Representation and Propagation of Uncertainty for Tabular Multiphase Equation-of-State Models

Allen C. Robinson, John H. Carpenter, Ann E. Mattsson,
Richard D. Drake, William J. Rider
Computational Shock and Multiphysics
Sandia National Laboratories
Albuquerque, NM 87185

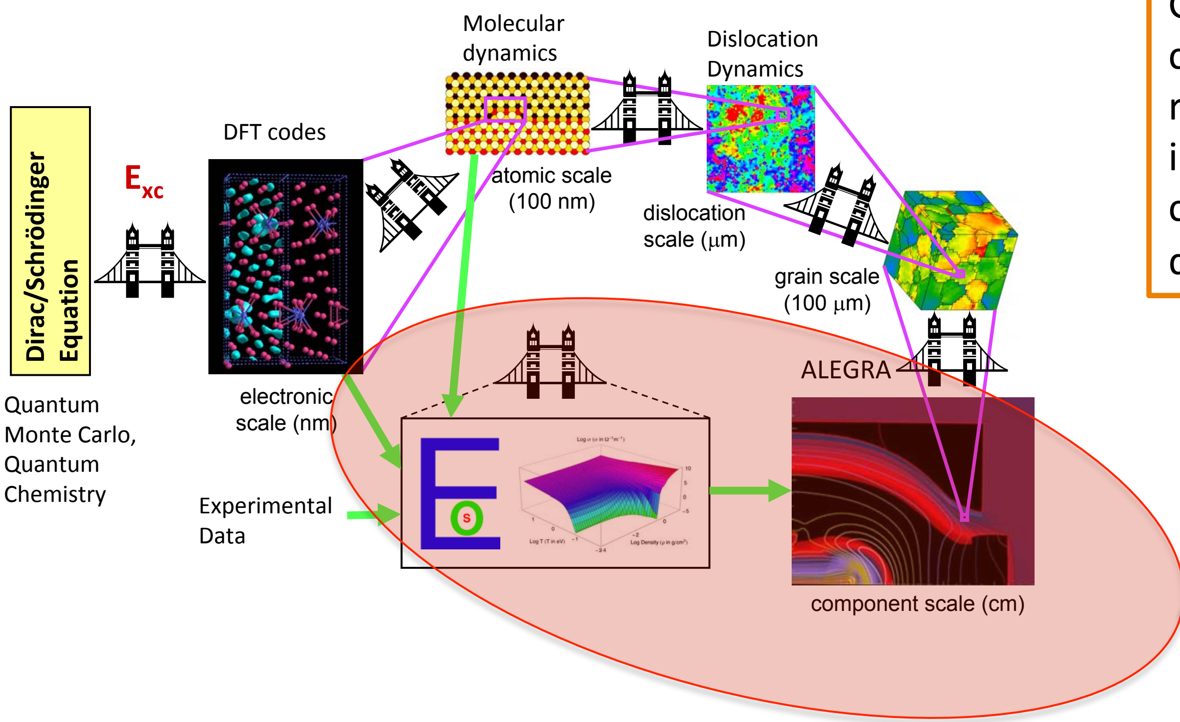
Bert Debusschere
Reacting Flow Research
Sandia National Laboratories
Livermore, CA 94551

MultiMat 2013
San Francisco, CA
Sept 2-6, 2013

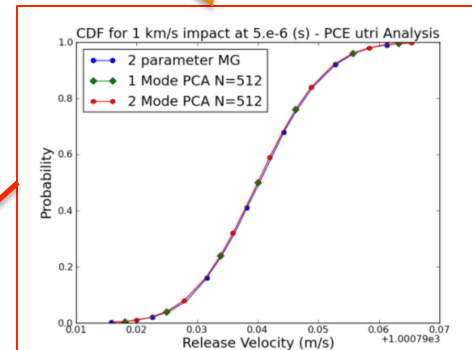
SAND2013-7181 C

Sandia National Laboratories is a multi program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
.

Building Bridges between the Fundamental Laws of Nature and Engineering



Goal: The analyst running the continuum code should easily get results that can be transformed into the equivalent of "50% chance of rain" to give to the decision maker.



Previous Focus: MultiMat 2011 Presentation

- Proof of Principle using Simple Equation of State:
- "Fundamental issues in the representation and propagation of uncertain equation of state information in shock hydrodynamics", Computers and Fluids, 83, (2013) p. 187–193

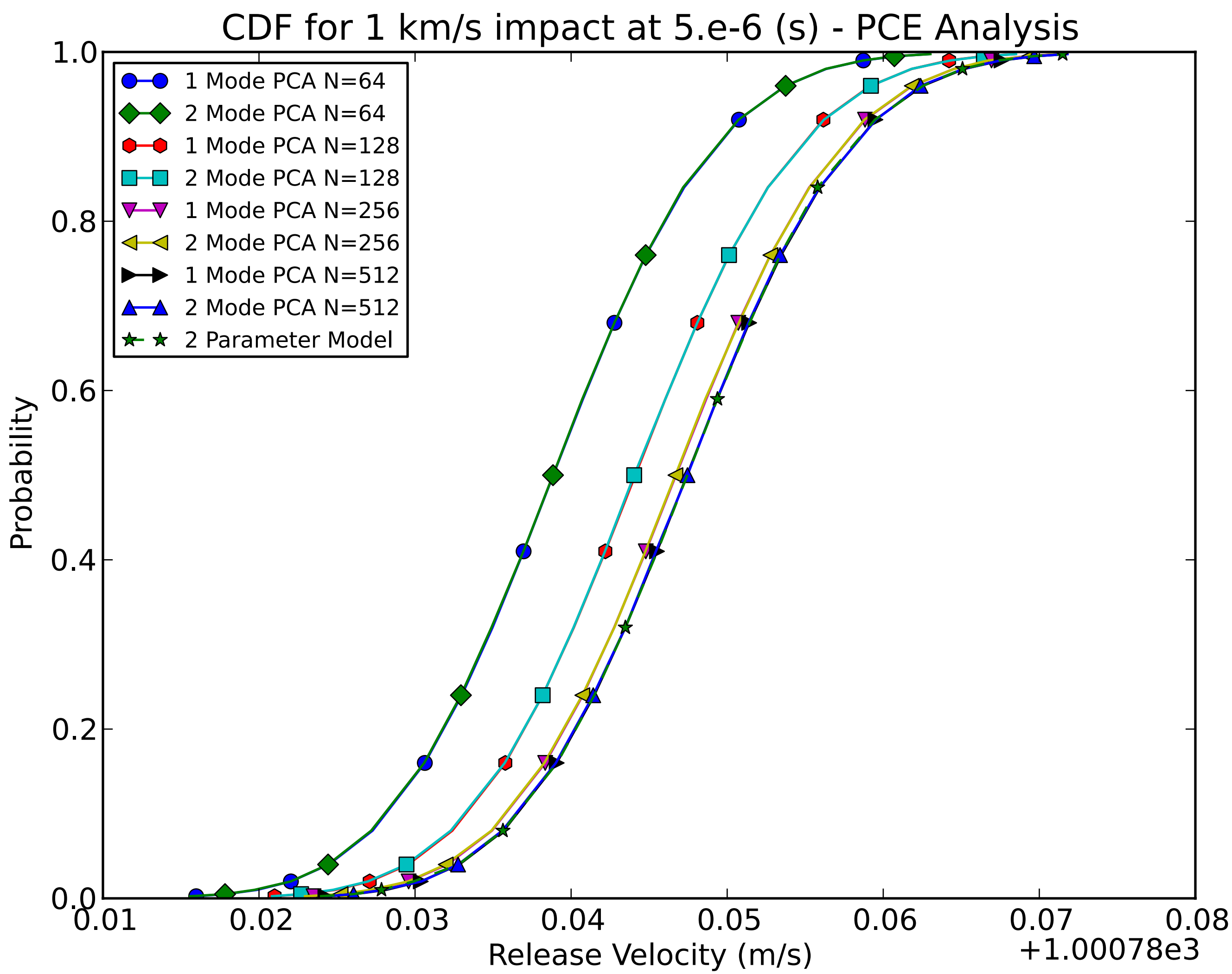
Current Focus: MultiMat 2013 : Much more difficult

Representation of uncertain multiphase tabular equations of state

Enabling Macroscale UQ Analyses

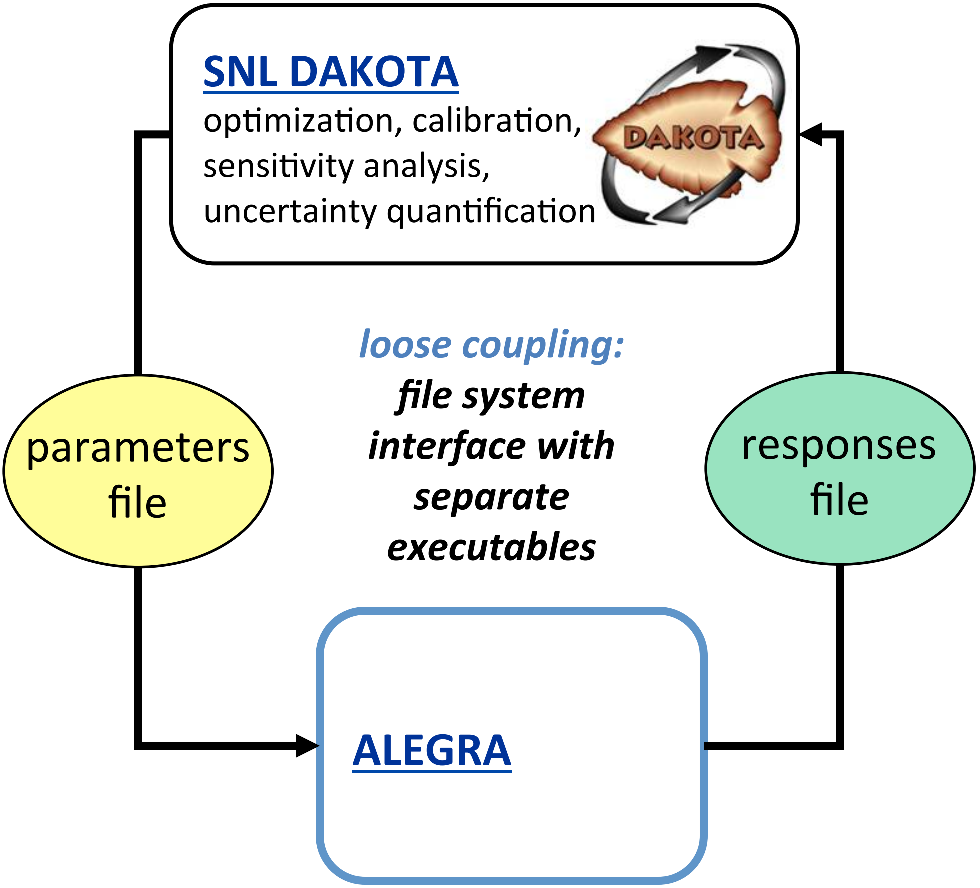
Software Package	Output
EOS model library and data	Proposal Model (XML input deck)
Bayesian Inference using Markov Chain Monte Carlo	Extensive Sampling of the posterior distribution function (PDF)
Multinormal PDF fit + Dakota	Reduced PDF samples
EOS Table Building	Topologically equivalent tables for each sample
PCA Analysis	Mean EOS table + most significant perturbations
Hydrocode + Dakota	Cumulative Distribution Function (CDF) for quantities of interest

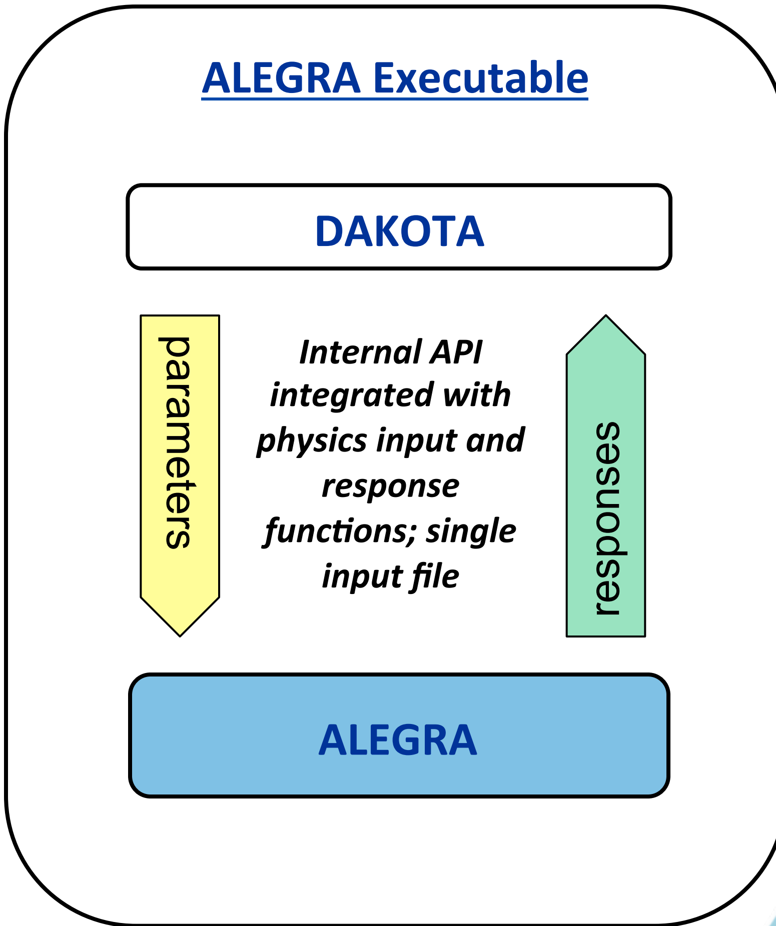
Key Ideas: Quantifiable tabular accuracy is important and Principal Components Analysis can be used to reduce the stochastic dimension

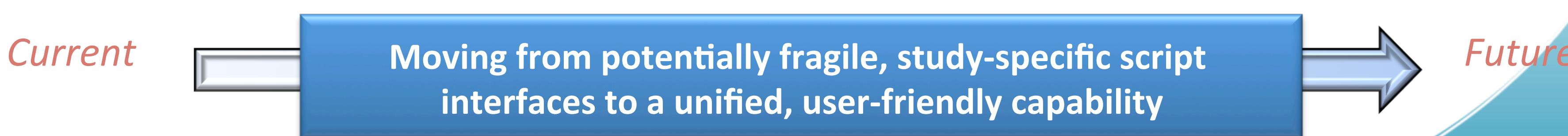


Sandia National Laboratories

Key Idea: Embedding a Dakota Interface in ALEGRA input syntax reduces "User Energy Barrier"







Major new issues for Bayesian inference for multiphase EOS tables with many parameters

Software Package	Output
EOS model library and data	Proposal Model (XML input deck)
Bayesian Inference using Markov Chain Monte Carlo	Extensive sampling of the posterior distribution function (PDF)
(Optional) Multinormal PDF fit + DAKOTA sampling	Reduced sample set
EOS Table Building	Topologically equivalent tables for each sample
PCA Analysis	Mean EOS table + most significant perturbations
Hydrocode + Dakota	Cumulative Distribution Function (CDF) for quantities of interest

Uncertain Equation of State (EOS) parameters are inferred from data using Bayesian inference

- Parametric EOS Model:

$$P = P(\rho, E; \lambda) \qquad T = T(\rho, E; \lambda)$$

- Bayes' rule updates prior belief in parameter values using data *D*

posterior, likelihood, prior, normalization

$$p(\lambda|D) = \frac{p(D|\lambda)p(\lambda)}{p(D)}$$

  - Prior: based on prior data or expert opinion
  - Likelihood: how likely is the data for given parameter values
    - Accounts for various sources of uncertainty
  - Posterior: Probability of parameter values after updating with data

Posterior distribution provides parameter values and their uncertainty

- Probability here represents the degree of belief in particular parameter values
- Various sources of uncertainty can be accounted for in the likelihood
  - Measurement noise
  - Model discrepancy
  - …

$$D = \{d_i\}_{i=1}^{N} \qquad y_i = f(x_i;\lambda)$$

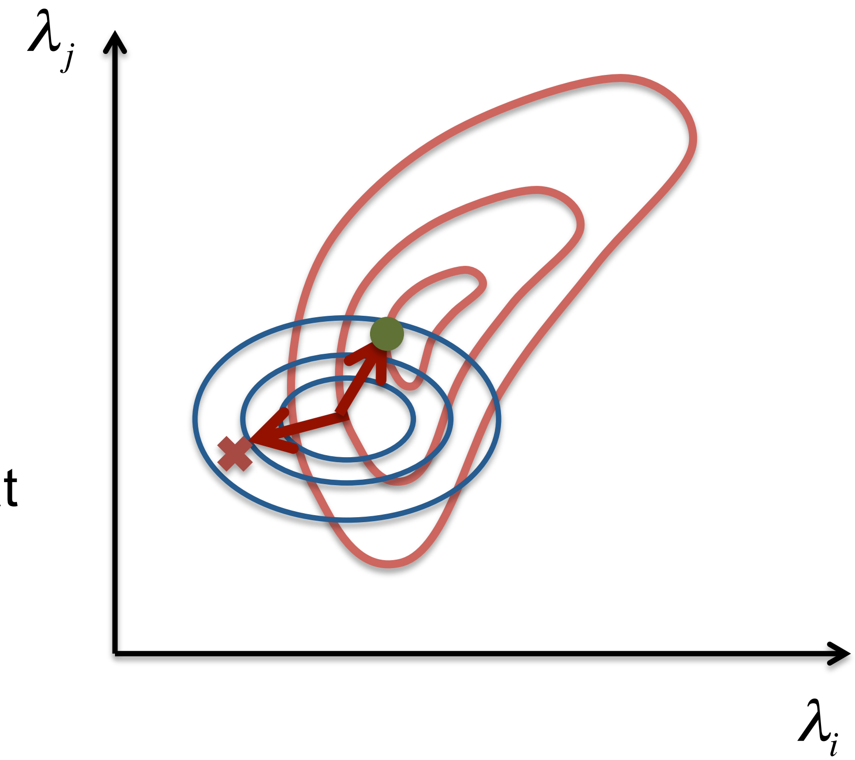
$$P(D|\lambda) = \frac{1}{(2\pi\sigma^2)^{N/2}} \prod_{i=1}^{N} \exp\left(-\frac{(d_i - y_i)^2}{2\sigma^2}\right)$$

$$\log(P(D|\lambda)) = -\frac{N}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{N} \frac{(d_i - y_i)^2}{2\sigma^2}$$

- As such, the role of data is to provide information to reduce the uncertainty
  - Multiple data sets can feed into the inference simultaneously
  - The more data, the narrower the posterior distribution

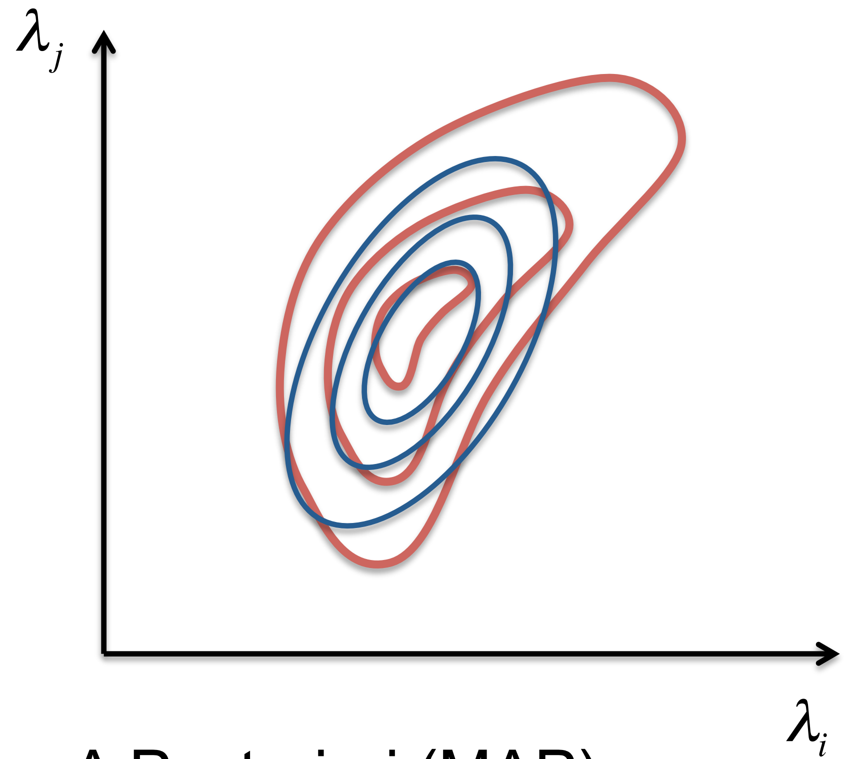
The posterior is explored with Markov Chain Monte Carlo (MCMC)

- Hard to explore
  - Often no explicit formula
  - High-dimensional
- Generally sampled with Markov Chain Monte Carlo (MCMC)
  - Generate proposal sample from Gaussian distribution centered at current state
    - Proposal distribution width determines mixing
  - Compute α as posterior ratio of new sample over old one
  - Accept new sample with probability min(α,1)



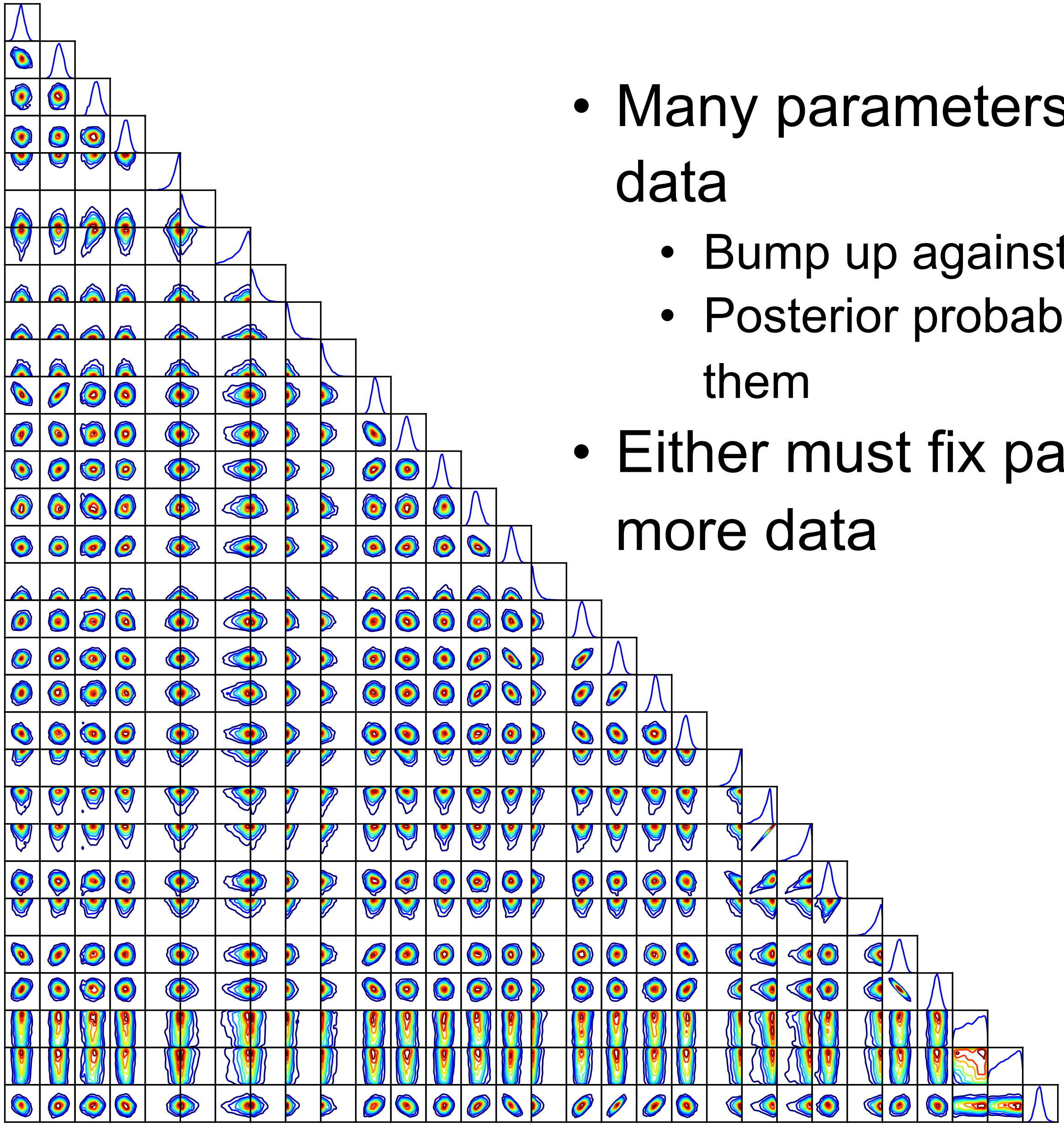
MCMC in high-dimensional, complex models is DIFFICULT and COSTLY

- Posterior shape is often non-Gaussian
  - Sometimes multimodal
- Posterior width can vary several orders of magnitude in different dimensions
- Good starting point hard to find
- Good mixing is hard to obtain
  - Needed to have good coverage



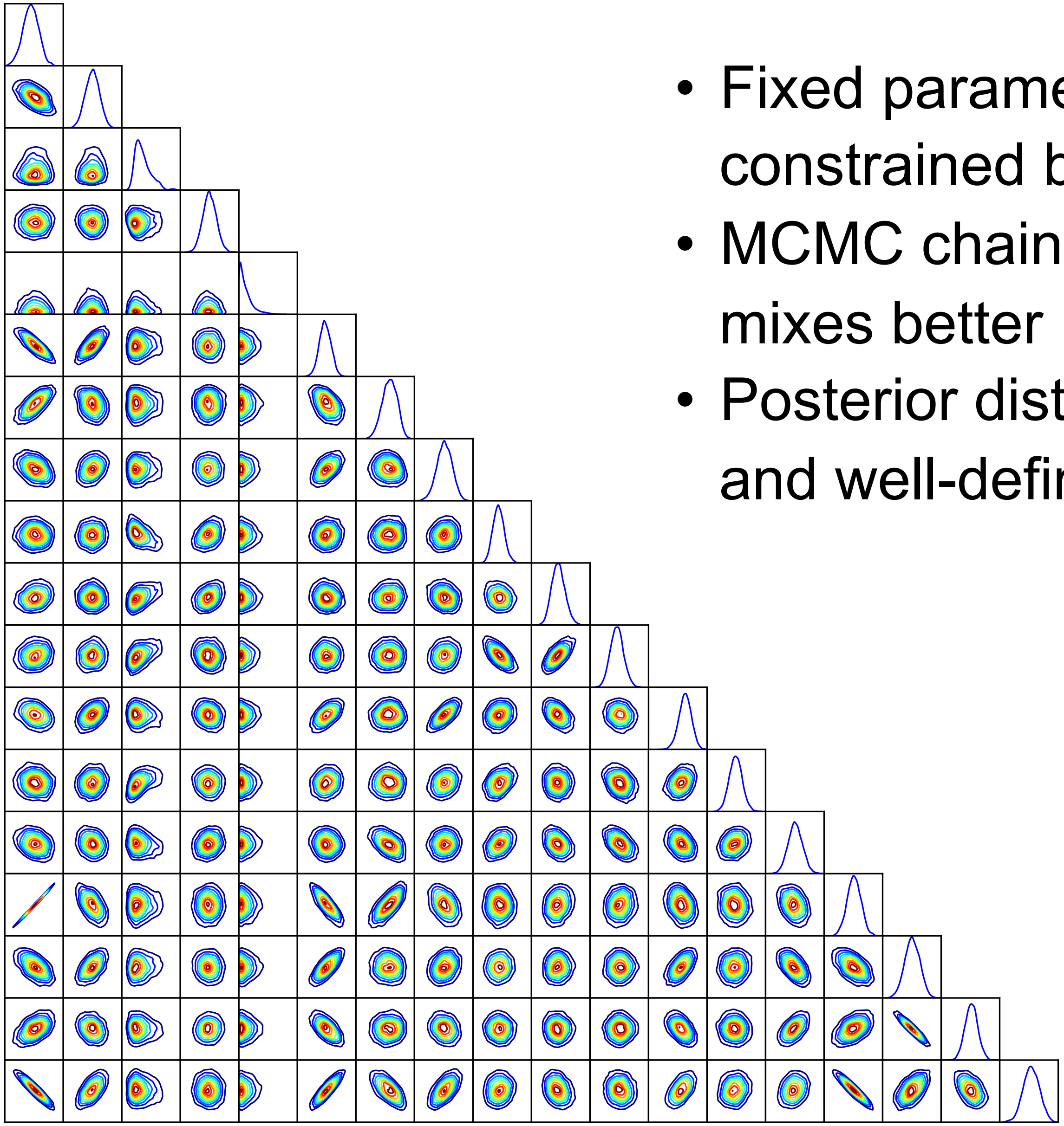
- Use optimization to find Maximum A Posteriori (MAP) parameter values to start chain from
- Use adaptive MCMC
  - Adjust proposal covariance based on previous samples
  - Use long burn-in time to ensure covariance is positive definite

30 parameter Al EOS inference



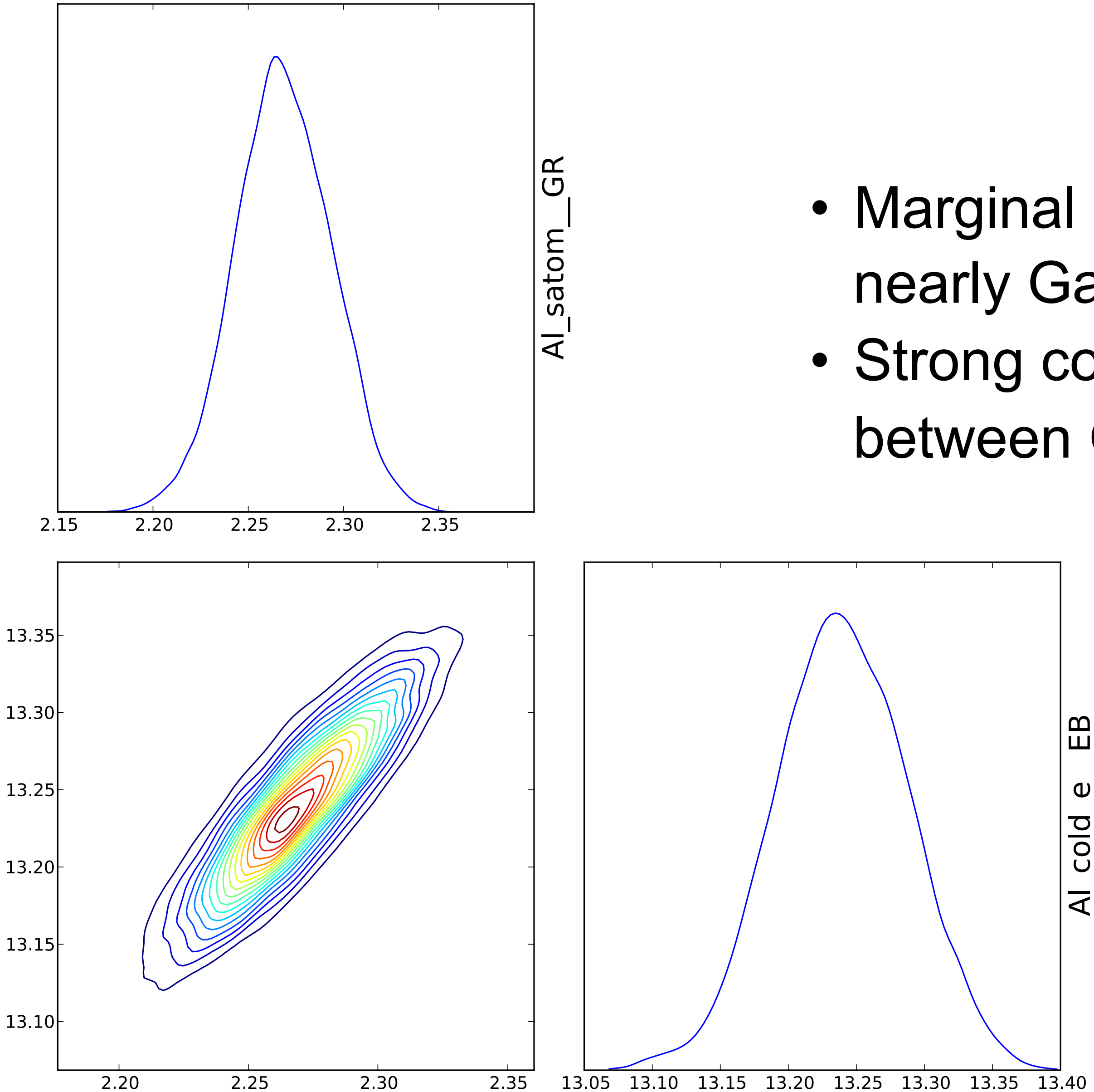
- Many parameters ill-constrained by data
  - Bump up against prior bounds
  - Posterior probability not very sensitive to them
- Either must fix parameters or add more data

18 parameter Al EOS inference



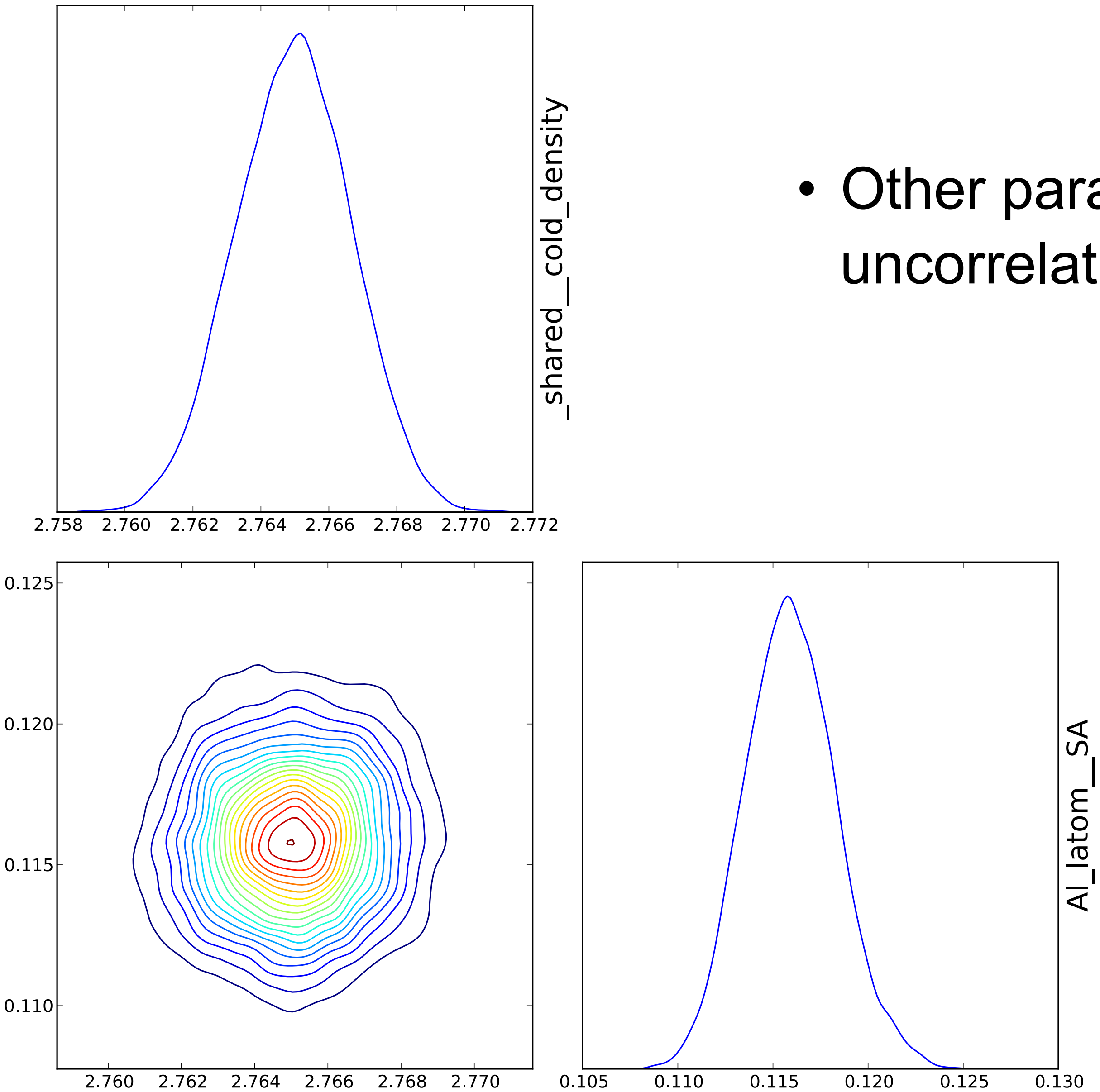
- Fixed parameters that are not well constrained by data
- MCMC chain over 18 parameters mixes better
- Posterior distributions are smoother and well-defined

Select marginal distributions



- Marginal distributions nearly Gaussian
- Strong correlation between GR and EB

Select marginal distributions



- Other parameters uncorrelated

Multiphase Table Generation and Representation

Software Package	Output
EOS model library and data	Proposal Model (XML input deck)
Bayesian Inference using Markov Chain Monte Carlo	Extensive Sampling of the posterior distribution function (PDF)
Multinormal PDF fit + Dakota	Reduced PDF samples
EOS Table Building	Topologically equivalent tables for each sample
PCA Analysis	Mean EOS table + most significant perturbations
Hydrocode + Dakota	Cumulative Distribution Function (CDF) for quantities of interest

Motivation for Tabulation

Computational and storage efficiency required by codes:

- Good scaling required from small to enormous systems
- Analytic models too slow for inline use
- Minimize data access and movement for quick start up
- Cost of tabulation amortized over many simulations

UQ analysis requires many simulation runs for sampled points in the parametric EOS space.

Several tabulation strategies proposed:

Strategy A On-the-fly generation from models

Strategy B On-the-fly generation from perturbed tables

Strategy C Ensemble of tables

Practical Table Requirements

Verified representation

- Tabulation errors cloud UQ analysis
- Represent analytic model better than data uncertainty
- Required for all variables used in codes

Efficient state look up

- One table look up per state evaluation
- Fast, vectorized interpolation

Compatibility with PCA process

- Topological equivalence between tables
- Smooth mapping between table meshes
- Consistency between independent variable spaces

New Table Format

Unstructured triangular grid (UTri)

- Easily follow phase boundaries
- Effectively capture discontinuous derivatives
- Allows adaptivity for storage efficiency
- Less efficient table look ups
- Linear interpolation on triangles

Tabulate both X(ρ, T) and X(ρ, E)

- X includes all desired thermodynamic variables
- Lose consistency only to tabulated accuracy
- No inverse look ups (i.e. iterations) required
- Special care required in PCA analysis

Determining Phase Boundaries

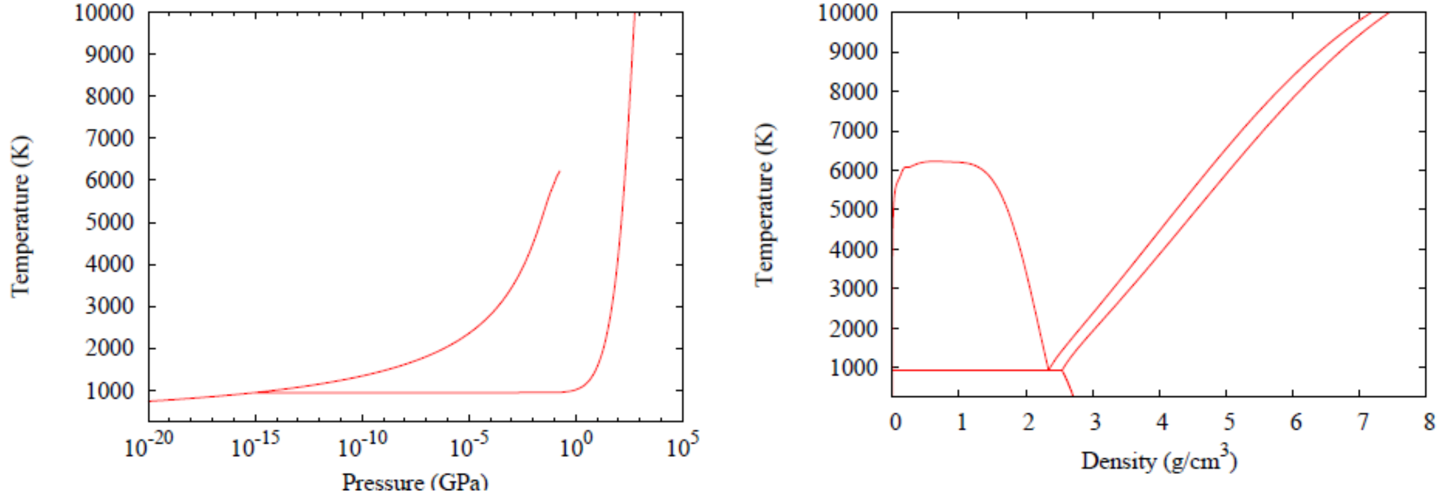
Phase boundary detection code built into EOS model library:

- Determined on demand by models
- Approximated by 1^{st}, 3^{rd}, or 5^{th} order splines
- Spline accuracy much greater than target table tolerance

Multiple uses for spline boundaries in tabulation:

- Aid in looking up states in EOS models
- Speed up state inversion algorithm
- Set boundaries for mesh regions

Multi-phase aluminum model example:

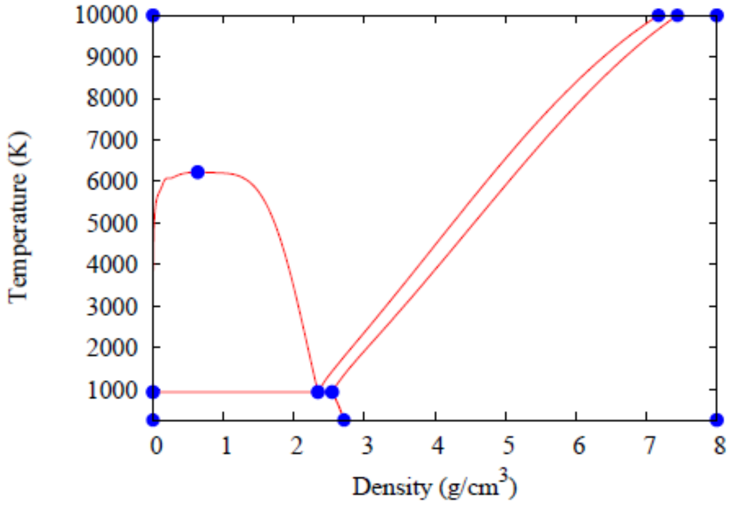


Topological Equivalence

Generate graph from phase and table boundaries in ρ-T space

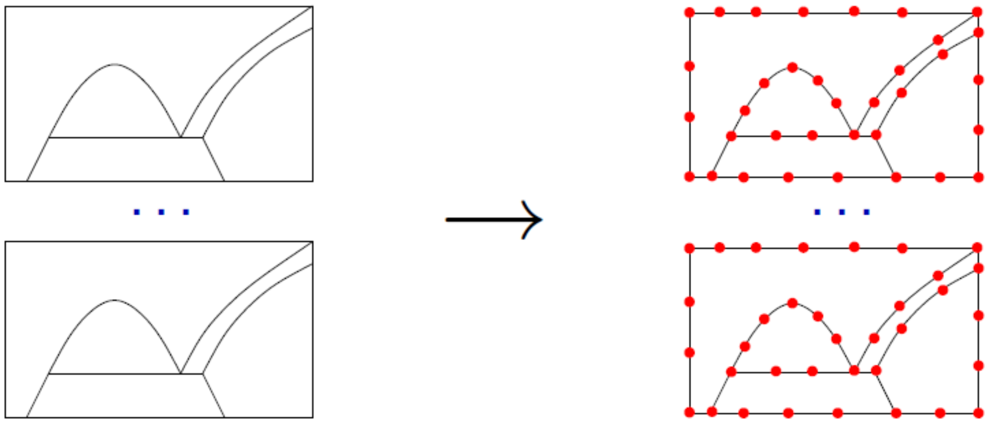
- Boundaries become edges in graph
- Nodes placed at intersections of phase boundaries
- Identical graphs between model samples indicate topological equivalence
  - PCA analysis enabled for variable node locations
  - Capture uncertainty in phase boundary behavior

Aluminum model example:



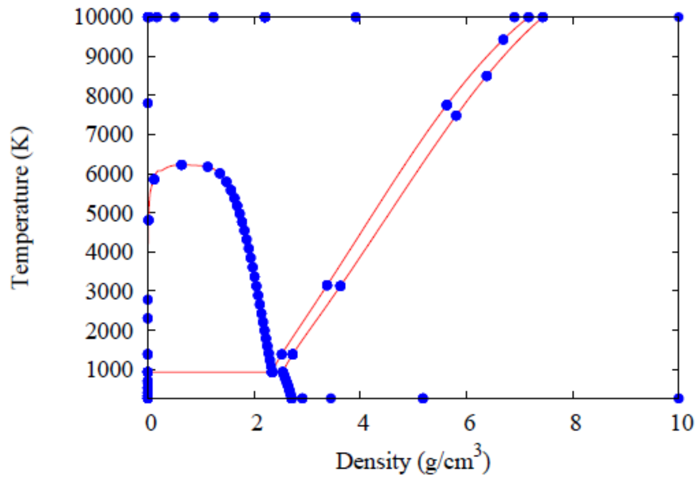
Mesh Boundaries

Adaptively mesh each boundary to a desired tolerance:



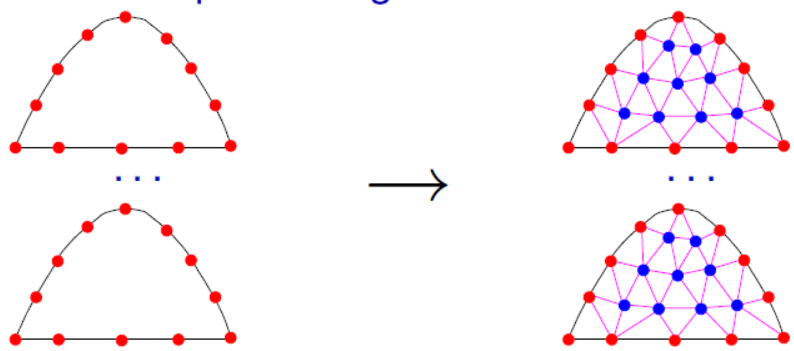
- Boundaries meshed concurrently for all N tables
- Linear map of node locations between tables

Aluminum model at 0.5 tolerance:



Mesh Regions

Adaptively mesh each phase region to a desired tolerance:



- Points placed into first table and mapped to others

Boundary point locations on each phase line segment are related by a linear mapping.

$$\frac{1}{\sqrt{g}}\frac{\partial}{\partial \xi^\alpha}\left(\sqrt{g}g^{\alpha\beta}\frac{\partial x^i}{\partial \xi^\alpha}\right) = 0$$

Or other smooth mappings?

Laplace-Beltrami smoother

Interior boundary point locations are smoothed to remove noise from PCA.
Fully consistent mesh topologies are required.

Tabular EOS UQ representation

Use Principal Component Analysis (PCA) to look for a tabular representation with reduced dimensionality:

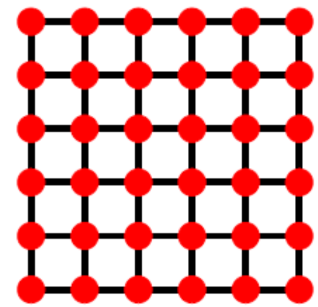
- Start with representative sample of tables (e.g. PCE integration points)
- Perform PCA:

$$\bar{z} = Z H^{1/2} \mathbf{1} / \mathbf{1}^T H \mathbf{1}$$

$$G^{1/2}(Z - \bar{z}\mathbf{1}^T) H^{1/2} = \tilde{U} \Sigma \tilde{V}^T$$

High precision floating point due to wide range table dynamic range requirements

Nx



$$z = \bar{z} + U\Sigma\xi = \bar{z} + G^{-1/2}\tilde{U}\Sigma\xi = \bar{z} + (Z - \bar{z}\mathbf{1}^T)H^{1/2}\tilde{V}\xi$$

- Choose a truncated set of modes to export in tabular form
- Mie-Grüneisen example has two significant modes:

$$T = \bar{T} + \xi_1 T_1 + \xi_2 T_2$$

- Random variables ξ are uncorrelated, with zero mean and unit standard deviation, but not necessarily independent
- Gaussian approximation for ξ exact if T is Gaussian process; accurate to second order otherwise

Building the first aluminum multiphase UQ enabled table

- The major issues with building the UQ enabled full multiphase EOS have to do with building AUTOMATIC robust EOS and table generation algorithms.
- Parameter sets can be chosen that give errors in phase boundary topology (kinks)
- Unstable EOS regions can occur if there is not enough data to constrain the automatic inference process.
- Any constraints that used to be dealt with in various informal ways by the EOS expert MUST now be automatically detected and removed from the Bayesian inference process as representing an impossible parameter state.
- The inference must include sufficient data and more extensive prior constraints.
- The end result will in fact be a much more robust and automated EOS table building system with the added benefit of UQ enablement.

$$T = \bar{T} + \xi_1 T_1 + \xi_2 T_2 + \xi_3 T_3 + \cdots$$

Multiphase Tabular Generation and Representation: Initial AL UQ enabled table

$$T = \bar{T} + \xi_1 T_1 + \xi_2 T_2 + \xi_3 T_3 + \cdots$$

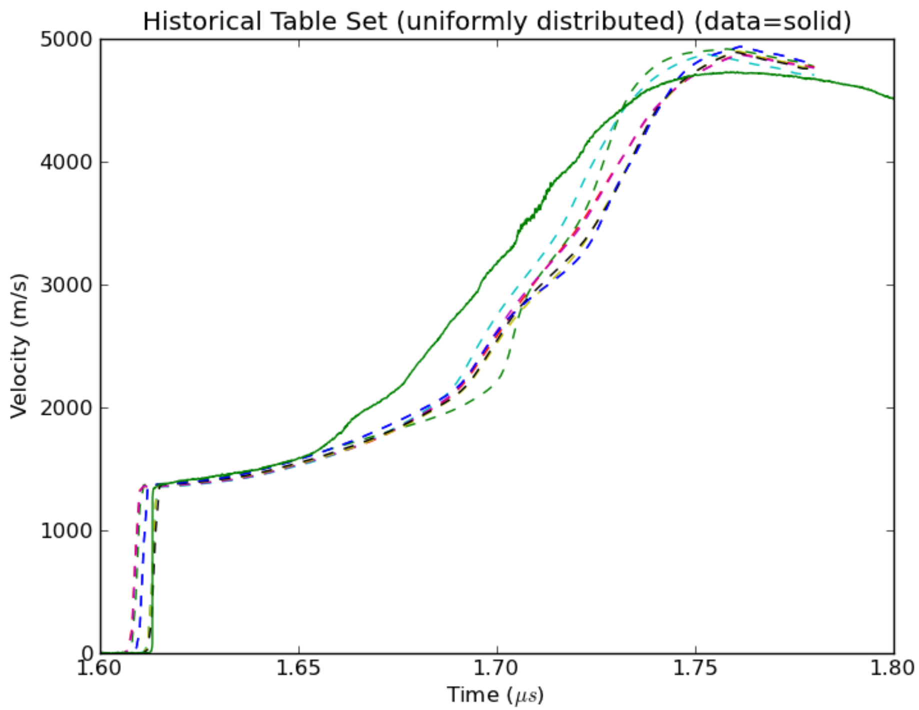
- First wide range AL EOS tables with 6 phase regions in the density-energy table and 5 phase regions in the density-temperature table have been generated. (Triple point collapses)
- With the current data set there are 36 free parameters.
- 13 parameters were fixed due to insufficient information.
- The MCMC inference samples 23 parameters
- We took 400 samples from the chain. Currently we are seeing only 1 significant mode at 1-e6 cutoff in PCA analysis. However, we had to throw out some samples due to problems with the mesh smoothing in the melt region.
- Accuracy of the tables is set at a relative tolerance of .5.
- Solver currently scales as N^2 so this limits practical number of samples.

Enabling Macroscale UQ Analyses

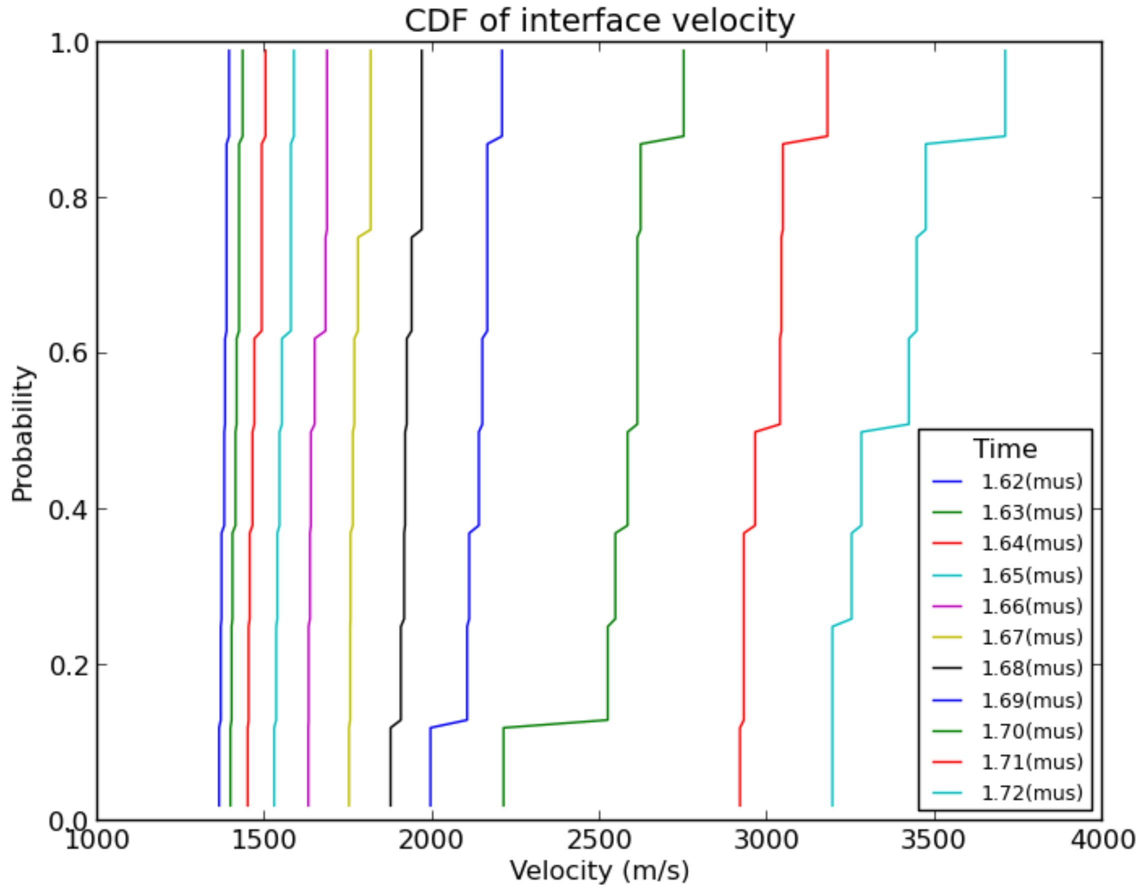
Software Package	Output
EOS model library and data	Proposal Model (XML input deck)
Bayesian Inference using Markov Chain Monte Carlo	Extensive Sampling of the posterior distribution function (PDF)
Multinormal PDF fit + Dakota	Reduced PDF samples
EOS Table Building	Topologically equivalent tables for each sample
PCA Analysis	Mean EOS table + most significant perturbations
Hydrocode + Dakota	Cumulative Distribution Function (CDF) for quantities of interest

Al Shock Ramp – Naïve approach

- Z data and initial ALEGRA modeling setup courtesy Matt Martin.
- Demonstrate uncertainty analysis with multiple AL EOS tables.
- Assume 8 tables each occur with .125 probability
- Dakota sampling and response functions are set up in one control input file with no external user scripting.



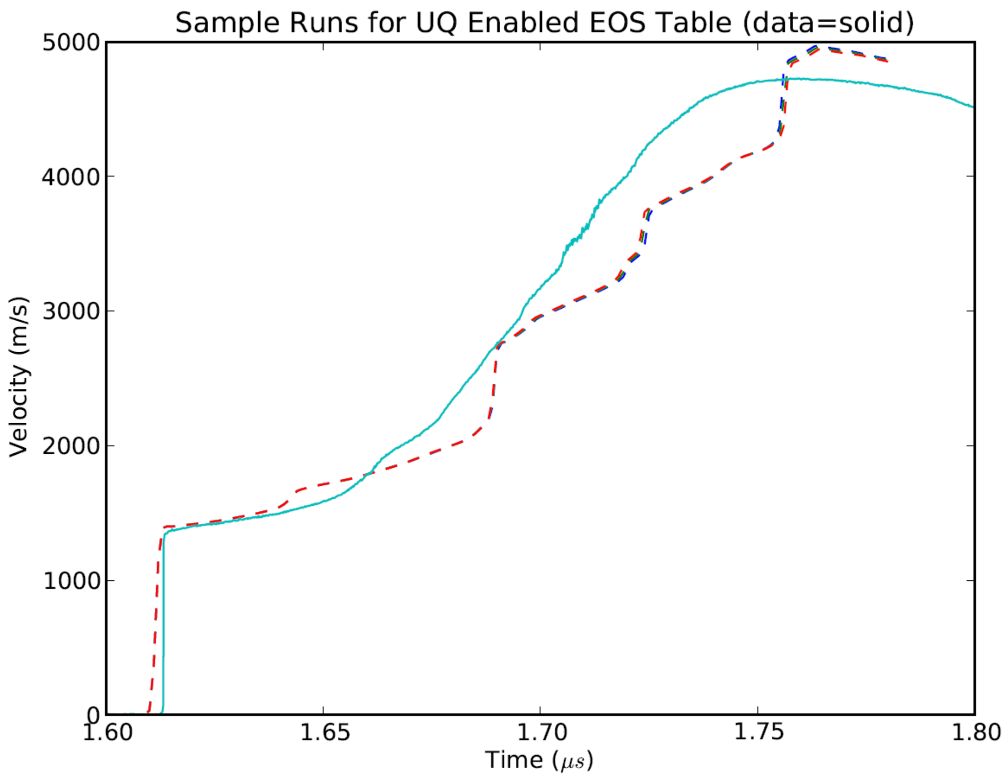
CDF profiles using a set of 8 classical tables and a hypothesized uniform discrete distribution.



Al Shock Ramp Experiment – UQ AL EOS

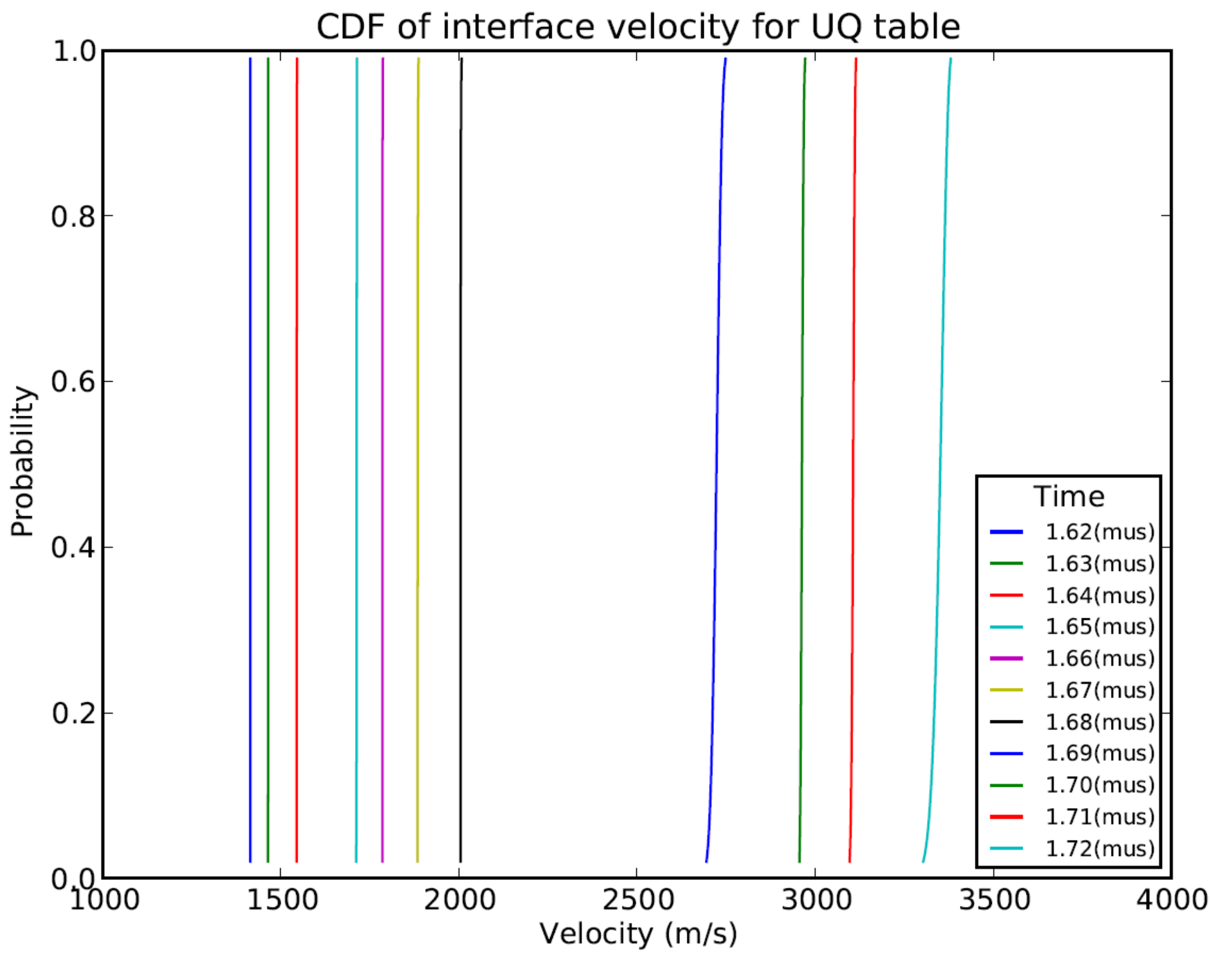
- UQ enabled AL EOS from 202 sample tables with .5 relative tolerance results in one primary mode.
- PCE sampling of first mode with N(0,1) distribution

$$T = \bar{T} + \xi_1 T_1 + \xi_2 T_2 + \xi_3 T_3 + \cdots$$



Sandia National Laboratories

CDF profiles from PCE expansion of velocities using UQ enabled AL EOS



Sandia National Laboratories

Summary

- We have developed a practical approach to represent uncertainty in multiphase EOS for production delivery to design analysts working at the continuum level.
- The Bayesian inference framework seems to be a good framework for thinking about both the parameter estimate problem and for developing the distributions on input parameters.
- A robust high dimensional inference methodology is essential.
- Robust control of prior definitions which include physical constraints as well as sufficient data to constrain the parameter set is essential.
- The new UTRI tabular format gives more detailed phase boundary definitions, direct control over tabular accuracy and enables UQ representation.
- Delivery for easy usage in continuum codes is essential in order to provide immediate impact at the continuum modeling scale.
- Good stochastic compression has been observed so far for AL EOS models.

$$T = \bar{T} + \xi_1 T_1 + \xi_2 T_2 + \xi_3 T_3 + \cdots$$

Sandia National Laboratories