SAND2016-6461PE

# Accounting for Data Uncertainty in Making and Using an EOS

JOWOG32Mat EOS Workshop
July 13, 2016

John H. Carpenter, Allen C. Robinson,
Bert J. Debusschere, Ann E. Mattsson

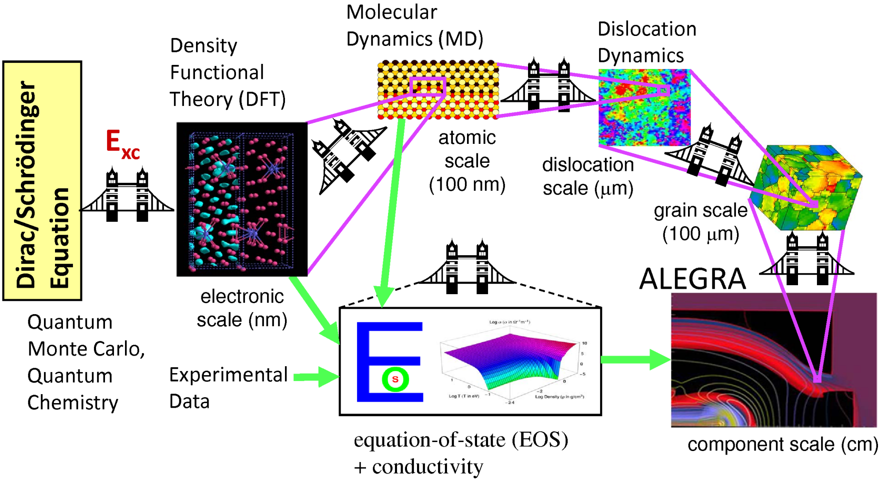
Center for Computing Research

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.

National Nuclear Security Administration

Sandia National Laboratories

# EOS Development Paradigm




- Provide quantitative error estimates to continuum analysts based upon fundamental measurements and calculations of the EOS.
- Preserve model providence throughout the process.
- Usable system for generation and use of the EOS.

# Tabular EOS UQ System

| Software Package | Output |
|---|---|
| EOS model library and data | Proposal model (XML input deck) |
| Bayesian inference using Markov Chain Monte Carlo | Extensive sampling of Posterior Distribution Function (PDF) |
| EOS Table building | Topologically equivalent tables for each sample |
| PCA analysis | Mean EOS table + most significant table perturbations |
| Hydrocode + Dakota | Cumulative Distribution Function for quantities of interest |

- First half of system utilizes analytic EOS models.
- Last half of system utilizes tabulated EOS models.
- Uncertainty information transferred from parametric to thermodynamic space.
- Focus herein will be on the first steps involving data

Sandia National Laboratories

# What Does “EOS Model” Mean?

## XML input deck is the (meta-)EOS Model:

| | |
|---|---|
| `<EOSModel>` | Traditional EOS model definition |
| `<EOSData>` | EOS data and uncertainties used for model calibration |
| `<Inference>` | Controls for the inference |
| `<Tabulation>` | Controls for the tabulation |

## EOS table and interpolation scheme is the real “EOS model”

- Codes actually query it for thermodynamic closure states
- Example: pick on Kerley’s 3700 and Sandia’s codes
  - ALEGRA simulation with 3700 and backup linear interpolation
  - ALEGRA simulation with 3700 and sound speed modifications
  - CTH simulation with 3700 and bad state clipping
  - Saying “Aluminum 3700” describes none of these accurately
  - They are not even the same as Kerley’s model used to build 3700

## Can the XML input really be the “EOS model”?

- Tabulation must be representative of original models
- Consistency between EOS build tools and hydrocode interpolation
- System must be automated – no by-hand modifications
- Provide no incentives for fiddling by the hydrocode/analysts

# Recording the Art of EOS Building

## Expert modeler decisions (often visually based):

- Appropriate models to use
- Stability and physicality requirements
- Relative weighting of data sets
- How well models should agree with data

## Bayes' rule allows inferring parameters' posterior distribution function (PDF) using data and prior knowledge:

$$\overbrace{p(\lambda|D)}^{\text{posterior}} = \frac{\overbrace{p(D|\lambda)}^{\text{likelihood}}\,\overbrace{p(\lambda)}^{\text{prior}}}{\underbrace{p(D)}_{\text{normalization}}}$$

## The expert's art must be encoded in the XML input for automation of PDF sampling and table building.

- All invalid parameter sets must be rejected in the inference
- Likelihood contains weighting of data
- Prior contains conditions on physicality (rejection criteria)
- Expert still must guide the system to a good starting point, the Maximum A Posteriori (MAP) value

# Bayesian Inference

$$\overbrace{p(\lambda|D)}^{\text{posterior}} = \frac{\overbrace{p(D|\lambda)}^{\text{likelihood}}\overbrace{p(\lambda)}^{\text{prior}}}{\underbrace{p(D)}_{\text{normalization}}}$$

Bayesian "objective" is the posterior distribution function (PDF):

- Attempt to find the most likely parameters and their distribution, allowing UQ practices to be employed
- Often sampled with Monte Carlo methods but can also be optimized directly with gradient methods
- Typical to work with logarithm of PDF and ignore normalization

# Constructing Bayesian Likelihood

Calculate likelihood:

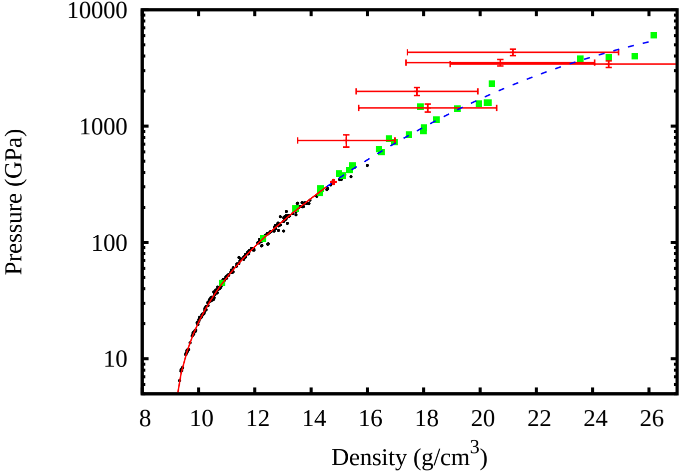
$$\log(p(D|\lambda)) = -\frac{1}{2}\left[\frac{D-\bar{D}}{\sigma_D}\right]^2 - \frac{1}{2}\log\sigma_D^2$$

- Simple Gaussian noise models currently supported
- Noise model hyperparameter $\sigma_D$ often can be fixed
- $\sigma_D$ may be specified either as specific values or a relative weight
- Gaussian statistics with fixed hyperparamters gives least-squares
- Every data source is cast into this framework, from cold curves to shock data to phase boundaries
- Using surrogate models for data can make sense when many and/or scattered point are available

# Surrogate Data Example

Principal Hugoniot data for copper:

- Surrogate model is linear or quadratic $u_s$-$u_p$ form
- Points sampled from surrogate for inclusion in likelihood
- $\sigma_D$ is chosen based upon spread of data from surrogate



# Constructing Bayesian Prior

Calculate prior:

$$\log(p(\lambda)) = -\frac{1}{2}\left[\frac{\lambda-\bar{\lambda}}{\sigma_\lambda \bar{\lambda}}\right]^2 + \texttt{conditions}(\lambda)$$

- Simple Gaussian priors and bound constraints currently supported for parameters
- Various uses include constraining parameters to certain ranges or eliminating unphysical models

Conditions:

- Act as rejection criteria (very small probability)
- Encode physicality requirements on the model
- Requirements ranked in prior contribution to aid in sampling
- Can still make sampling and optimization difficult
- Moving from discrete to smooth engagement of conditions would likely improve these difficulties

# Physicality Conditions

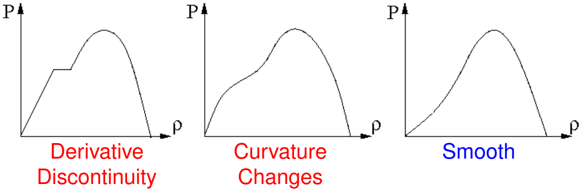
## Thermal and mechanical stability:

- Heat capacity and bulk modulus must be positive
- Applied to pure phase regions (i.e. not transitions)
- Particularly important for regions without calibration data

## Smoothness of phase boundary lines:

- Boundaries must not have discontinuities
  - They indicate problems with model parameters or solvers
  - Prevents later problems with curve approximation and tabulation
- Boundaries should not have multiple curvature changes
  - One sign change allowed along boundaries
  - Physical for vapor dome, other lines may still be questionable
  - With few exceptions more changes are unphysical
- Applies to vaporization, melt, polymorphic transitions
- Derivatives sampled along phase boundaries

## Vapor dome example:



# Aluminum Example Case

Test multi-phase aluminum model:

- Semi-empirical solid-liquid-gas model
  - Cold curve uses polynomial expansion form
  - FCC solid phase uses the Debye model
  - Fluid phase uses Bushman-Lomonosov-Fortov model
- 37 parameters in total
- Range of interest to 150 kK and 20 g/cm$^3$

Multiple sets of data used for calibration:

- Isobaric enthalpy and density for solid and liquid
- Shock data for solid and liquid
- Isothermal compression data for solid
- QMD calculations of critical point plus melt and vaporization data

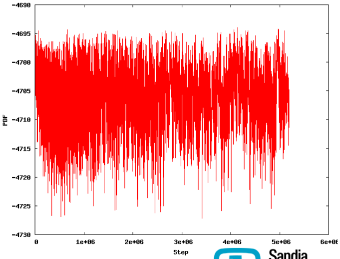
# Optimizing the Posterior

## Optimization and Sampling:

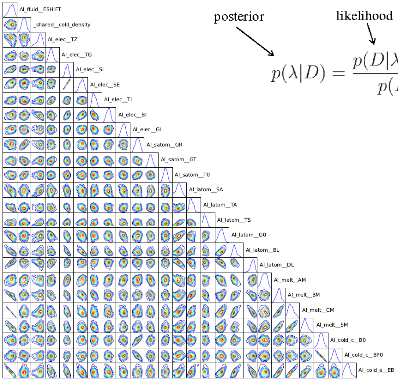
- PDF is optimized as in normal EOS development
- Hand tuning of parameters or automated methods used
- Methods: gradient based, genetic algorithms, simulated annealing
- All suffer curse of dimensionality in finding global optimum

## Maximum A Posteriori (MAP) parameters:

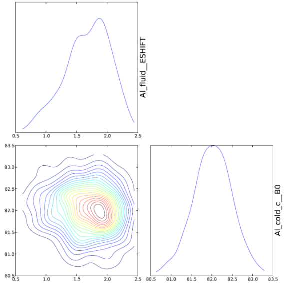
- Normal EOS development gives the expert opinion “best” model.
- Sampling the PDF via Markov Chain Monte Carlo (MCMC) verifies this opinion is represented (the “art” has been recorded) in the statement of the likelihood and prior probabilities.
- MCMC burn-in with significant change in the PDF indicates a “better” model.
- The verified best model is the starting point for the remaining UQ process.



# Al EOS Model Inference




$$\underbrace{p(\lambda|D)}_{\text{posterior}} = \frac{\overbrace{p(D|\lambda)}^{\text{likelihood}}\,\overbrace{p(\lambda)}^{\text{prior}}}{\underbrace{p(D)}_{\text{normalization}}}$$




Al EOS 25 parameter inference

A marginal distribution

- Use adaptive Markov Chain Monte Carlo scheme to reduce number of steps
- Start chain from optimized MAP parameters
- PDF evaluations may be parallelized to enable long chains ($\sim$4.5M steps for this EOS, one serial evaluation is approximately 2 sec.)
- Each posterior evaluation is roughly equivalent to generating an entire EOS table and having an expert check it for correct behavior.

Sandia National Laboratories

# Summary

Dealing with data uncertainty in creating EOS models:

- The data needs to be treated on the same footing as the rest of the modeling process
- Expert opinion on model agreement with data must be recorded
- Automated processes must reproduce the expert opinion
- Physical constraints on models are an implicit form of data in this system
- Sampling of the PDF provides the foundation for propagating the data uncertainty to inform uncertainty in engineering outputs.

Sandia National Laboratories