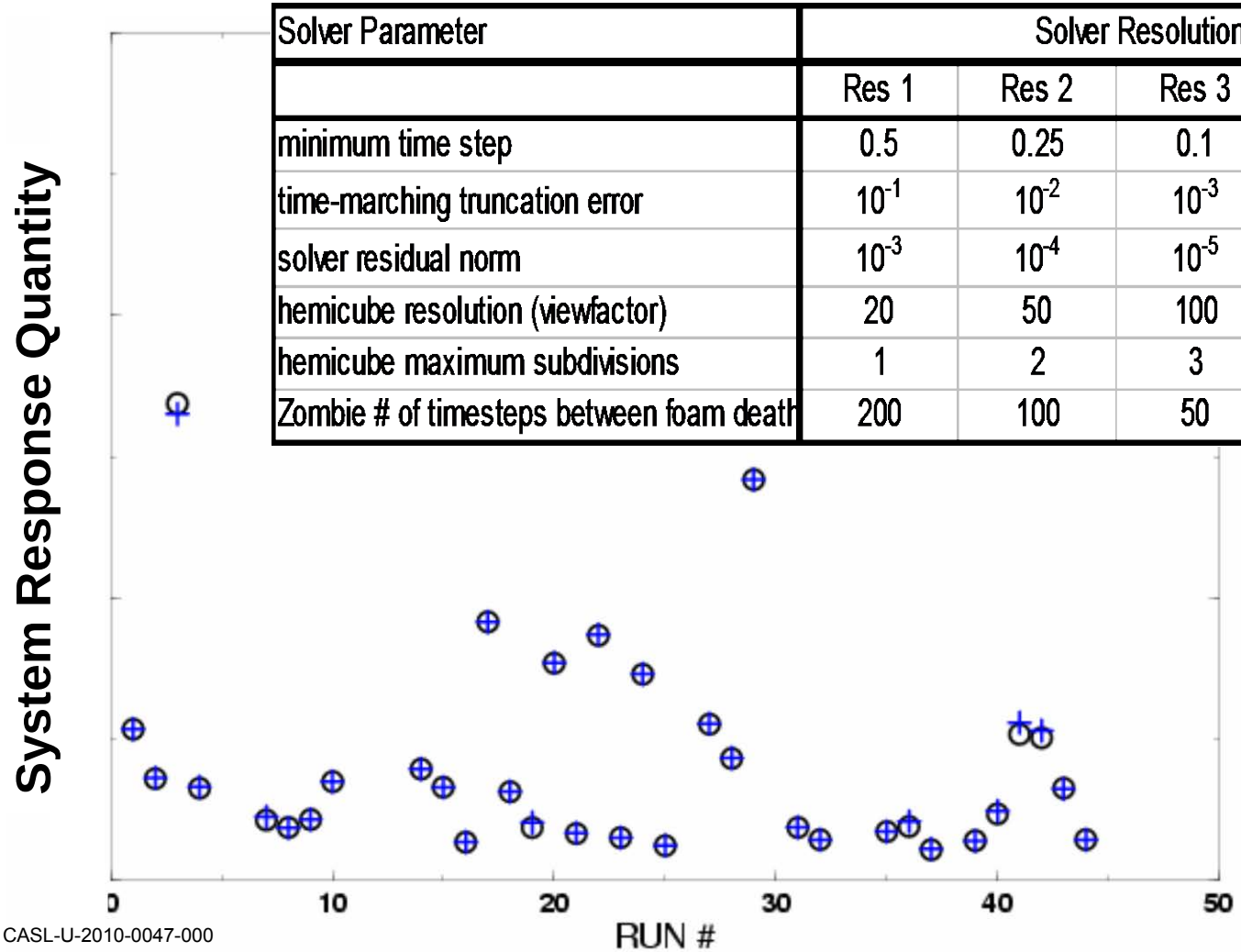
Threaded assembly



CASL-U-2010-0047-000

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



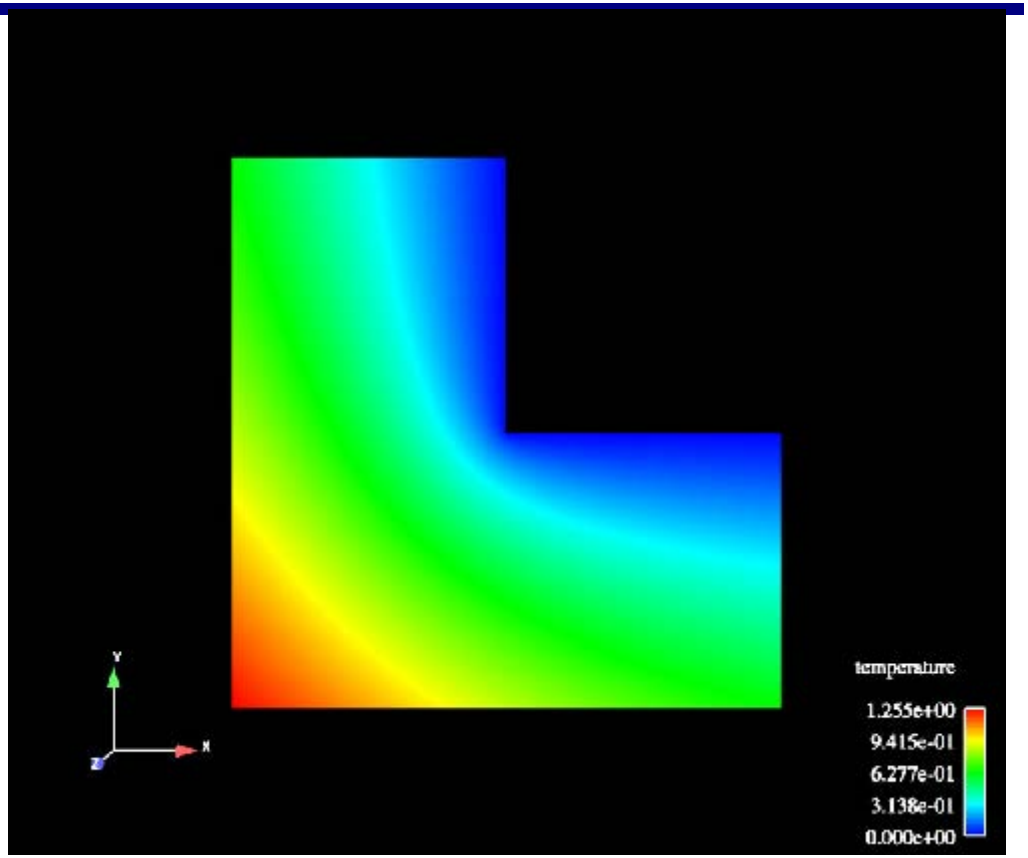
Verification of Error Estimator and Adaptive Algorithm

- 2D Exact Solution:

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

- Linear elements
- ZZ error estimator

- Feedback adaptive algorithm:

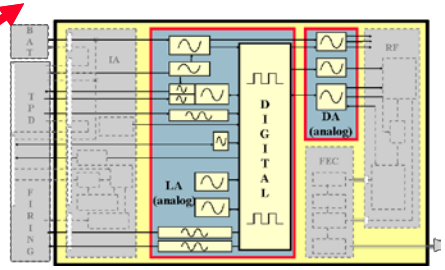


$$\text{if } \left(||e^*||_{H_1(\omega_j)} > 0.995 \max_{1 \le i \le N_\omega} ||e^*||_{H_1(\omega_i)} \right) \text{ then refine } \omega_j$$

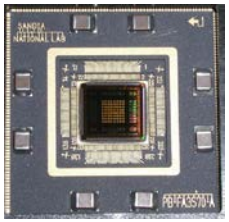
A Hierarchy of Science-Based Validation Experiments Ensures Models Get the Right Answer for the Right Reasons

Hierarchal Validation: Right answer for the right reason

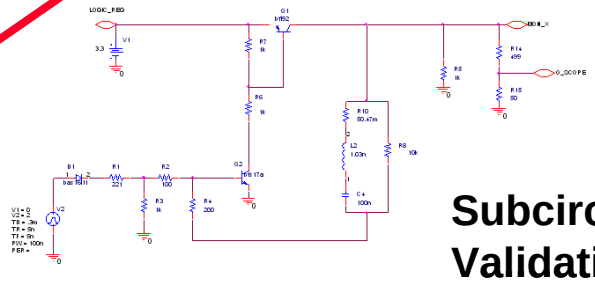
Increasing complexity, Decreasing number of tests



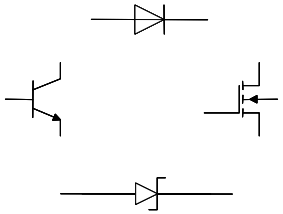
System-Level Circuit Validation



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**

CASL-U-2010-0047-000

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

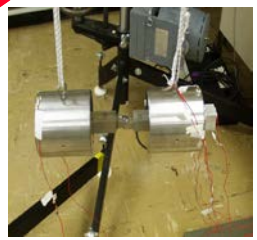
Hierarchal Validation: Right answer for the right reason

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**

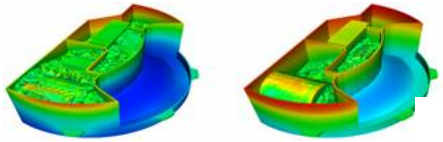
CASL-U-2010-0047-000

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

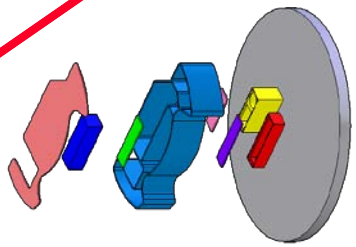
Hierarchal Validation: Right answer for the right reason

Increasing complexity, Decreasing number of tests

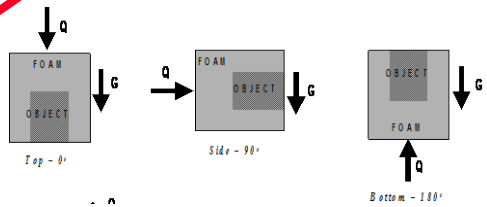
Full System Test



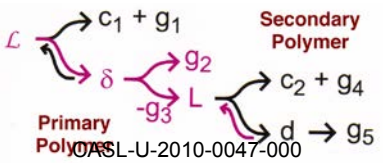
Validation Real Sub-systems



Validation with mockups



Foam recession

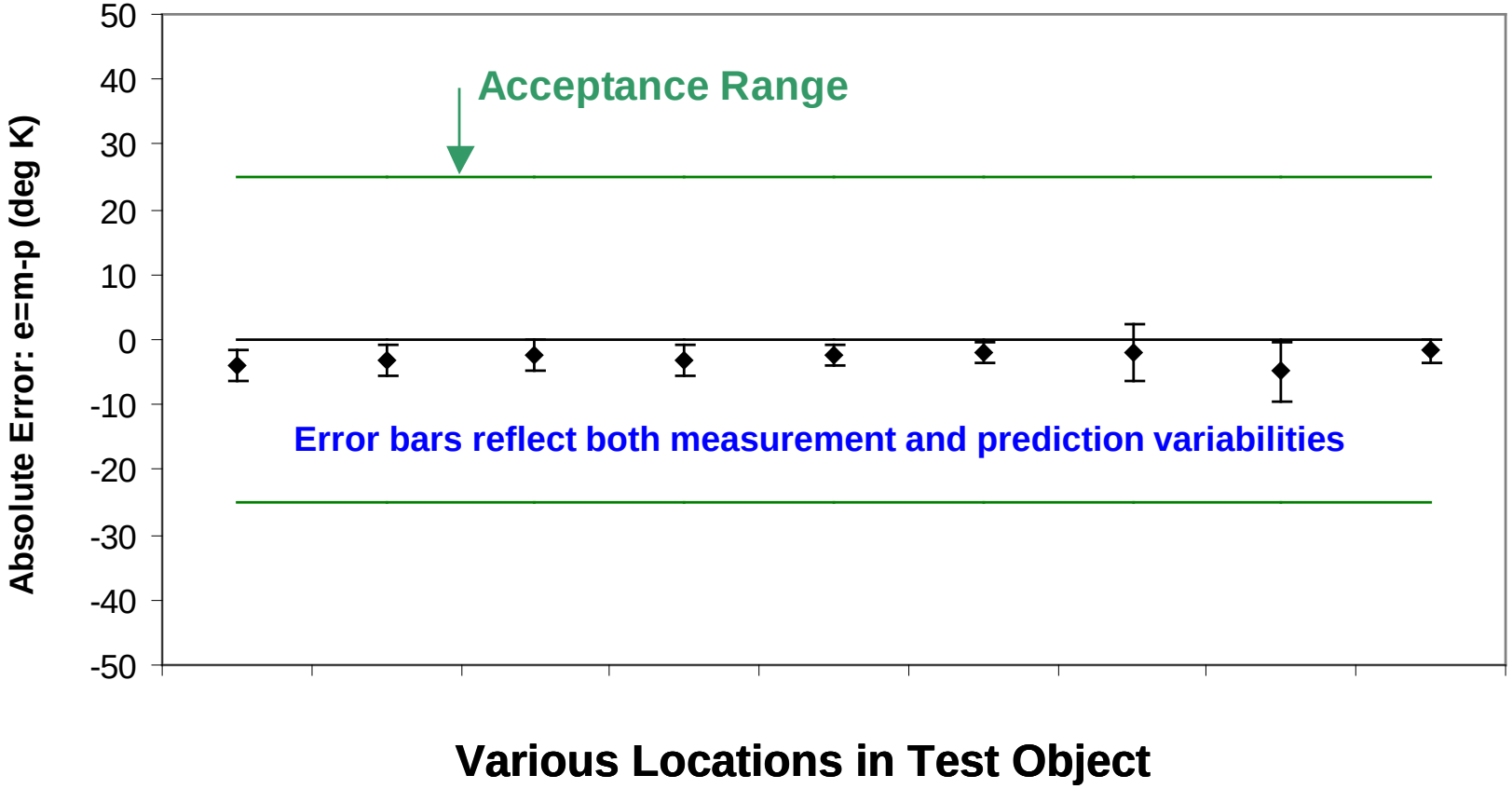


Chemistry characterization/validation

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

Well Established Physics Fidelity

e~2K for conduction



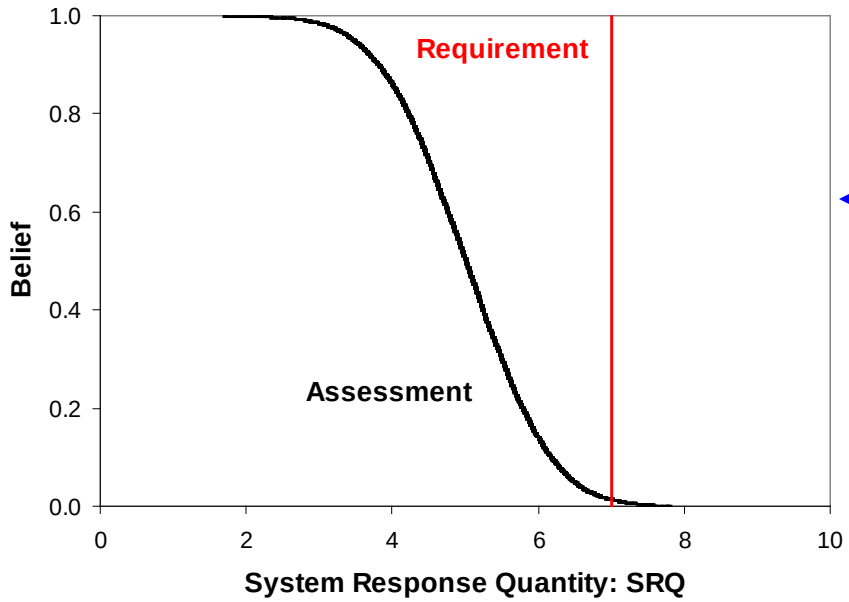
CASL-U-2010-0047-000

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

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



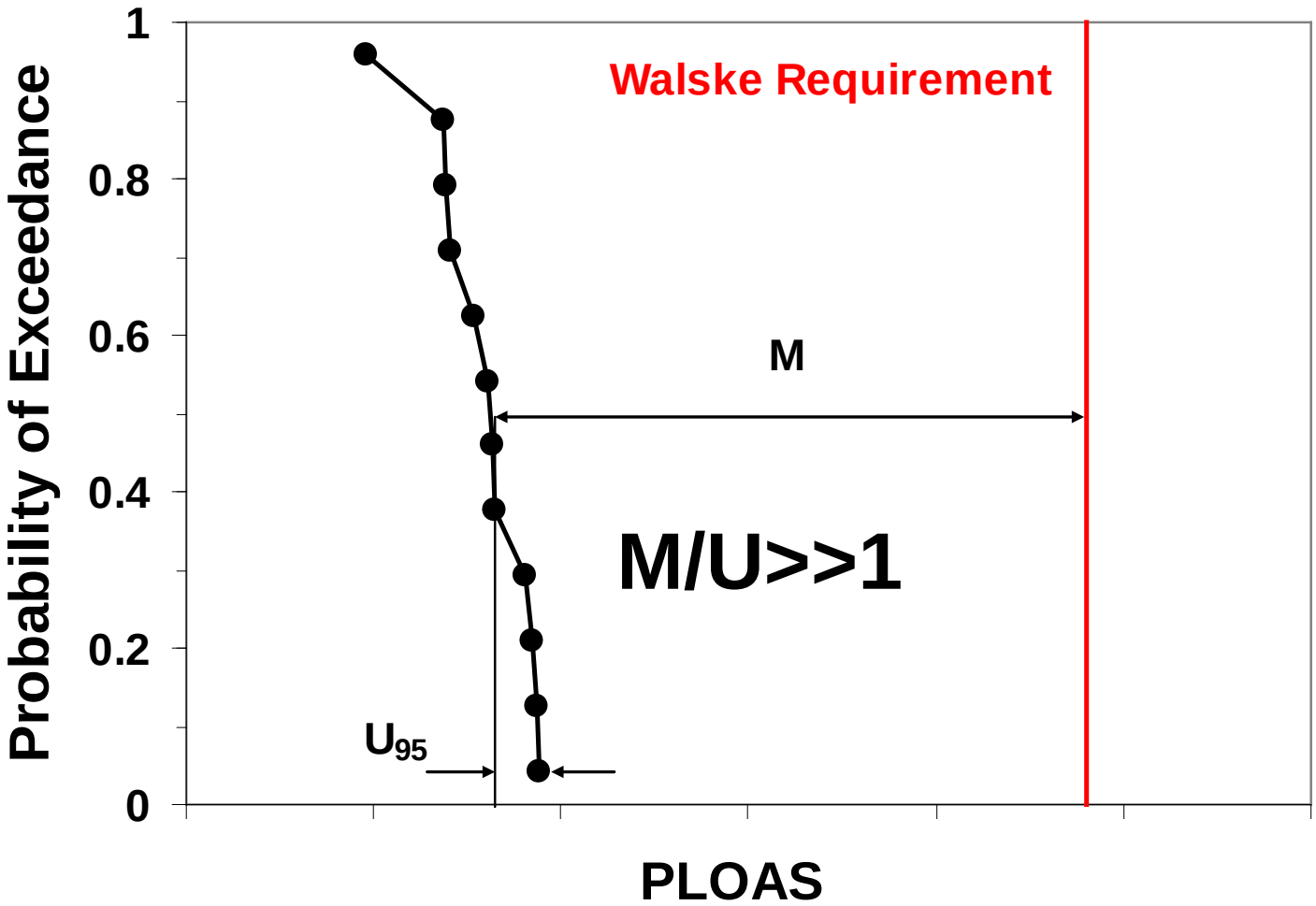
?

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

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

Uncertainty propagation should not make something from nothing

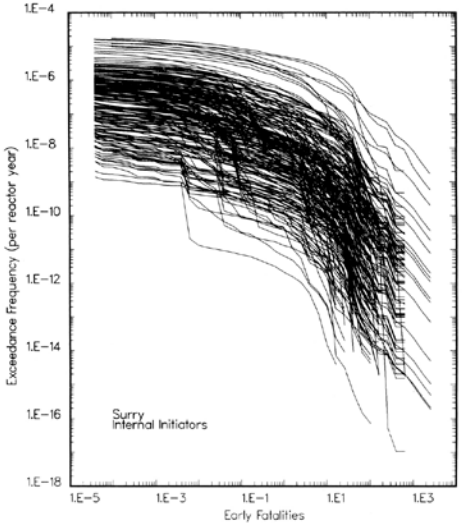
Quantifying Margins and Uncertainties (QMU aka QRA) Supports Risk-Informed Decisions



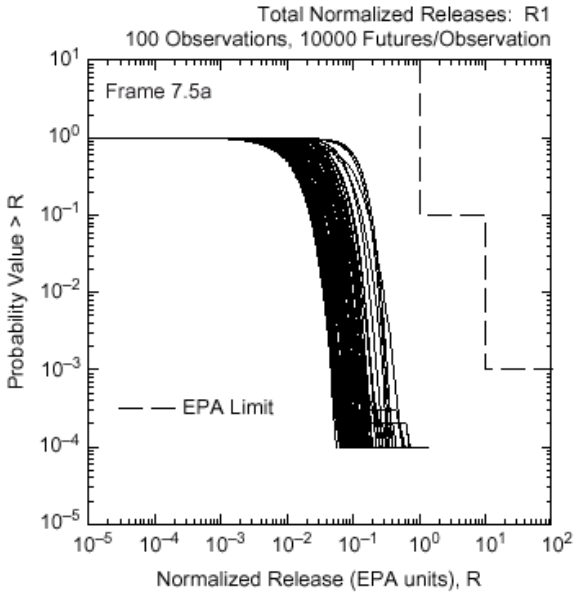
QRA = Quantitative Risk Assessment

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

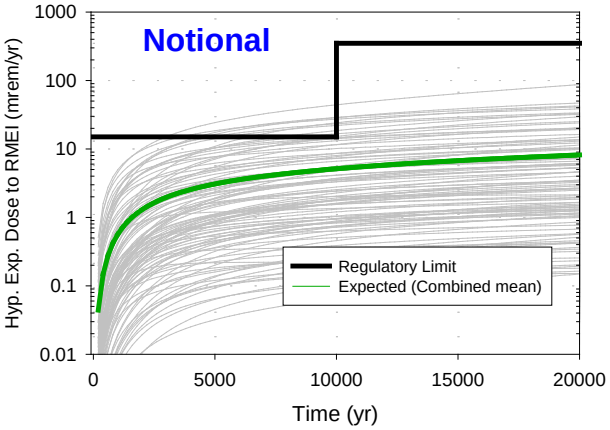
Reactor Safety NUREG-1150: 1990



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**

CASL-U-2010-0047-000