SAND2021-9041C

Bayesian calibration of interatomic potential models for binary alloys

Arun Hegde
(ahegde@sandia.gov)
Cosmin Safta
Habib Najm

Elan Weiss
Wolfgang Windl

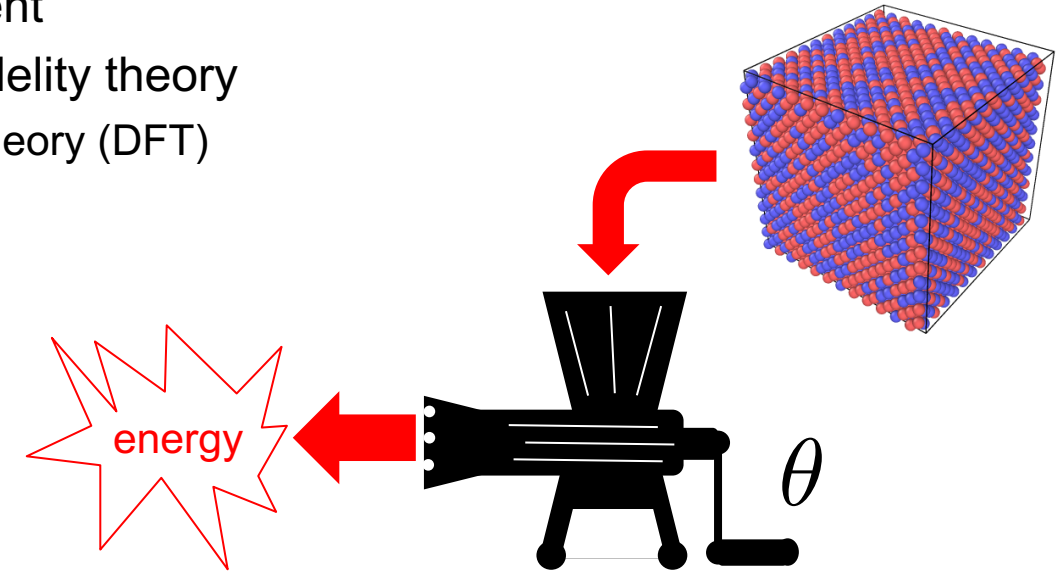
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Overview

- Interatomic potential models and UQ
- Problem setup (RAMPAGE potentials for binary alloys)
- Bayesian calibration for interatomic potential fitting
- Case study – Au-Cu binary alloy systems

Interatomic potentials and UQ

- Interatomic potentials
  - function that takes as input the positions of atoms and returns the energy of the system
  - contains unknown parameters that must be determined empirically
    - comparison with experiment
    - comparison with higher-fidelity theory
      - e.g. density functional theory (DFT)



Interatomic potentials and UQ

- Interatomic potentials
  - function that takes as input the positions of atoms and returns the energy of the system
  - contains unknown parameters that must be determined empirically
    - comparison with experiment
    - comparison with higher-fidelity theory
      - e.g. density functional theory (DFT)
- Reliable simulation with interatomic potentials requires quantified uncertainties*
  - for model validation and comparison
  - for predictivity
  - for decision-making

*S.L. Frederiksen, K.W. Jacobsen, K.S. Brown, J.P. Sethna, Phys. Rev. Lett. 93, 165501 (2004)

RAMPAGE potentials for binary alloy design

Rapid **A**lloy **M**ethod for **P**roducing **A**ccurate **G**eneral **E**mpirical potentials*

Embedded Atom Model (Finnis-Sinclair type) for systems with <u>two element types: **A** and **B**</u>

- atom i of type $\alpha \in \{A, B\}$
- atom j of type $\beta \in \{A, B\}$

energy at atom *i*

$$E_i = \frac{1}{2} \sum_{j \neq i} V_{\alpha\beta}(r_{ij}) + F_\alpha \left(\sum_{j \neq i} \rho_{\alpha\beta}(r_{ij}) \right)$$

component functions could each contribute to θ

*L. Ward, A. Agrawal, K.M. Flores, and W. Windl. Rapid production of accurate embedded-atom method potentials for metal alloys. (2012). arXiv:cond-mat.mtrl-sci/1209.0619

RAMPAGE potentials for binary alloy design

Rapid **A**lloy **M**ethod for **P**roducing **A**ccurate **G**eneral **E**mpirical potentials*

Embedded Atom Model (Finnis-Sinclair type) for systems with <u>two element types: **A** and **B**</u>

$$F_A \quad \rho_{AA} \quad V_{AA}$$

$$F_B \quad \rho_{BB} \quad V_{BB}$$

extracted from existing literature

already fitted, presumably valid

$$\rho_{BA} \quad \rho_{AB}$$

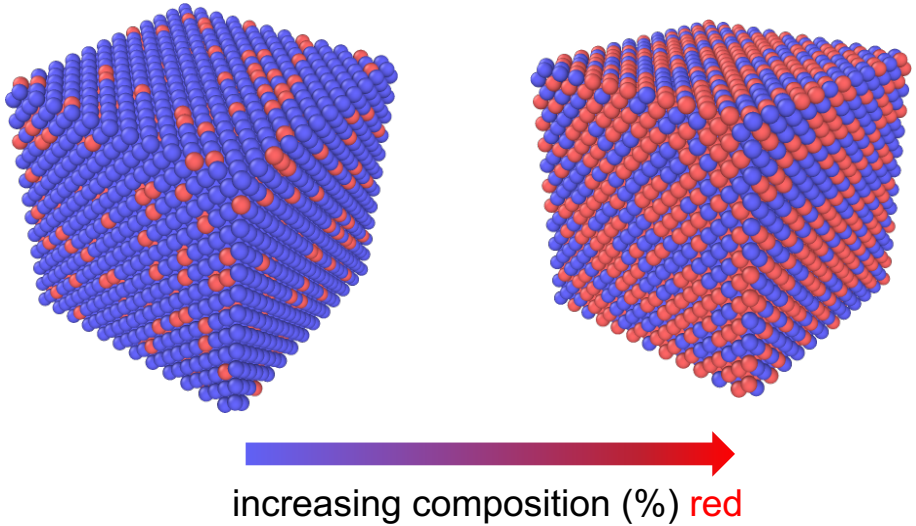
$$V_{BA} = V_{AB}$$

modeling and fitting effort

*L. Ward, A. Agrawal, K.M. Flores, and W. Windl. Rapid production of accurate embedded-atom method potentials for metal alloys. (2012). arXiv:cond-mat.mtrl-sci/1209.0619

RAMPAGE potentials for binary alloy design

Rapid **A**lloy **M**ethod for **P**roducing **A**ccurate **G**eneral **E**mpirical potentials*



- Generate DFT data for a variety of structures with compositions ranging from 100% A, 0%B to 0%A, 100% B.
- Use data to fit the cross-term components of the interatomic potential

*L. Ward, A. Agrawal, K.M. Flores, and W. Windl. Rapid production of accurate embedded-atom method potentials for metal alloys. (2012). arXiv:cond-mat.mtrl-sci/1209.0619

A 5-parameter model

Interatomic potential model

$$V_{AB}(r) = D\left(e^{-2\alpha(r-r_{eq})} - 2e^{-\alpha(r-r_{eq})}\right)$$

Morse pair potential

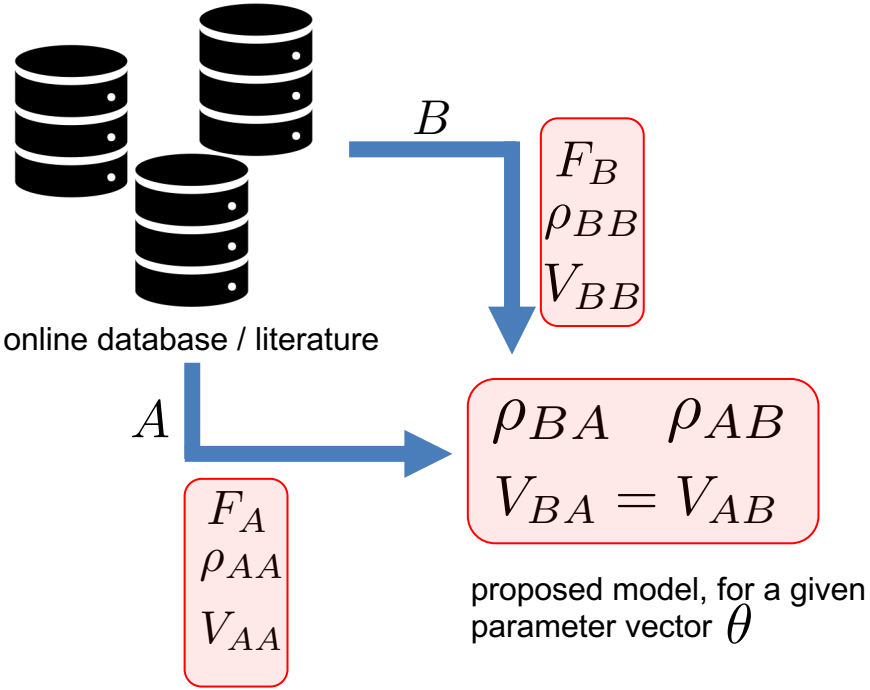
$$\rho_{BA}(r) = r^6\left(e^{-S_A r} + 2^9 e^{-2S_A r}\right)$$

$$\rho_{AB}(r) = r^6\left(e^{-S_B r} + 2^9 e^{-2S_B r}\right)$$

Voter electron densities

- **5** uncertain parameters: $\theta = [r_{eq}, D, \alpha, S_A, S_B]$
- **102** QOIs total
  - 17 compositions ranging from 3% A to 97% A
  - For each composition: **lattice parameter, mixing enthalpy, C11, C12, C44, bulk modulus**
- Higher-fidelity DFT data generated for each QOI
  - used for fitting the uncertain parameters

Workflow for computing QOIs



Step 1: download fitted potentials for the pure system, combine with proposed model

Workflow for computing QOIs

online database / literature

B

F_B
ρ_{BB}
V_{BB}

A

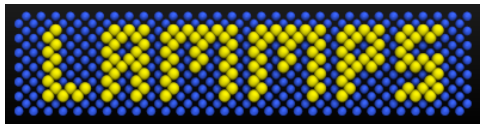
F_A
ρ_{AA}
V_{AA}

$\rho_{BA} \quad \rho_{AB}$
$V_{BA} = V_{AB}$

proposed model, for a given parameter vector θ

Step 2: compute quantities of interest using LAMMPS

other inputs:
- structures
- composition ranges
- optimization tolerances
- etc.



molecular dynamics simulation software*

computed properties & post processing

*image from https://www.lammps.org

Workflow for computing QOIs

online database / literature

B

F_B
ρ_{BB}
V_{BB}

A

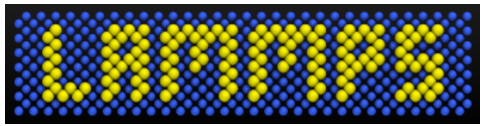
F_A
ρ_{AA}
V_{AA}

$$\rho_{BA} \quad \rho_{AB}$$
$$V_{BA} = V_{AB}$$

proposed model, for a given parameter vector θ

Step 3: compare to reference DFT data

other inputs:
- structures
- composition ranges
- optimization tolerances
- etc.



molecular dynamics simulation software*

computed properties & post processing

inference

*image from https://www.lammps.org

Bayesian calibration

finite composition space $x \in \mathcal{X}$

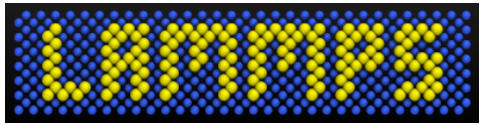
physical properties $i \in \{\text{lat}, \text{mix}, \text{C}_{11}, \text{C}_{12}, \text{C}_{44}, \text{bulk}\}$

notation used to index different **QOIs (102 total)**

$$y_i(x) = f_i(x;\theta) + \epsilon_i(x)$$

DFT data — simulation

simulation

$\theta \in \mathbb{R}^5$ → Interatomic potential model →  → $f_i(x;\theta)$

molecular dynamics simulation software*

*image from https://www.lammps.org

Bayesian calibration

finite composition space $x \in \mathcal{X}$

physical properties $i \in \{\text{lat}, \text{mix}, \text{C}_{11}, \text{C}_{12}, \text{C}_{44}, \text{bulk}\}$

notation used to index different **QOIs (102 total)**

$$y_i(x) = f_i(x;\theta) + \epsilon_i(x)$$

DFT data — simulation — discrepancy

$$\epsilon_i(x) \sim \mathcal{N}\left(0, \sigma^2 f_i(x;\theta)^2\right) \qquad i \in \{\text{lat}, \text{C}_{11}, \text{C}_{12}, \text{C}_{44}, \text{bulk}\}$$

$$\epsilon_i(x) \sim \mathcal{N}\left(0, \sigma^2 f_i(x;\theta)^2 + \tau^2\right) \qquad i \in \{\text{mix}\}$$

$$\lambda = [\theta, \log\sigma, \log\tau]$$

Bayesian calibration

finite composition space $x \in \mathcal{X}$

physical properties $i \in \{\text{lat}, \text{mix}, \text{C}_{11}, \text{C}_{12}, \text{C}_{44}, \text{bulk}\}$

notation used to index different **QOIs (102 total)**

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

$\epsilon_i(x) \sim$

$\epsilon_i(x) \sim$

$, \text{C}_{12}, \text{C}_{44}, \text{bulk}\}$

$[\theta, \log\sigma, \log\tau]$

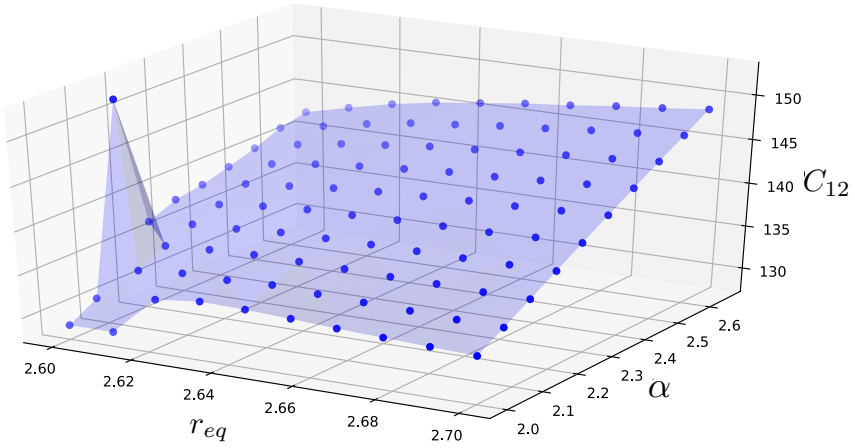
Inference strategy

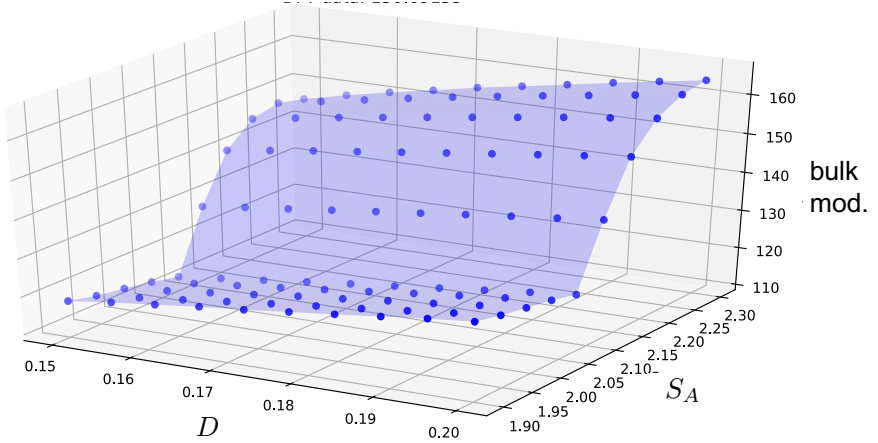
- Standard strategy
  1. specify a reasonable prior, e.g. uniform over a plausible range
  2. perform MCMC with the full model (i.e. LAMMPS)
  3. analyze posterior samples
- Challenges
  1. not always clear how to specify prior parameter ranges
  2. portions of the parameter space lead to unphysical results
     - unconverged minimizations, flat QOI response, kinks, etc. (next slide)
  3. simulation runtime depends on the input parameter choices
     - single evaluation ~15 minutes – 1hrs+ on a single cpu

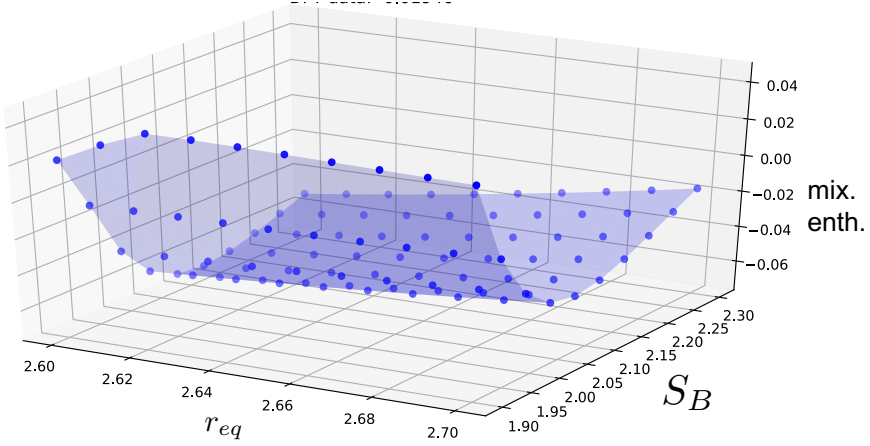
Inference strategy

- Standard strategy
  1. specify a reasonable prior, e.g. uniform over a plausible range
  2. perform MCMC with the full model (i.e. LAMMPS)
  3. analyze posterior samples
- Challenges
  1. not always clear how to specify prior parameter ranges
  2. portions of the parameter space lead to unphysical results
     - unconverged minimizations, flat QOI response, kinks, etc. (next slide)
  3. simulation runtime depends on the input parameter choices
     - single evaluation ~15 minutes – 1hrs+ on a single cpu

Illustrations of challenge #2







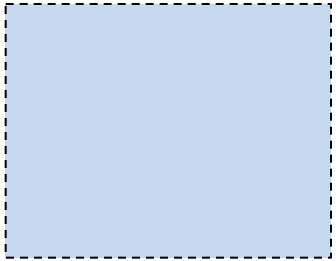
2D slices of the simulation response

Inference strategy

Strategy

1. Find a good initial box in the parameter space.
2. Initialize a set of training/test samples.
3. Fit Gaussian process surrogates.
4. Perform MCMC (with surrogates, uniform prior).
5. Adapt box based on posterior samples.
6. Append posterior samples to training set.
7. Repeat steps **3-6** until:
   - surrogate error on training/test samples is small
   - posterior samples strictly contained

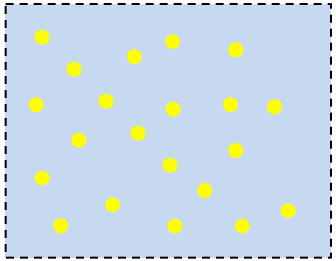
(found through optimization or "expert opinion")



Inference strategy

Strategy

1. Find a good initial box in the parameter space.
2. Initialize a set of training/test samples.
3. Fit Gaussian process surrogates.
4. Perform MCMC (with surrogates, uniform prior).
5. Adapt box based on posterior samples.
6. Append posterior samples to training set.
7. Repeat steps **3-6** until:
   - surrogate error on training/test samples is small
   - posterior samples strictly contained

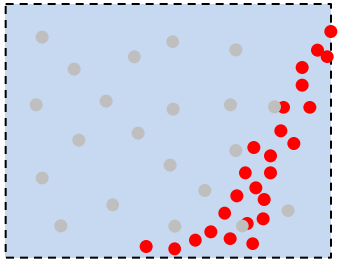


LAMMPS evaluations performed “offline” in a highly parallelized HPC setting

Inference strategy

Strategy

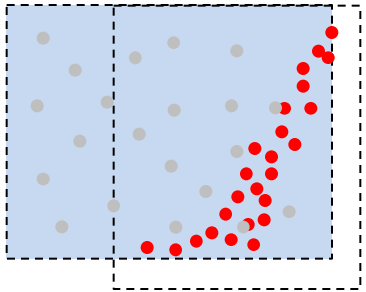
1. Find a good initial box in the parameter space.
2. Initialize a set of training/test samples.
3. Fit Gaussian process surrogates.
4. Perform MCMC (with surrogates, uniform prior).
5. Adapt box based on posterior samples.
6. Append posterior samples to training set.
7. Repeat steps **3-6** until:
   - surrogate error on training/test samples is small
   - posterior samples strictly contained



Inference strategy

Strategy

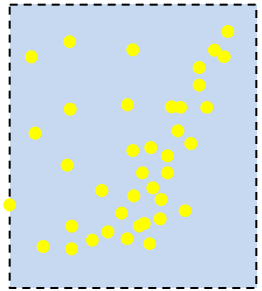
1. Find a good initial box in the parameter space.
2. Initialize a set of training/test samples.
3. Fit Gaussian process surrogates.
4. Perform MCMC (with surrogates, uniform prior).
5. Adapt box based on posterior samples.
6. Append posterior samples to training set.
7. Repeat steps **3-6** until:
   - surrogate error on training/test samples is small
   - posterior samples strictly contained



Inference strategy

Strategy

1. Find a good initial box in the parameter space.
2. Initialize a set of training/test samples.
3. Fit Gaussian process surrogates.
4. Perform MCMC (with surrogates, uniform prior).
5. Adapt box based on posterior samples.
6. Append posterior samples to training set.
7. Repeat steps **3-6** until:
   - surrogate error on training/test samples is small
   - posterior samples strictly contained

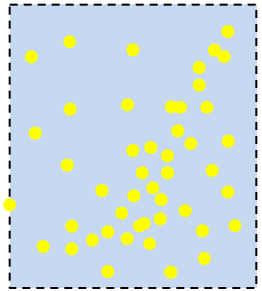


LAMMPS evaluations of new samples performed "offline" in a highly parallelized HPC setting

Inference strategy

Strategy

1. Find a good initial box in the parameter space.
2. Initialize a set of training/test samples.
3. Fit Gaussian process surrogates.
4. Perform MCMC (with surrogates, uniform prior).
5. Adapt box based on posterior samples.
6. Append posterior samples to training set.
7. Repeat steps **3-6** until:
   - surrogate error on training/test samples is small
   - posterior samples strictly contained



(may also add new samples)

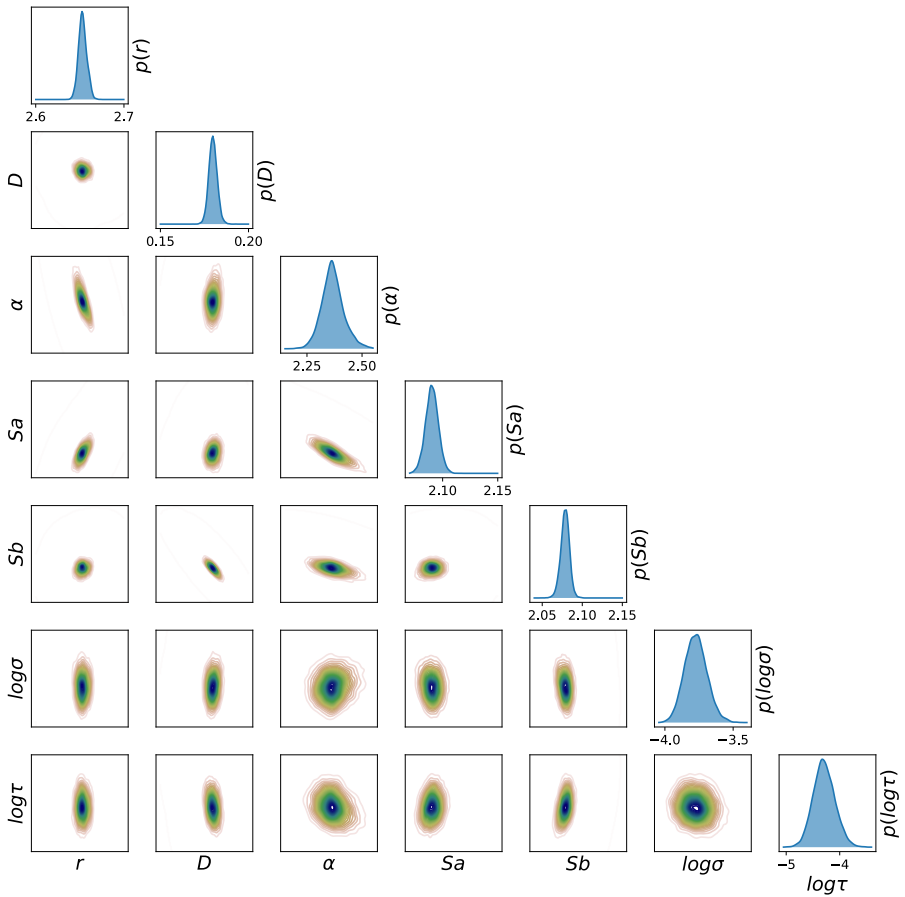
LAMMPS evaluations of new samples performed “offline” in a highly parallelized HPC setting

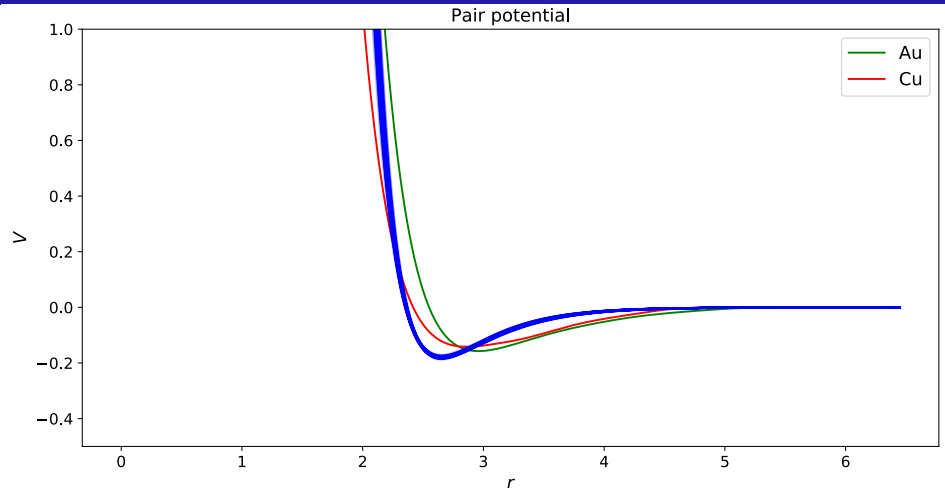
Case study: Au-Cu system

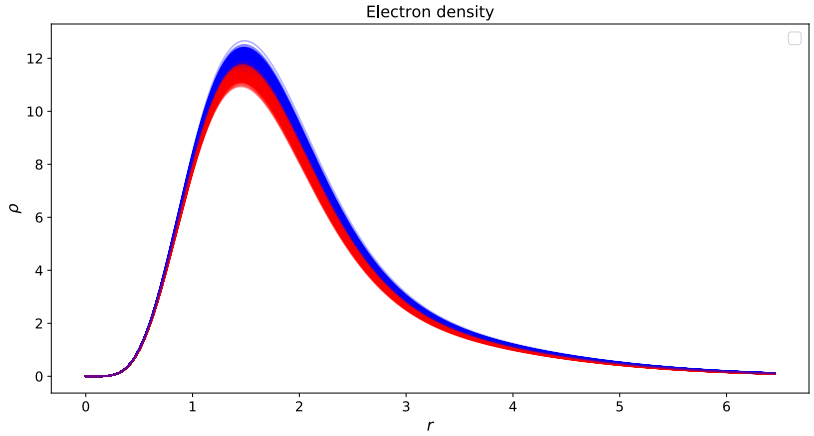
- Interatomic potential model
  - RAMPAGE potential, single element terms for Au* and Cu*
  - 5-parameter cross-term model
- 102 QOIs in total
  - 17 compositions ranging from 3% Au to 97% Au
  - for each composition: lattice parameter, mixing enthalpy, C11, C12, C44, bulk modulus
- Higher-fidelity DFT data generated for each QOI
  - used for fitting the uncertain parameters
- MCMC algorithm: Adaptive Metropolis

*X. W. Zhou, R. A. Johnson, H. N. G. Wadley, Phys. Rev. B, 69, 144113 (2004)

Results: posterior marginals

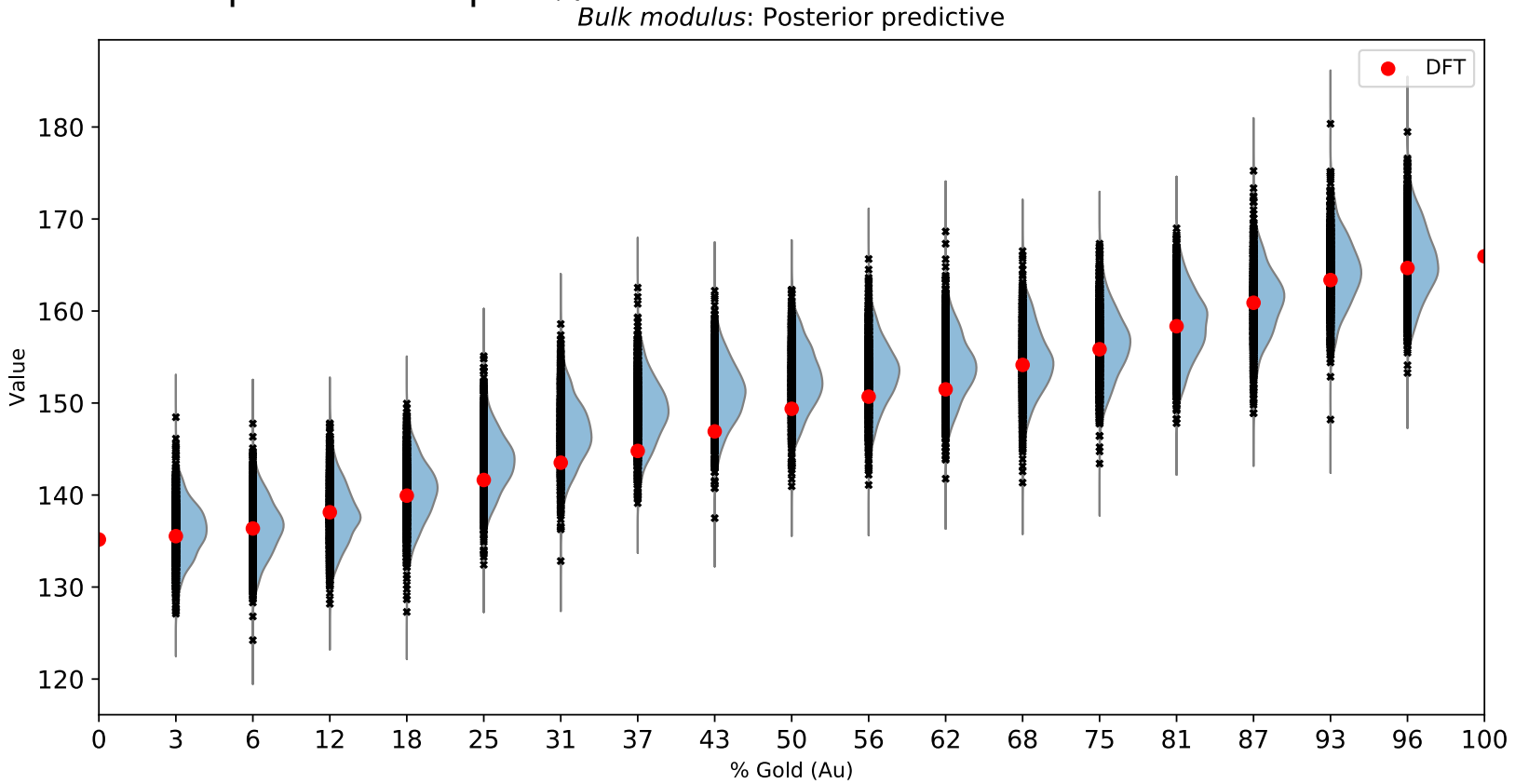






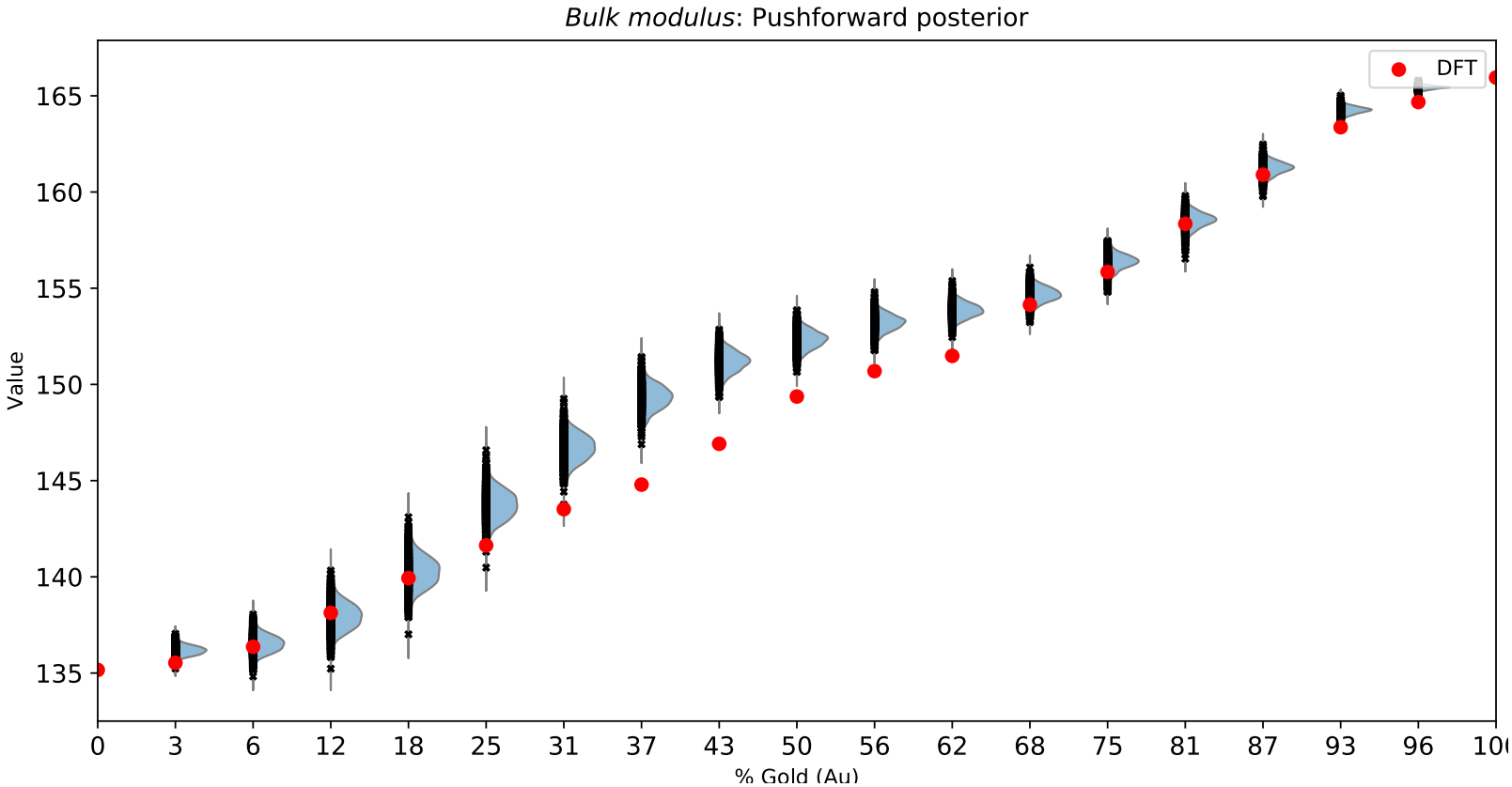
Results: snapshot of posterior predictive

QOI evaluates of posterior samples λ



Results: snapshot of pushforward posteriors

QOI evaluates of marginal posterior samples θ



A 7-parameter model

What happens if we change the model? +2 parameters

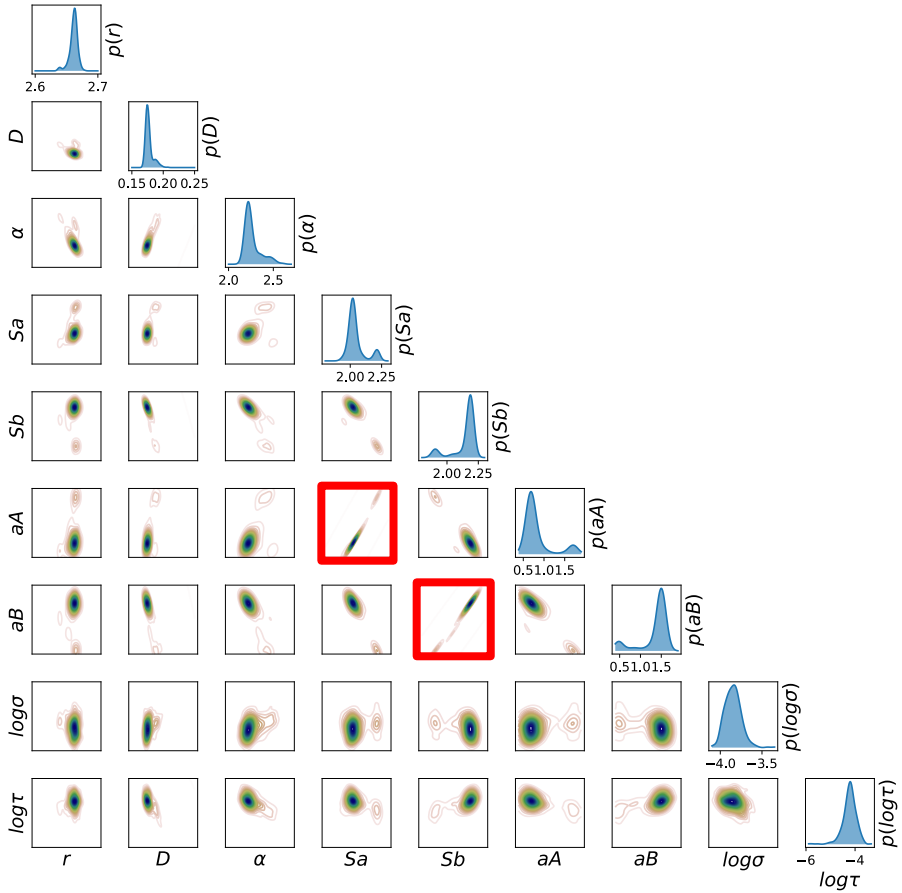
Interatomic potential model

$$V_{AB}(r) = D\left(e^{-2\alpha(r-r_{eq})} - 2e^{-\alpha(r-r_{eq})}\right)$$

$$\rho_{BA}(r) = a_A r^6 \left(e^{-S_A r} + 2^9 e^{-2S_A r}\right)$$

$$\rho_{AB}(r) = a_B r^6 \left(e^{-S_B r} + 2^9 e^{-2S_B r}\right)$$

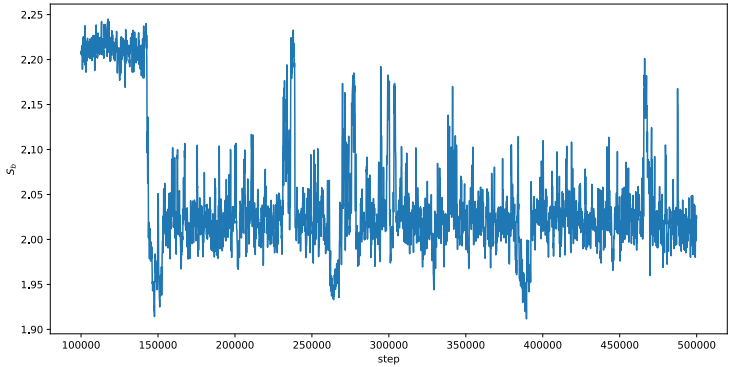
A 7-parameter model



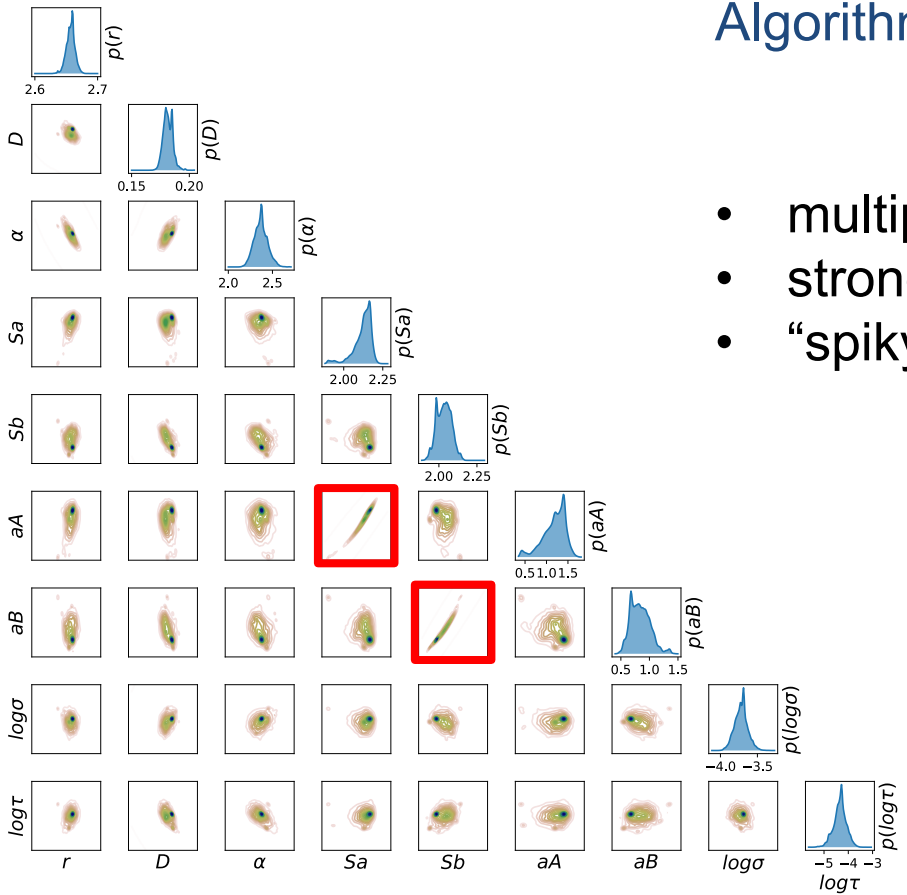
Algorithm: Adaptive Metropolis

- multiple modes
- strong correlations

Diagnosing convergence in such settings can be challenging.



A 7-parameter model



Algorithm: Transitional MCMC

sequential monte carlo algorithm that evolves an ensemble of particles

- multiple modes
- strong correlations
- “spiky” posteriors imply particle degeneracy

Practical considerations:

Success of “out of the box” algorithms depends on the tuning parameters.

Observed correlations can inform modeling decisions and adjustments.

Summary

- Investigated RAMPAGE potential model for fitting Au-Cu systems
- Bayesian framework for fitting interatomic potentials
  - sequential strategy based on adaptively exploring the parameter space with locally fit surrogate models.
- Success story: 5-parameter model
- Ongoing story: 7-parameter model
    - strong correlations among parameters, multimodal posteriors
    - sampling strategies tested: adaptive metropolis, transitional MCMC, parallel tempering
- Key topics for future work: model selection, prediction of other QOIs

Acknowledgements

Work supported by the U.S. Department of Energy, Office of Science

- Advanced Scientific Computing Research (ASCR)
- Scientific Discovery through Advanced Computing (SciDAC) program through the FASTMath Institute
- Basic Energy Sciences (BES)
- National Energy Research Scientific Computing Center (NERSC)

National Energy Research Scientific Computing Center

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the presentation do not necessarily represent the views of the U.S. Department of Energy or the United States Government.