SAND2009-0591P

# Verification, Validation, Uncertainty Quantification, Predictive Modeling and Simulation:

## *Integration of NW Capabilities into NEAMS*

**Jim Stewart, SNL**

**Cetin Unal, LANL**

**François Hemez, LANL**

# Three Main Themes

1. The Born-Assessed Framework can provide a science-based approach to risk-informed decisions for supporting licensing.

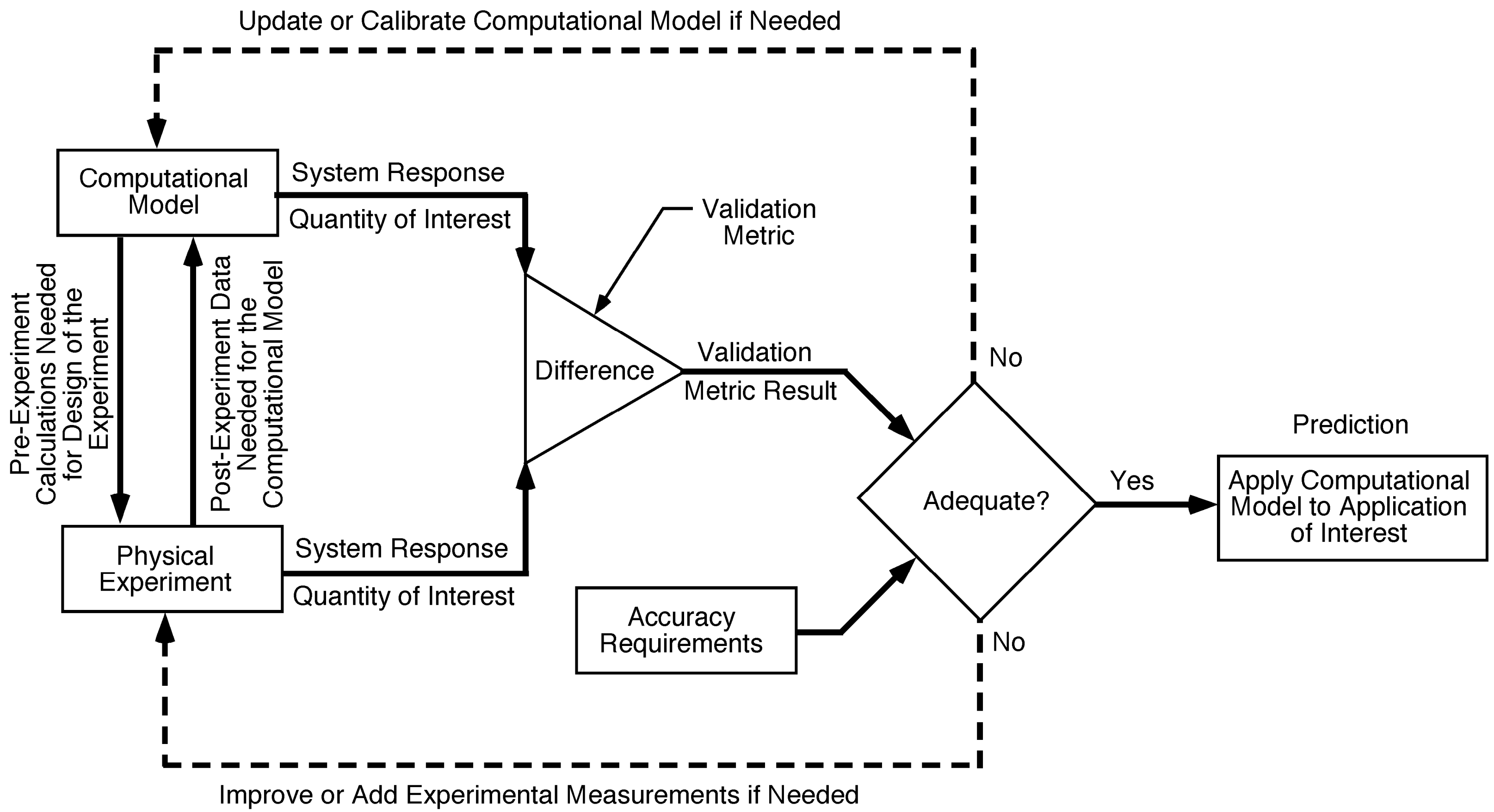
2. The NW program has provided leading tools and capabilities that can serve as starting points for the NEAMS VU program. However, *significant new VU tools and capabilities are broadly needed* to achieve the NEAMS goals.

3. Integration of VU with the other parts of NEAMS is essential from Day 1.

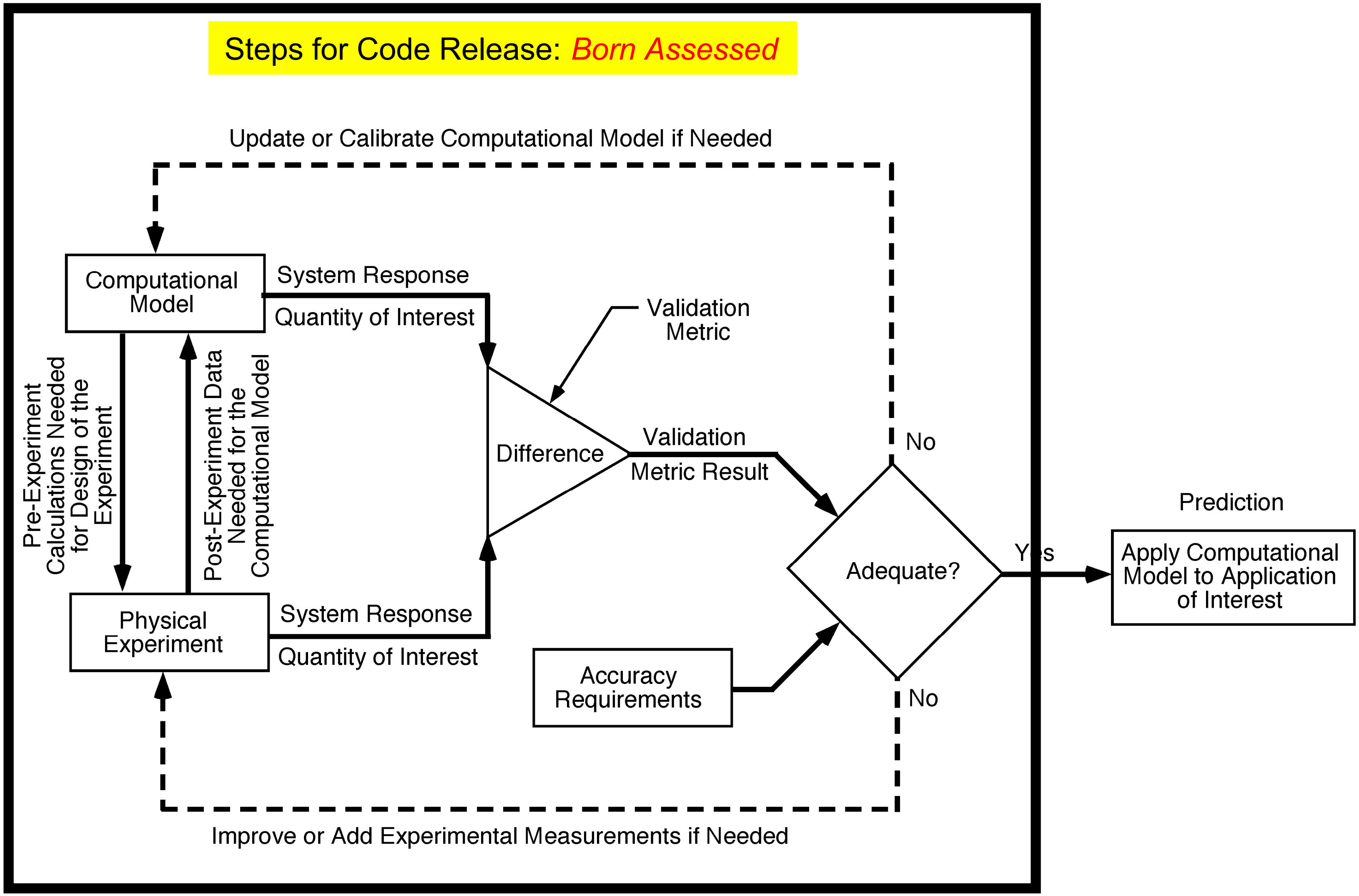
# Outline

- *V&V-based Born-Assessed Framework*
- Verification
- VU Tools and Algorithms
  - Capability Gaps
  - Integration
- Towards Licensing

# *V&V-Based Born-Assessed Framework:* Validation, Calibration, and Prediction




*Reference:* W. L. Oberkampf and M. F. Barone, "Measures of Agreement Between Computation and Experiment: Validation Metrics," SAND2005-4302, August 2005.




*Remark:* Conceptually similar to "Born Assessed" process documented in: R. A. Nelson, A. R. Larzelere, and S. Runnels, "GNEP Modeling and Simulation: An Improved Applications Development Paradigm for Rapid Deployment," Los Alamos National Laboratory Report LA-UR-07-1865.

# Further Details of Validation/Calibration Process

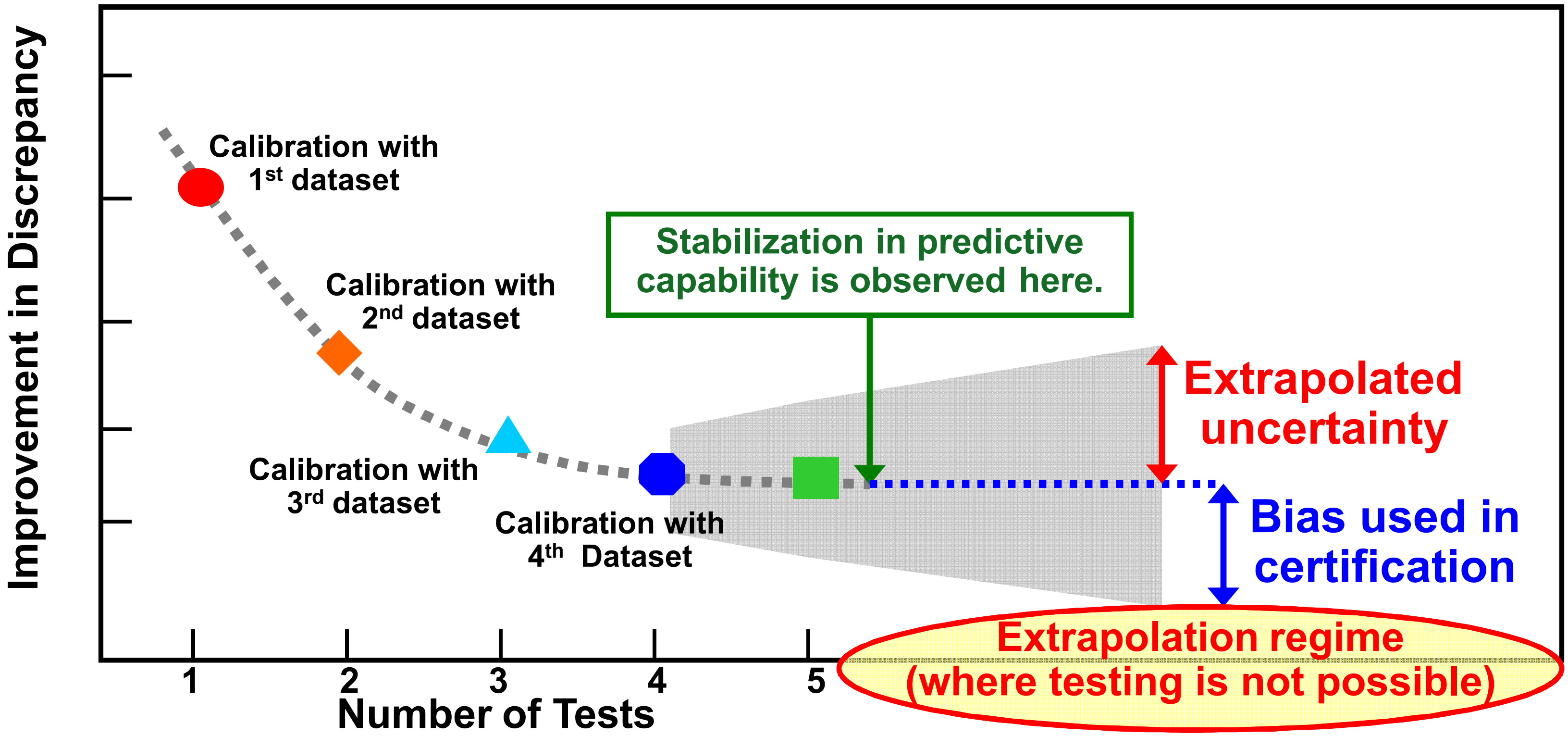
*(Prerequisite for Code Release)*

- Physical Experiments May Include
  - Separate Effects Tests (SETs)
  - Integral Effects Tests (IETs)
  - Uncertainty Quantification
- Application of Computational Model Involves
  - Model Setup: Geometry and Mesh
  - Solution Verification
  - Sensitivity Analysis and Uncertainty Quantification
- *Validation Metrics*: Comparison of computational model results to experimentally measured system response quantities (SRQs)
- *Updates/improvements* to model and/or physical experiments
- *Assessment of adequacy* for intended application (including possible interpolation/extrapolation of model)

Primary Goal: *Born Assessed*

# More on Assessment of Adequacy: Seeking Robustness and Maturity

- We propose to develop models of predictive maturity based on the concept of *stabilization* (or robustness).

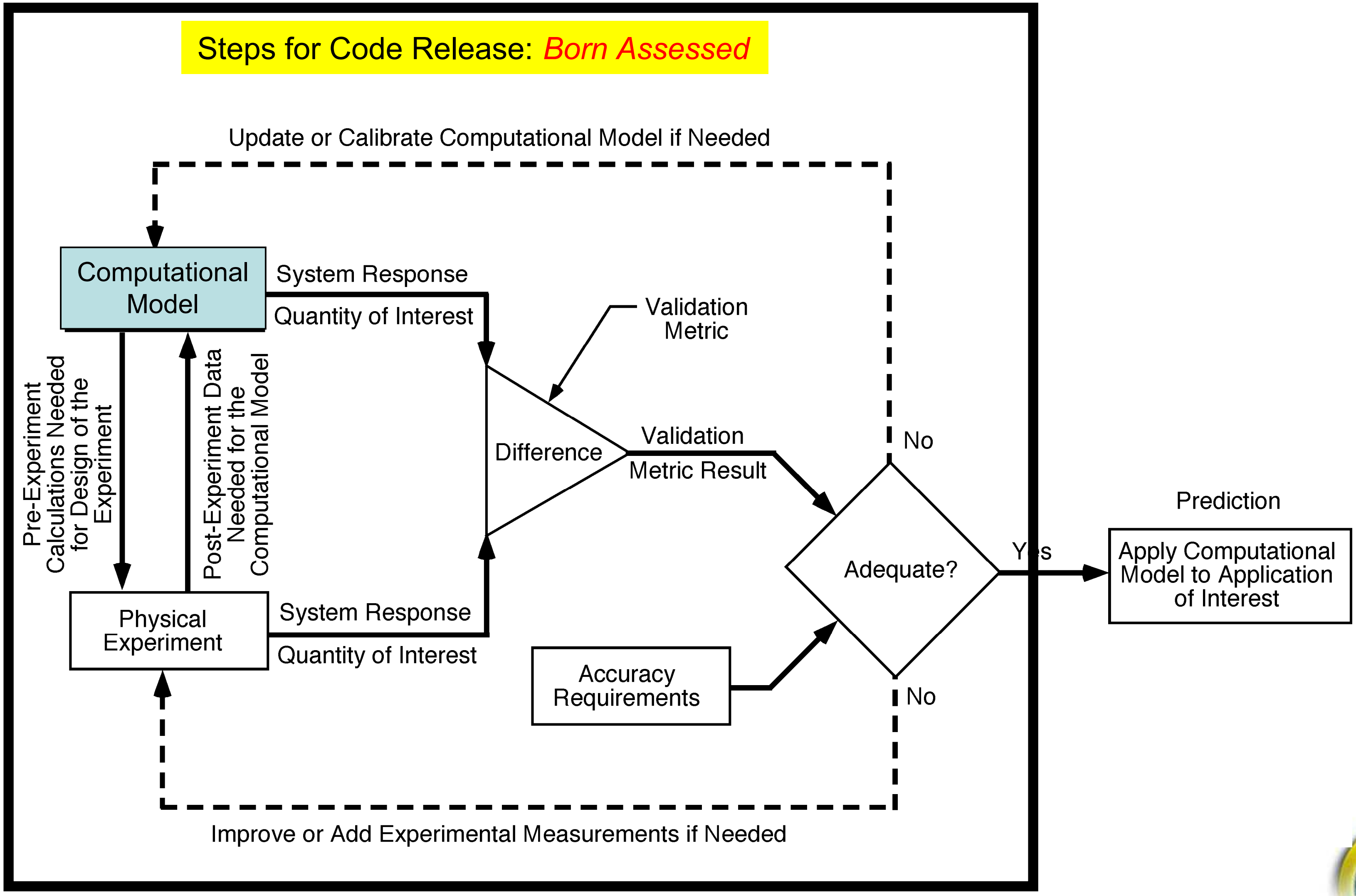



Reference: Hemez, F.M., Atamturktur, S., Unal, C., “Defining Predictive Maturity for Validated Numerical Simulations,” *Technical Report of the Fiscal Year 2008 Global Nuclear Energy Partnership Program*, Los Alamos National Laboratory, Los Alamos, New Mexico, September 2008. LA-UR-08-6741.

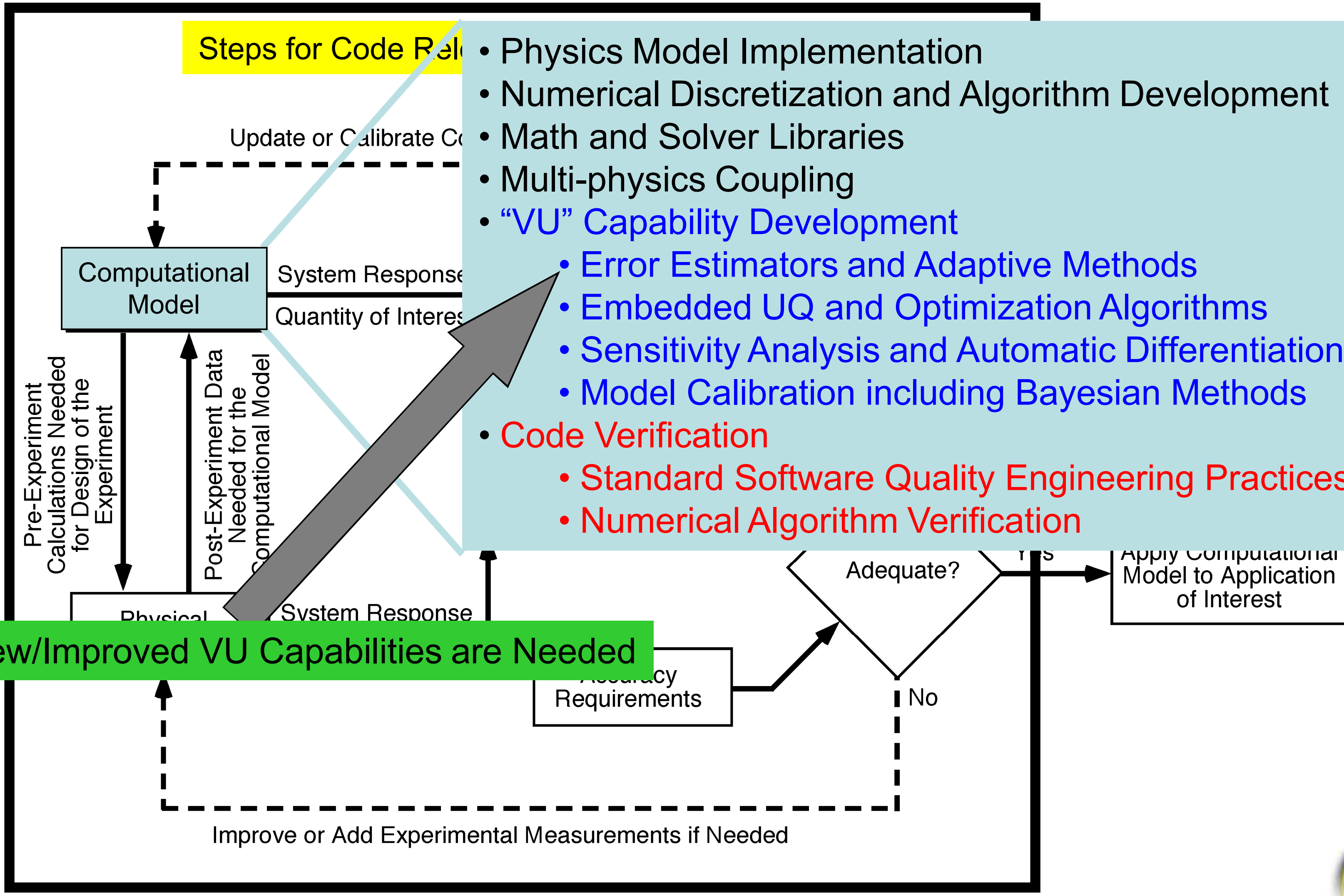
# How Is “Born Assessed” Different From the CSAU Methodology?

- Born Assessed is similar to CSAU in some ways, e.g.,
  - Formal quantification of uncertainties
  - Use of SETs and IETs
- However, Born Assessed builds on modern V&V ideas, e.g.,
  - Verification explicitly called out (to be discussed more later…)
    - Code verification
    - Solution verification
  - Formal role of validation: Model calibration and improvement through feedback of new experimental data
- Dynamic “Born Assessed” process vs. CSAU process that was developed for legacy, frozen software tools

# What’s Inside the Computational Model?



# What's Inside the Computational Model?




- Physics Model Implementation
- Numerical Discretization and Algorithm Development
- Math and Solver Libraries
- Multi-physics Coupling
- "VU" Capability Development
  - Error Estimators and Adaptive Methods
  - Embedded UQ and Optimization Algorithms
  - Sensitivity Analysis and Automatic Differentiation
  - Model Calibration including Bayesian Methods
- Code Verification
  - Standard Software Quality Engineering Practices
  - Numerical Algorithm Verification

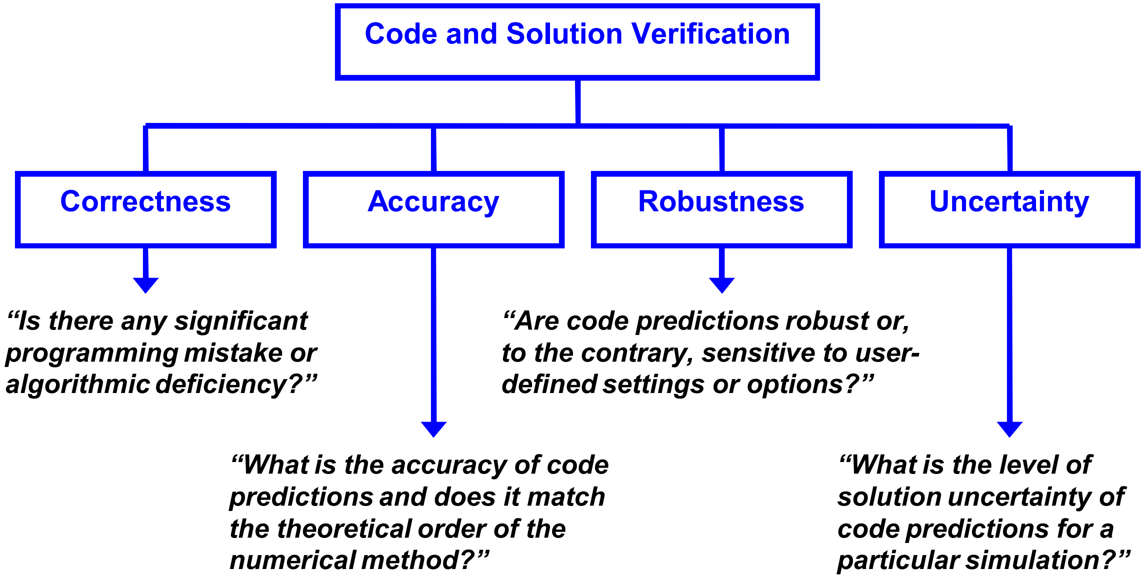
New/Improved VU Capabilities are Needed

# Outline

- V&V-based Born-Assessed Framework
- *Verification*
- VU Tools and Algorithms
  - Capability Gaps
  - Integration
- Towards Licensing

# Verification

- ***Goal:*** *To demonstrate convergence to the correct answer for the intended application (i.e., the actual system of mathematical equations)*
- Numerical errors can contribute to incorrect stockpile decisions or validation assessments
- Challenges
  - Multi-physics, multi-scale, non-smooth solutions, contact, singularities, etc.



# Verification:
## Two Main Components

- Code Verification
- Solution Verification

# Code Verification

**Traditional SQE Activities**

- Design review/inspection
- Code Review/inspection
- Pair programming
- Configuration management
- Requirements management
- Defect tracking
- Unit testing
- Regression testing
- Integration testing
- Low volume beta testing
- High volume beta testing
- Release & distribution management

**Activities for scientific software**

- Exact solutions
  - Open form
  - Closed form
  - Manufactured
- Order verification
  - Single physics
  - Tightly coupled multiphysics
  - Loosely coupled multiphysics
- Application-focused test coverage analysis

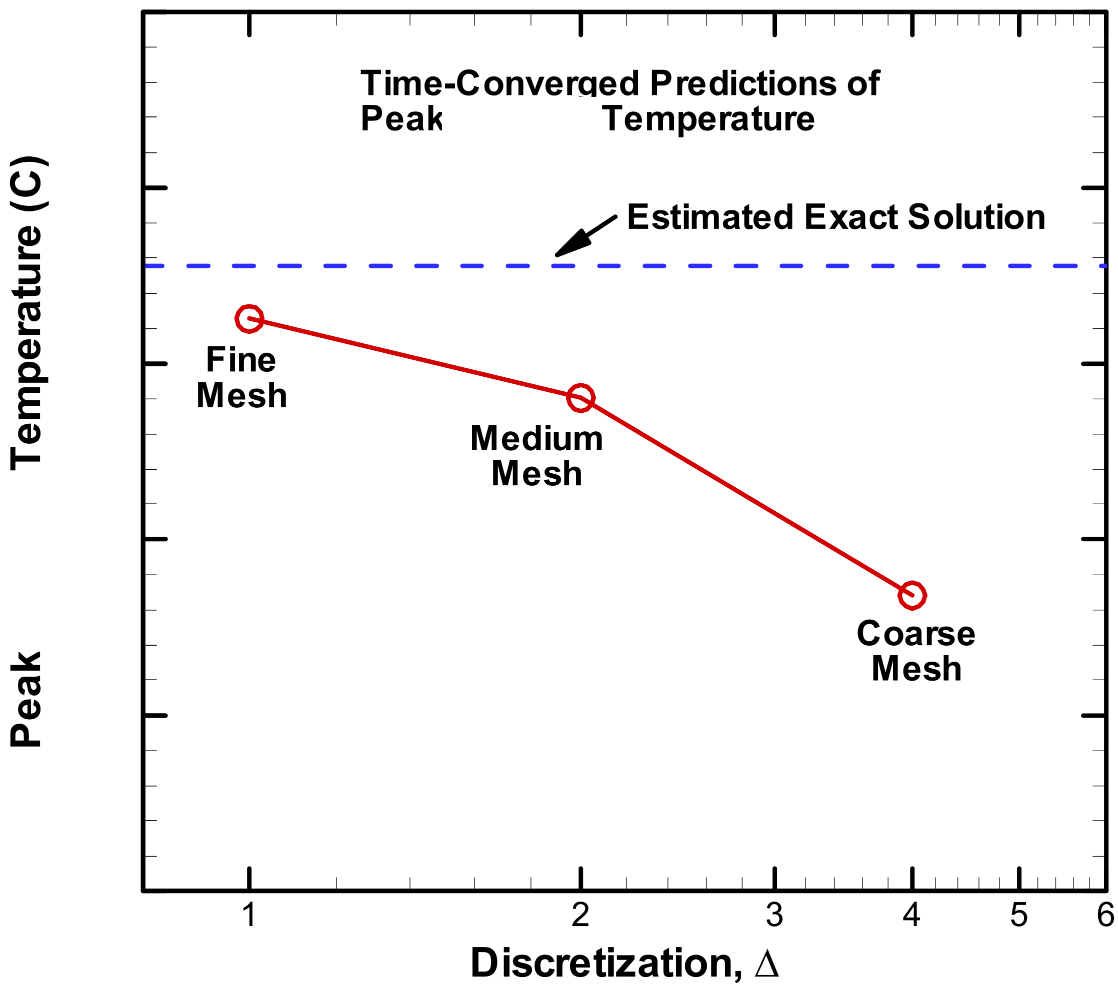
# Solution Verification

Solution verification addresses the following questions:

- In the context of *model validation*:
  - Are numerical errors obscuring or undermining comparisons of calculations with experimental data?
- In the context of *predictive simulation*:
  - Is the solution accuracy adequate for the intended application?

# Weak Link Failure Model (SNL's Calore)

## Solution verification: Is the discretization adequate?




**Component is highly non-isotropic and (initially) not adequately mesh converged**

# Outline

- V&V-based Born-Assessed Framework
- Verification
- *VU Tools and Algorithms*
  - *Capability Gaps*
  - *Integration*
- Towards Licensing

# VU Capability Gaps

- *Main Issue:* Standard “Legacy” VU Methods are generally *not robust*, *too expensive* and/or *not scalable* for our new large-scale, multi-physics IPSCs; e.g.,
  - *Sensitivity Analysis*: Sampling, Response Surfaces, etc.
  - *UQ*: Monte-Carlo and Latin Hypercube Sampling
  - *Solution Verification*: Mesh Extrapolation (e.g., Richardson)

• The NW program has delivered leading tools (e.g., DAKOTA) for VU.
• However, these tools are *starting points* for NEAMS, and are not (yet) sufficient for licensing support.
• Additional investments in VU tools and capabilities are needed.

# VU Capability Gaps: Why?

- In the ASC program, the V&V program developed *alongside* the other programs; i.e., V&V didn't exist when ASC started!
- Many "VU" algorithms have been developed and/or advanced within the last few years
  - Many require some degree of algorithm/software *integration* and *inter-operability* with the codes
  - Such integration can be expensive to retro-actively apply
- Much VU algorithm and tool development is still needed to address current deficiencies (e.g., robustness, efficiency, scalability, etc.)

# VU Capability Gap:

## A Posteriori Error Estimation

**Goal: Accurately estimate or bound the mesh discretization error in the quantity of interest**

• Based on *adjoints* which can be intrusive to the code
• Must be integrated with the code design and part of the code development activities

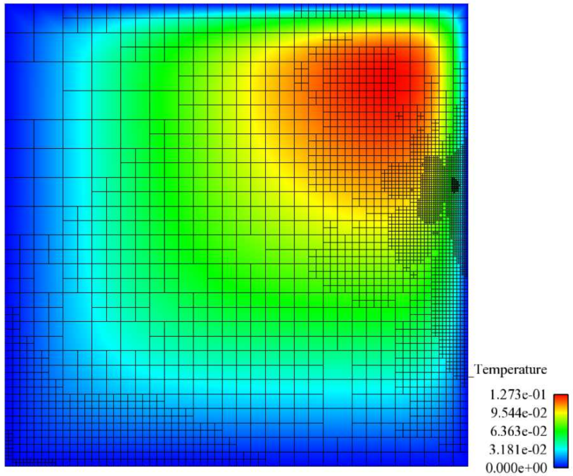
# SNL's *Encore:* Toolkit for Verification

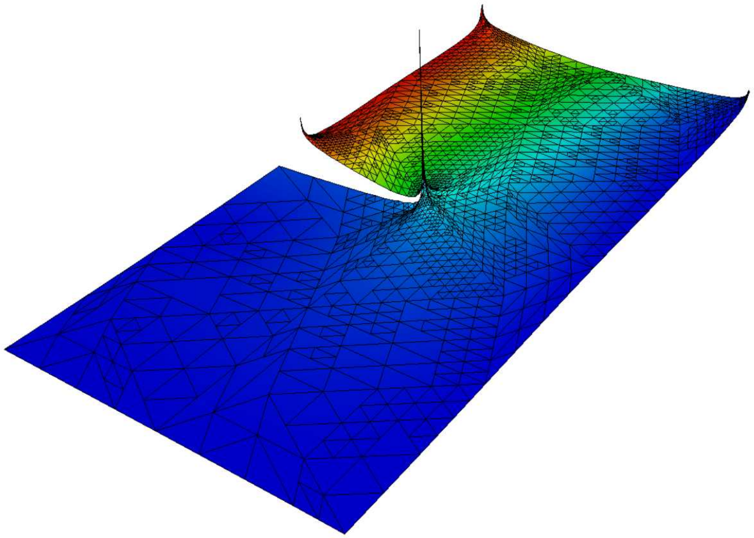
- Strategic goal: *To enable predictive simulations*
  - Unified, modular services for *code and solution verification*
  - Bridge between application codes (e.g., SIERRA Mechanics, RAMSES) and UQ tools (Trilinos, DAKOTA)
- Code verification
  - Analytical and manufactured solutions
  - Grid transfers (for comparing solutions)
  - Norms, derived quantities of interest
- Solution verification
  - Developing support for *adjoint-based* error estimators
  - Flexible, user-driven adaptivity system

| MATURITY / ELEMENT | **Maturity Level 0**<br>Low Consequence,<br>Minimal M&S Impact,<br>e.g., Scoping Studies | **Maturity Level 1**<br>Moderate Consequence,<br>Some M&S Impact,<br>e.g., Design Support | **Maturity Level 2**<br>High-Consequence,<br>High M&S Impact,<br>e.g., Qualification Support | **Maturity Level 3**<br>High-Consequence,<br>Decision Making Based on M&S,<br>e.g.. Qualification or Certification |
|---|---|---|---|---|
| **Representation and Geometric Fidelity**<br>What features are neglected because of simplifications or stylizations? | • Judgment only<br>• Little or no representational or geometric fidelity for the system and boundary conditions (BCs) | • Significant simplification or stylization of the system and BCs<br>• Geometry or representation of major components is defined | • Limited simplification or stylization of major components and BCs<br>• Geometry or representation is well defined for major components and some minor components<br>• Some peer review conducted | • Essentially no simplification or stylization of components in the system and BCs<br>• Geometry or representation of all component is at the detail of "as built", e.g., gaps, material interfaces, fasteners<br>• Independent peer review conducted |
| **Physics and Material Model Fidelity**<br>How fundamental are the physics and material models and what is the level of model calibration? | • Judgment only<br>• Model forms are either unknown or fully empirical<br>• Few, if any, physics informed models<br>• No coupling of models | • Some models are physics based and are calibrated using data from related systems<br>• Minimal or ad hoc coupling of models | • Physics-based models for all important processes<br>• Significant calibration needed using separate-effects tests (SETs) and integral-effects tests (IETs)<br>• One-way coupling of models<br>• Some peer review conducted | • All models are physics based<br>• Minimal need for calibration using SETs and IETs<br>• Sound physical basis for extrapolation and coupling of models<br>• Full, two-way coupling of models<br>• Independent peer review conducted |
| **Code Verification**<br>Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results? | • Judgment only<br>• Minimal testing of any software elements<br>• Little or no SQE procedures specified or followed | • Code is managed by SQE procedures<br>• Unit and regression testing conducted<br>• Some comparisons made with benchmarks | • Some algorithms are tested to determine the observed order of numerical convergence<br>• Some features & capabilities (F&Cs) are tested with benchmark solutions<br>• Some peer review conducted | • All important algorithms are tested to determine the observed order of numerical convergence<br>• All important F&Cs are tested with rigorous benchmark solutions<br>• Independent peer review conducted |
| **Solution Verification**<br>Are numerical solution errors and human procedural errors corrupting the simulation results? | • Judgment only<br>• Numerical errors have unknown or large effect on simulation results | • Numerical effects on relevant SRQs are qualitatively estimated<br>• Input/output (I/O) verified only by the analysts | • Numerical effects are quantitatively estimated to be small on some SRQs<br>• I/O independently verified<br>• Some peer review conducted | • Numerical effects are determined to be small on all important SRQs<br>• Important simulations are independently reproduced<br>• Independent peer review conducted |
| **Model Validation**<br>How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy? | • Judgment only<br>• 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<br>• Large or unknown experimental uncertainties | • Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs<br>• Experimental uncertainties are well characterized for most SETs, but poorly known for IETs<br>• 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<br>• Experimental uncertainties are well characterized for all IETs and SETs<br>• Independent peer review conducted |
| **Uncertainty Quantification and Sensitivity Analysis**<br>How thoroughly are uncertainties and sensitivities characterized and propagated? | • Judgment only<br>• Only deterministic analyses are conducted<br>• Uncertainties and sensitivities are not addressed | • Aleatory and epistemic (A&E) uncertainties propagated, but without distinction<br>• Informal sensitivity studies conducted<br>• Many strong UQ/SA assumptions made | • A&E uncertainties segregated, propagated and identified in SRQs<br>• Quantitative sensitivity analyses conducted for most parameters<br>• Numerical propagation errors are estimated and their effect known<br>• Some strong assumptions made<br>• Some peer review conducted | • A&E uncertainties comprehensively treated and properly interpreted<br>• Comprehensive SAs conducted for parameters and models<br>• Numerical propagation errors are demonstrated to be small<br>• No significant UQ/SA assumptions made<br>• Independent peer review conducted |

# Adjoint-Based Error Estimators and Adaptivity in SIERRA Mechanics

- Temperature field from thermal advection-diffusion example.
- *Quantity of interest:* Temperature at a point near the right boundary.
- The adjoint error estimator produces adaptivity that is optimal for this output.



- Nonlinear quasi-statics example.

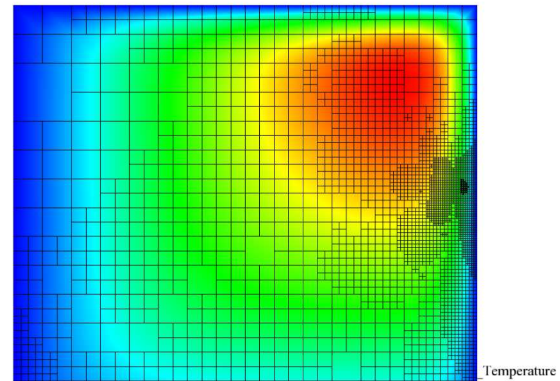
Elevation of Von Mises stress field colored by magnitude of adjoint displacement field.

*Quantity of interest:* Integral surface traction on the upper left surface.

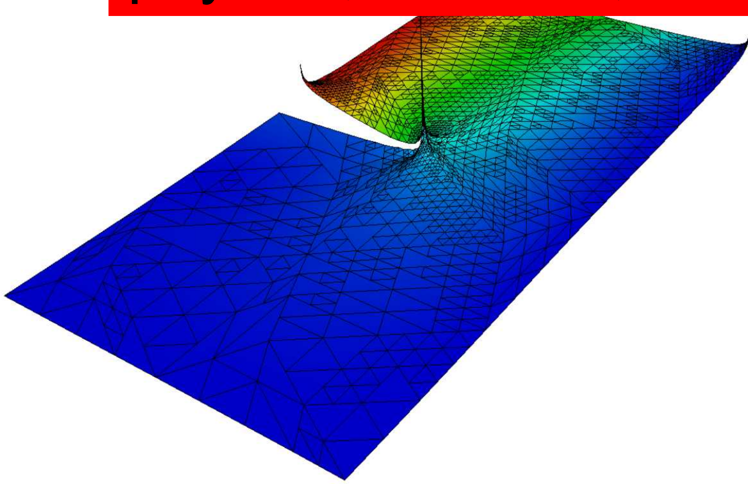
The adaptivity resolves stress singularities critical to calculation of an accurate force-displacement curve.

# Adjoint-Based Error Estimators and Adaptivity in SIERRA Mechanics

- Temperature field from thermal advection-diffusion example.
- *Quantity of interest:* Temperature at a point near the right boundary.
- The adjoint error estimator produces ada




ISSUE: Only available for simple problems; much capability development still needed (nonlinear, multi-physics, scalable, etc.)



Elevation of von Mises stress field colored by magnitude of adjoint displacement field.

*Quantity of interest:* Integral surface traction on the upper left surface.

The adaptivity resolves stress singularities critical to calculation of an accurate force-displacement curve.

# VU Capability Gap:
# Embedded Algorithms

- Algorithm Infrastructure
  - Automatic Differentiation (AD)
  - Adjoint solution technology
- Applications of *Embedded Algorithms*
  - Sensitivity analysis
  - Uncertainty Quantification
  - Goal-oriented (adjoint-based) error estimation

| MATURITY / ELEMENT | **Maturity Level 0**<br>**Low Consequence, Minimal M&S Impact, e.g., Scoping Studies** | **Maturity Level 1**<br>**Moderate Consequence, Some M&S Impact, e.g., Design Support** | **Maturity Level 2**<br>**High-Consequence, High M&S Impact, e.g., Qualification Support** | **Maturity Level 3**<br>**High-Consequence, Decision Making Based on M&S, e.g.. Qualification or Certification** |
|---|---|---|---|---|
| **Representation and Geometric Fidelity**<br>**What features are neglected because of simplifications or stylizations?** | • Judgment only<br>• Little or no representational or geometric fidelity for the system and boundary conditions (BCs) | • Significant simplification or stylization of the system and BCs<br>• Geometry or representation of major components is defined | • Limited simplification or stylization of major components and BCs<br>• Geometry or representation is well defined for major components and some minor components<br>• Some peer review conducted | • Essentially no simplification or stylization of components in the system and BCs<br>• Geometry or representation of all components is at the detail of "as built," e.g., gaps, material interfaces, fasteners<br>• Independent peer review conducted |
| **Physics and Material Model Fidelity**<br>**How fundamental are the physics and material models and what is the level of model calibration?** | • Judgment only<br>• Model forms are either unknown or fully empirical<br>• Few, if any, physics-informed models<br>• No coupling of models | • Some models are physics based and are calibrated using data from related systems<br>• Minimal or ad hoc coupling of models | • Physics-based models for all important processes<br>• Significant calibration needed using separate-effects tests (SETs) and integral-effects tests (IETs)<br>• One-way coupling of models<br>• Some peer review conducted | • All models are physics based<br>• Minimal need for calibration using SETs and IETs<br>• Sound physical basis for extrapolation and coupling of models<br>• Full, two-way coupling of models<br>• Independent peer review conducted |
| **Code Verification**<br>**Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results?**<br><br>**Solution Verification**<br>**Are numerical solution errors and human procedural errors corrupting the simulation results?** | • Judgment only<br>• Minimal testing of any software elements<br>• Little or no SQE procedures specified or followed<br><br>• Judgment only<br>• Numerical errors have unknown or large effect on simulation results | • Code is managed by SQE procedures<br>• Unit and regression testing conducted<br>• Some comparisons made with benchmarks<br><br>• Numerical effects on relevant SRQs are qualitatively estimated<br>• Input/output (I/O) verified only by the analysts | • Some algorithms are tested to determine the observed order of numerical convergence<br>• Some features & capabilities (F&Cs) are tested with benchmark solutions<br>• Some peer review conducted<br><br>• Numerical effects are quantitatively estimated to be small on some SRQs<br>• I/O independently verified<br>• Some peer review conducted | • All important algorithms are tested to determine the observed order of numerical convergence<br>• All important F&Cs are tested with rigorous benchmark solutions<br>• Independent peer review conducted<br><br>• Numerical effects are determined to be small on all important SRQs<br>• Important simulations are independently reproduced<br>• Independent peer review conducted |
| **Model Validation**<br>**How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy?** | • Judgment only<br>• 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<br>• Large or unknown experimental uncertainties | • Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs<br>• Experimental uncertainties are well characterized for most SETs, but poorly known for IETs<br>• 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<br>• Experimental uncertainties are well characterized for all IETs and SETs<br>• Independent peer review conducted |
| **Uncertainty Quantification and Sensitivity Analysis**<br>**How thoroughly are uncertainties and sensitivities characterized and propagated?** | • Judgment only<br>• Only deterministic analyses are conducted<br>• Uncertainties and sensitivities are not addressed | • Aleatory and epistemic (A&E) uncertainties propagated, but without distinction<br>• Informal sensitivity studies conducted<br>• Many strong UQ/SA assumptions made | • A&E uncertainties segregated, propagated, and identified in SRQs<br>• Quantitative sensitivity analyses conducted for most parameters<br>• Numerical propagation errors are estimated and their effect known<br>• Some strong assumptions made<br>• Some peer review conducted | • A&E uncertainties comprehensively treated and properly interpreted<br>• Comprehensive SAs conducted for parameters and models<br>• Numerical propagation errors are demonstrated to be small<br>• No significant UQ/SA assumptions made<br>• Independent peer review conducted |

**Adjoints and AD are key capabilities for increased efficiency and accuracy!**

# **VU Challenge:** Integration

- Many of these new algorithms require tight integration with codes and models
- *Many ASC codes already too "mature" to capitalize on new VU capabilities without significant investments in software re-architecting (e.g., adjoints and sensitivities)*

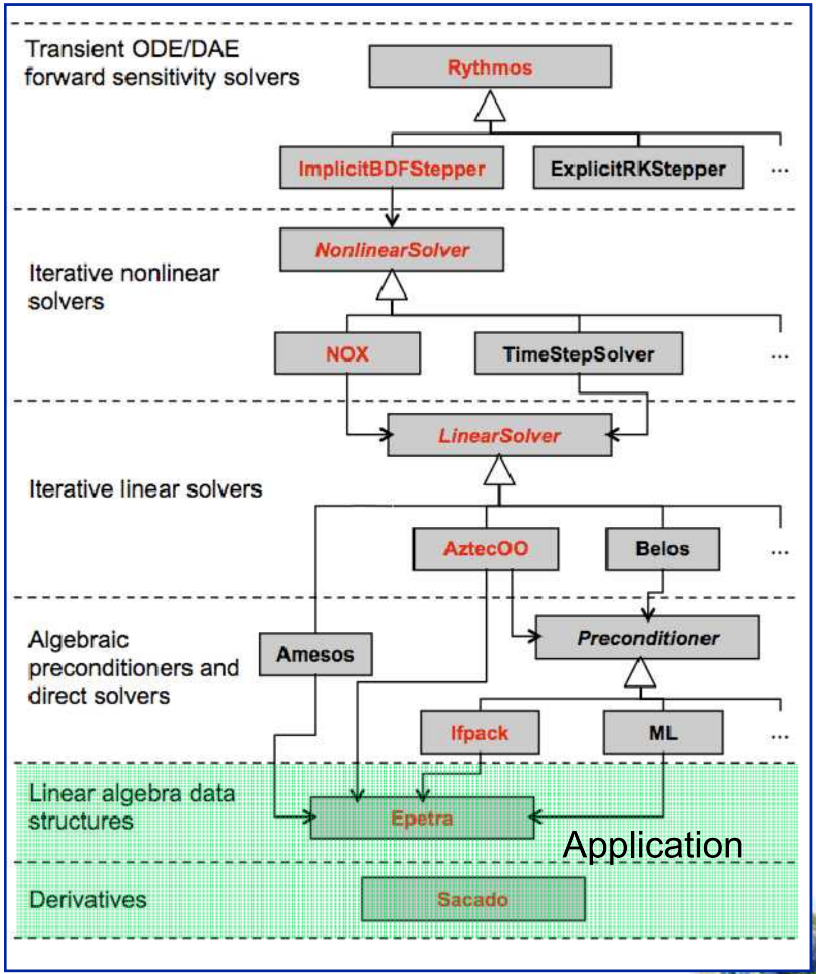
# SNL's Sacado Package: *Automatic Differentiation*

- *Sacado:* Operator overloading AD tools for C++ applications
- Part of the Trilinos library
- Impacting UQ:
  - Stochastic Galerkin methods (ongoing)
  - Polynomial chaos methods (near term)
  - Epistemic UQ
    - Intervals (mid term)
    - Probability boxes (long term)

# **VU Integration:** Using Sacado's Advance Capabilities for Sensitivity Analysis

- Requires integration with
  - Solvers
  - Application code

Vertical integration of Trilinos capabilities



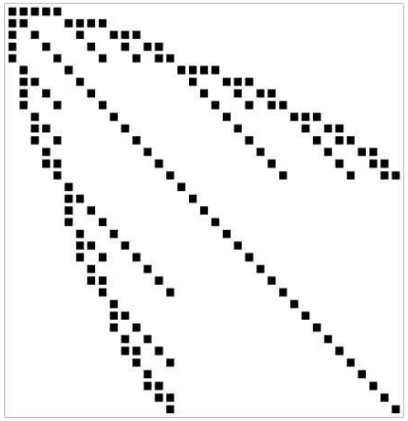
# VU Capability Gap: *UQ*

## Implementing Stochastic Galerkin Methods in Nonlinear Applications is Challenging

- Code transformation from deterministic code to SG code
  - Need tools/libraries to automate computation of SG residual and Jacobian entries

$$F(u;p) = 0 \rightarrow \bar{F}(\bar{u}) = \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_{N_{PC}} \end{bmatrix} = 0$$

<u>Large linear systems: Each block is the size of a single “deterministic” block</u>



**SNL’s Trilinos provides powerful capabilities here**

# SNL's DAKOTA
# Optimization and UQ Toolkit

- Wide array of non-embedded (black box) capabilities
- Available as an *early delivery vehicle* for the VU capabilities developed in NEAMS (…along with other software available at LANL and SNL)
- *DAKOTA/FRAPCON-3 Coupling*
  - FY07 GNEP Fuels Campaign project at SNL
  - Goal: Apply DAKOTA to assess the sensitivity and quantify uncertainty of a FRAPCON-3 response variable relative to uncertain model parameters
  - Included extensive PIRT analysis

# New DAKOTA UQ Capabilities

## *Advanced Components for Born-Assessed Codes*

| | Production | New | Under dev. | Planned | Collabs. |
|---|---|---|---|---|---|
| Sampling | LHS/MC, QMC/CVT | IS/AIS/MMAIS, Incremental LHS | | Bootstrap, Jackknife | Gunzburger |
| Reliability | $1^{st}/2^{nd}$-order local: MVFOSM/SOSM, x/u AMV/AMV$^2$/ AMV+/AMV$^2$+, x/u TANA, FORM/SORM | Global: EGRA | | | Renaud, Mahadevan |
| Polynomial Chaos | | Wiener-Askey gPC: sampling, quadrature, pt collocation | Cubature | Adaptivity, Wiener-Haar | Ghanem |
| Other probabilistic | | | | Dimension reduction | Youn |
| Epistemic | Second-order probability | Dempster-Shafer evidence theory | | Bayesian, Imprecise probability | Higdon, Williams, Ferson |
| Metrics | Importance factors, Partial correlations | Main effects, Variance-based decomposition | Stepwise regression | | Storlie |

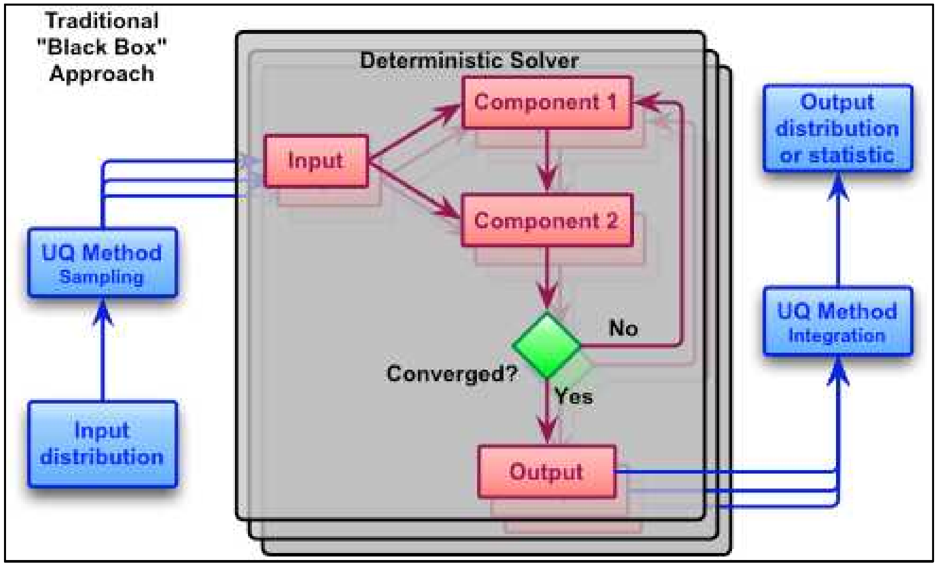
# **VU Integration:** Some Remarks

- DAKOTA can be used as a starting point for NEAMS UQ, optimization, and sensitivity analyses, but must be further developed and appropriately tailored to support science-base licensing applications.
- NW-funded tools such as DAKOTA and Trilinos can be leveraged for developing Born-Assessed codes for NEAMS.

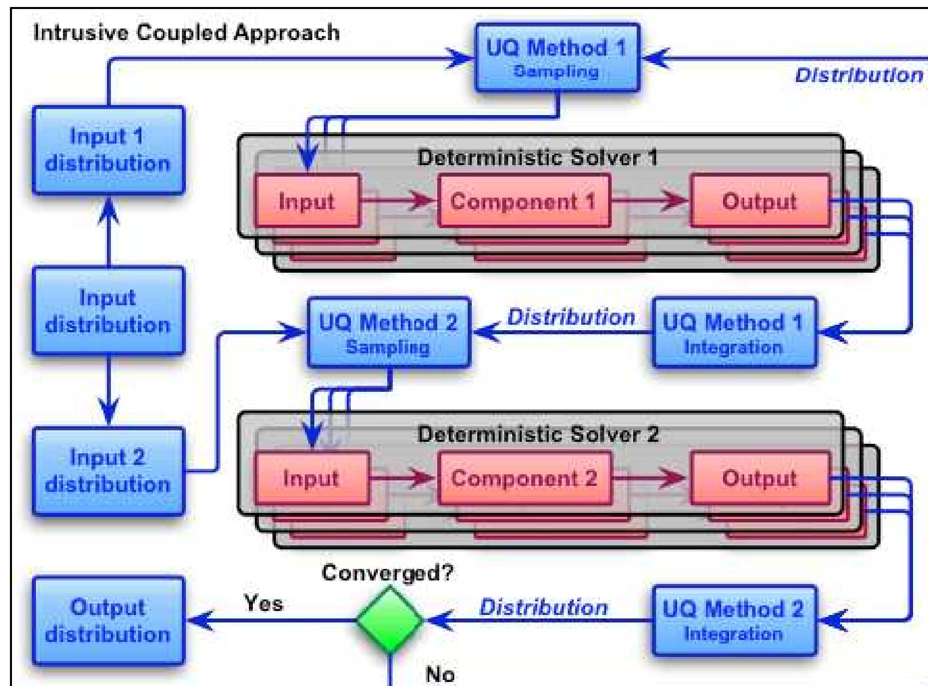
# *NEAMS Leveraging Opportunity:*

# Coupled System Embedded UQ Research Project

(Led by SNL, Teaming with Ghanem at USC)




- Invert layering of UQ around system simulation
  - Apply UQ to each component separately
  - Stochastic coupled solver technology
- Potentially orders of magnitude savings
  - Heterogeneous UQ
  - Stochastic dimension reduction




- FY08 SNL Late-Start LDRD
  - Mathematical feasibility demonstrated
- Successful ASCR complex systems proposal (Beginning FY09, Post-CR)
  - Sandia ~$600K/yr, USC ~$200K/yr
  - *Emphasis on NE applications (reactor core to entire plant)*
  - Applicability to NEAMS

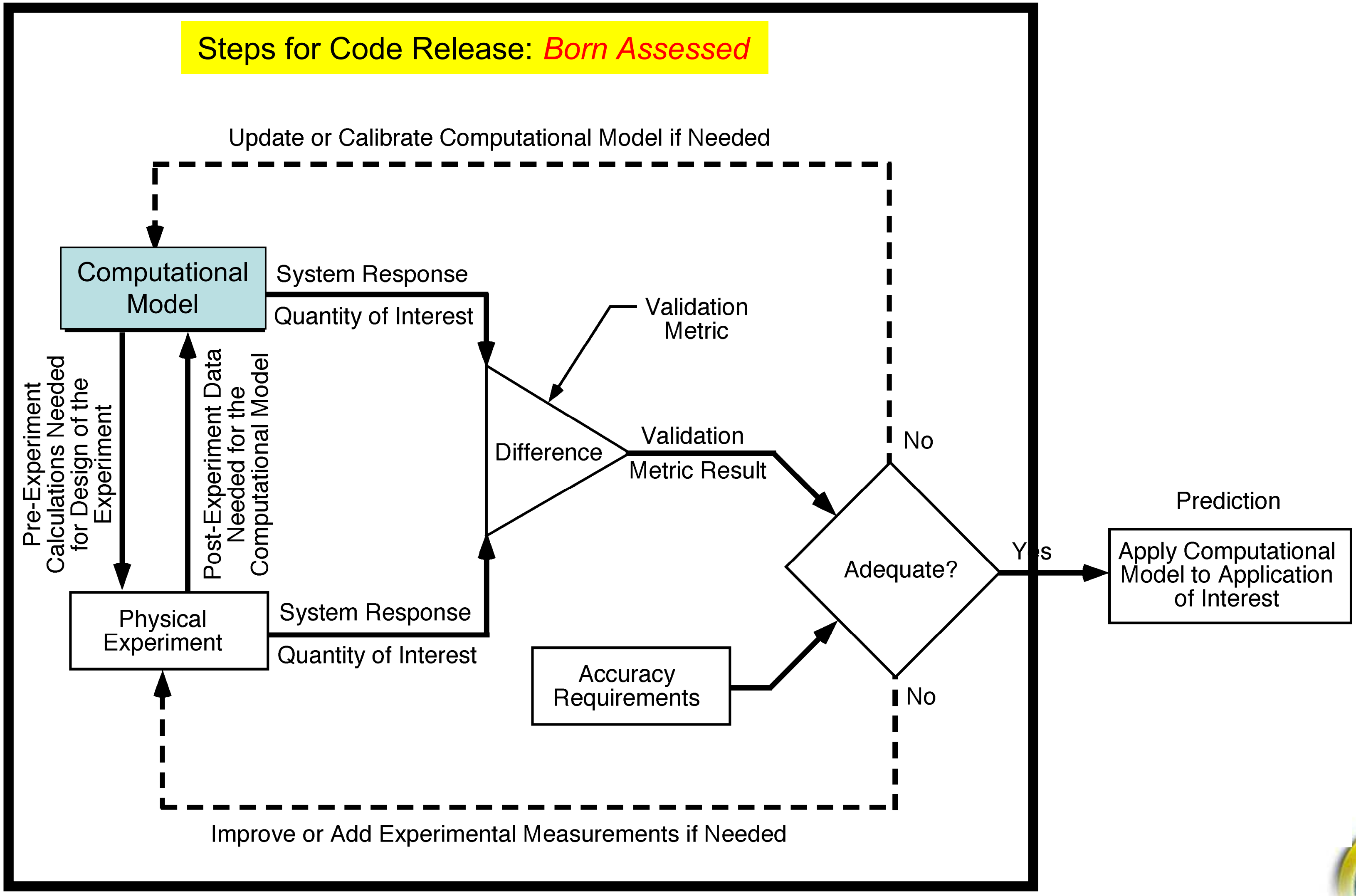
# Outline

- V&V-based Born-Assessed Framework
- Verification
- VU Tools and Algorithms
  - Capability Gaps
  - Integration
- *Towards Licensing*

# Towards Licensing:
## Using Science-Based Modeling and Simulation for Risk-Informed Decisions

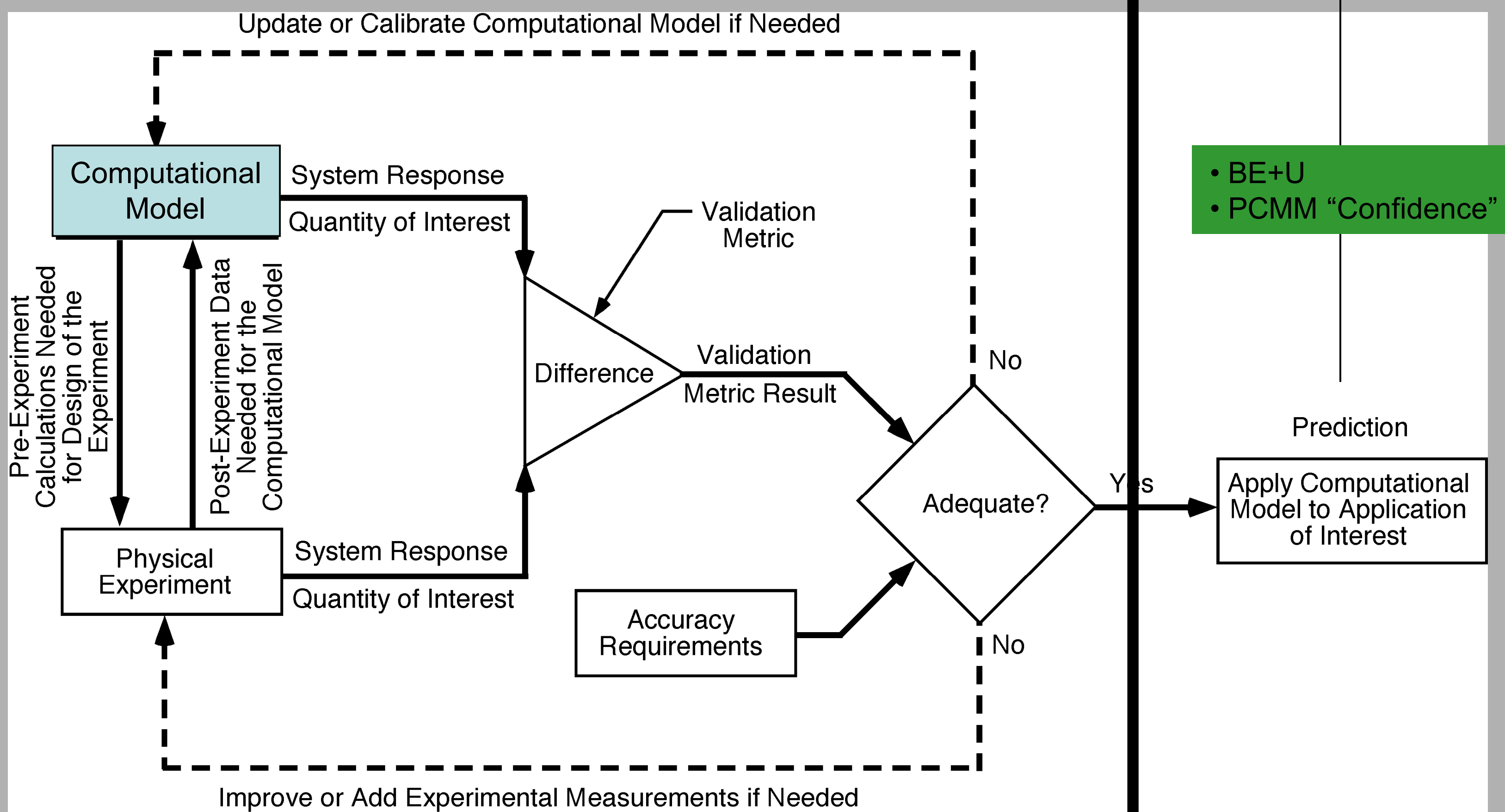
- Key Ingredients
  - Best Estimate + Uncertainty
  - Measure of Pedigree or “Confidence”: Predictive Capability Maturity Model (PCMM)
  - Treatment of other factors including “Unknown Unknowns”
- **There is more to risk-informed decision making than computational science**

# "Informing" the Licensing Process



# “Informing” the Licensing Process

## Steps for Code Release: *Born Assessed*



# Summary:
# Three Main Themes

1. The Born-Assessed Framework can provide a science-based approach to risk-informed decisions for supporting licensing.

2. The NW program has provided leading tools and capabilities that can serve as starting points for the NEAMS VU program. However, *significant new VU tools and capabilities are broadly needed* to achieve the NEAMS goals.

3. Integration of VU with the other parts of NEAMS is essential from Day 1.

**Much can be leveraged from the NW program.**

# Key VU Challenge:
## *Integration Driven by the VU Program Element*

- Integration of VU capabilities with the codes
- Integration of model development with
  - Physical experiments
  - Code development
- Integration of Born Assessed process
- Integration of V&V and UQ within the Licensing Framework

***Key Lesson Learned:***
***Integration is Essential from Day 1***