Uncertainty Quantification with Surrogate Models in Alloy Modeling

Jesper Kristensen,
(joint work with Prof. N. Zabaras*)

Applied and Engineering Physics &
Materials Process Design & Control Laboratory
Cornell University
271 Clark Hall, Ithaca, NY 14853-3501
and

*Warwick Centre for Predictive Modelling
University of Warwick, Coventry, CV4 7AL, UK

Warwick Centre for Predictive Modelling Seminar Series, University of Warwick, Coventry, UK, January 8, 2015
Finding the Best Materials

Search for target material/property

- Single candidate property (one column to the left)
  ➢ Hours

- Million+ candidates
  ➢ Infeasible!

Finding the Best Materials

Search for target material/property

- Single candidate property (one column to the left)
  - Hours

- Million+ candidates
  - Infeasible!

- Need for surrogate models
  - Rapid configuration space exploration
  - Allows design of materials

Using Surrogates in Alloy Modeling
The Cluster Expansion

- Alloy surrogate model
  - The cluster expansion

- Cluster with \( n \) points:
  - \( n \)-pt cluster

- Expansion coefficients \( J_k \): ECI
  - Effective cluster interactions

- Clusters similar under space group symmetries
  - Same ECI
  - "High symmetry: few unknowns"

Can Information Theory Improve Thermodynamic Alloy Modeling with Surrogates?

**COMMON METHODOLOGY**

\[ \mathcal{D} = \{ \sigma^{(i)}, E^{(i)} \}_{i=1}^{N} \]

\[ \mathcal{L}(\gamma) = \| \mathbf{E}^{(1:N)} - \Phi \gamma \|^2 \]

- Expensive data set
- Much-used approach: Least squares

**Can we do better if the objective is to obtain the ground states?**
Fitting the Boltzmann Distribution

\[
Z = \sum_{\sigma} \exp(-\beta E(\sigma))
\]

Partition function

\(\sigma\): Configurational states of the system

\[
p(\sigma) = \frac{\exp(-\beta E(\sigma))}{Z}
\]

Ab initio energy: expensive

Replace Boltzmann with surrogate distribution

\[
p(\sigma | \gamma) = \frac{\exp(-\beta E(\sigma; \gamma))}{Z(\gamma)}
\]

\(E(\sigma; \gamma) = \gamma^T \phi(\sigma)\)

We aim to **match distributions** rather than energies!
Quantify Information Loss

\[ p(\sigma) = \frac{\exp(-\beta E(\sigma))}{Z} \]

\[ p(\sigma|\gamma) = \frac{\exp(-\beta E(\sigma; \gamma))}{Z(\gamma)} \]

“Distance” of choice:

\[ S[\gamma] = \int_{\mathcal{M}} p(\sigma) \ln \left( \frac{p(\sigma)}{p(\sigma|\gamma)} \right) \, d\sigma \geq 0 \]

Relative Entropy (Rel Ent)

Choose ECI to minimize area!
**REL ENT VS. LEAST SQUARES**

**Gaussian approximation of**

\[
S[\gamma] = \int_{\mathcal{M}} p(\sigma) \ln \left( \frac{p(\sigma)}{p(\sigma;\gamma)} \right) d\sigma
\]

We ideally minimize

\[
S[\gamma] \approx \frac{\beta^2}{2} \text{Var}[E(\sigma) - E(\sigma;\gamma)]
\]

Least squares ideally minimizes

\[
\mathcal{L}[\gamma] = \sum_{\sigma} (E(\sigma) - E(\sigma;\gamma))^2
\]

Matching distributions becomes a weighted least squares problem (from minimizing \(S\) above) with weights

\[
(I_N - p_N 1_N) \text{diag}(p_N) (I_N - p_N 1_N)^t
\]

\[
p_N := \left( \frac{\exp(-\beta E(\sigma^{(1)}))}{Z_N}, \ldots, \frac{\exp(-\beta E(\sigma^{(N)}))}{Z_N} \right)
\]

\(t = \text{transpose}\)
Relative Entropy

Low $T$
- Ignore high energy states

High $T$
- All states equally likely

Least Squares

Energy

States
How to Compute Phase Transitions
Thermodynamics using MCMC

Phase Space Probability

\[ p(\sigma | \beta_0) \approx \sum_{i=1}^{N_P} w_0^{(i)} \delta(\sigma - \sigma^{(i)}) \]

Particle approximation

More steps taken

\[ \langle Q(T) \rangle = \sum_{j=1}^{N_P} w_s^{(j)} Q(\sigma_s^{(j)}) \]

Thermodynamic quantity \( Q \) (specific heat in case of phase transition)

Fully parallelizable: Each particle on its own core
ASMC Algorithm: Implementation

1. Set $s = 0$ and $\gamma_s = 0$. Sample a particle approximation from $p_{\gamma_0}(\sigma)$, $\left\{w_0^{(i)}, \sigma_0^{(i)}\right\}_{i=1}^N$.

2. Determine $\gamma_{s+1} \in [\gamma_s, 1]$ s.t:

$$ESS(\gamma_{s+1}) = \frac{1}{\sum_{i=1}^N (w_{s+1}^{(i)})^2} = \zeta ESS(\gamma_s),$$

where $w_{s+1}^{(i)}$ is the normalized version of $W_{s+1}^{(i)} = w_s^{(i)} \hat{w}_{s+1}^{(i)}$ with:

$$\hat{w}_{s+1}^{(i)} = \frac{p_{\gamma_{s+1}}\left(\sigma_{s}^{(i)}\right)}{p_{\gamma_s}\left(\sigma_{s}^{(i)}\right)} = e^{-(\beta_{\gamma_{s+1}} - \beta_{\gamma_s})E_s^{(i)} + (\beta_{\gamma_{s+1}}\mu_{\gamma_{s+1}} - \beta_{\gamma_s}\mu_{\gamma_s}) \sum_j \sigma_{s,j}^{(i)}}.$$

3. If $ESS(\gamma_{s+1}) < ESS_m$, then resample.

4. Draw samples $\left\{\sigma_{s+1}^i\right\}_{i=1}^N$ from $p_{\gamma_{s+1}}(\sigma)$.

5. If $\gamma_s = 1$, STOP. Otherwise, set $s \leftarrow s + 1$ and go to 2.
CASE STUDY: SILICON GERMANIUM

Predict two-phase coexistence to disorder phase transition at 50 % composition
50 % Si-Ge: Fit to Energies

- Use ASMC to obtain phase transitions

**Existing literature**


**Method:**
Least Squares (ATAT) + traditional MCMC
Silicon Germanium

Why the similarity?

Note: This is not probabilistic, but gives an idea of the behavior versus temperature.
CASE STUDY: MAGNESIUM LITHIUM

Predict order/disorder phase transitions at 33 %, 50 %, and 66 % Mg
### Magnesium Lithium

<table>
<thead>
<tr>
<th>Mg Composition</th>
<th>Temperature Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>33% Mg</td>
<td>~190 K</td>
</tr>
<tr>
<td>50% Mg</td>
<td>~300-450 K</td>
</tr>
<tr>
<td>66% Mg</td>
<td>~210 K</td>
</tr>
</tbody>
</table>

Fitting observed energies (cluster expansion)

---


All compositions: ~140-200 K

Experimental (non-conclusive)

---

Method:
Genetic Algorithm + MCMC

Experiments:
140-200 K

33% Mg

Experiments:
140-200 K

50% Mg

Experiments:
140-200 K
(largest error)

66% Mg

Method:
Genetic Algorithm + MCMC
Why the difference?

Large Difference (~0.5 meV)

Region of transitions
### Summary table

<table>
<thead>
<tr>
<th>Alloy</th>
<th>$x$</th>
<th>Fit energies (our work)</th>
<th>Fit energies (literature)</th>
<th>Relative Entropy</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si$<em>x$Ge$</em>{1-x}$</td>
<td>50 %</td>
<td>~339 K</td>
<td>~325 K</td>
<td>~339 K</td>
<td>N/A</td>
</tr>
<tr>
<td>Mg$<em>x$Li$</em>{1-x}$</td>
<td>33 %</td>
<td>~226 K</td>
<td>~190 K</td>
<td>~170 K</td>
<td>~140-200 K</td>
</tr>
<tr>
<td></td>
<td>50 %</td>
<td>~304 K</td>
<td>~300-450 K</td>
<td>~214 K</td>
<td>~140-200 K</td>
</tr>
<tr>
<td></td>
<td>66 %</td>
<td>~207 K</td>
<td>~210 K</td>
<td>~240 K</td>
<td>~140-200 K       (largest error)</td>
</tr>
</tbody>
</table>
Bayesian Approach to Predicting Materials Properties

Propagating Uncertainty From A Surrogate To, e.g., A Phase Transition
The former part of this presentation does not *per se* offer ways of answering central questions such as:

- What is the uncertainty in a quantity of interest (e.g., a phase transition) given that we do not know the best cluster expansion and that we have limited data?

- When computing a phase transition we want to know how uncertain we are about its value

- Unknown whether what we predict is OK

The Bayesian approach can provide answers to such questions

We show how *surrogate models* can be used to accomplish this
**Fully Bayesian Approach**

- Probability means a reasonable degree of belief*

- **Prior** belief on clusters + ECI

- **Likelihood** function
  - Given a model (i.e., set of clusters + ECI)
    - how likely is $D$

- **Posterior** belief on clusters + ECI

- Use Bayes theorem** to update degree of belief upon receiving new evidence $D$ (what $D$ is depends on the application)

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

*Note:

$D$ is limited, we can only see so many observations

*Laplace, Analytical Theory of Probability (1812)*

**Bayes, Thomas. Philosophical Transactions (1763)*
Propagating Uncertainty

Prior on property (quantity of interest) “$I$”

\[ p(I) = \int d\theta \delta(I[f(\cdot; \theta)] - I)p(\theta) \]

Notice how we integrate out the clusters and the ECI!
(In principle) all cluster and ECI choices (models) consistent with $D$ are considered

Then we observe an expensive data set $D$ which helps us to learn more about the clusters and the ECI

- In this work the data set was expensive energy computations

$$D = \{ x_i, f(x_i) \}$$

Likelihood: What information the data contains about the clusters and ECI
The Quantity of Interest

- Different quantities of interest $I$ can require different data sets $D$
- This framework allows for very general quantities of interest $I$

- Some examples:
  - $I = \text{phase transition}$
    - The phase transition is found from the internal energy
    - The data set $D$ consists of high-accuracy (expensive) energies
  - $I = \text{ground state line}$
    - The ground state line is found from the internal energy as well
    - The data set $D$ consists of expensive energies
  - $I = \text{maximum band gap structure}$
    - The maximum band gap structure is found, e.g., from knowing the band gap of each structure or the entire band diagram
    - The data set $D$ consists of expensive band gaps
How likely is the cluster expansion upon seeing $D$ including what we knew before? (Bayes theorem!)

We stress here that we now have a probability distribution on the property—not a single estimate.

From this, the uncertainty estimate follows.

We now have a way to propagate uncertainty from the cluster expansion to the quantity of interest.

Next, we select a Bayesian posterior.
Choose Bayesian posterior*

\[ p(\theta|D) \propto \Gamma(k)B(k, p - k + 1) \left\| J \right\|_1^{-k} \left\| y - XJ \right\|_2^{-n} \]

- \( k = \text{model complexity (\# of clusters)} \)
- \( p = \text{arbitrary max set of clusters to be used} \)

Based on LASSO-inspired priors
- Models describing physics are typically sparse**

Expectation values with this posterior not in closed form!
- Resolution: MCMC Sampling

---

* C. Xiaohui, J. Wang, and M. McKeown (2011)
Motivating Model Selection

\[ p(I | D) = \int d\theta \delta(I[f(\cdot; \theta)] - I)p(\theta | D) \]

- There is an infinite number of cluster expansions (each symbolized by its own \( \theta \) in the integral above)
  - Which ones are \textit{most relevant} to determining the value of the integral?
- We now explore model selection as an option
Can Model Selection be Used to Quantify Epistemic Uncertainties With Limited Data?

J. Kristensen and N. Zabaras. Computer Physics Communications 185.11 (2014)
MODEL SELECTION
SELECTION OF BOTH BASIS FUNCTIONS AND EXPANSION COEFFICIENTS
We used reversible jump Markov chain Monte Carlo (RJMCMC)* to perform the model selection.

Define 3 move types:
- Birth step (+1)
- Death step (-1)
- Update step (0)

Practically speaking it behaves like a standard MCMC chain:
- Use 50% burn-in
- Use thinning if you want to (for memory reasons, e.g.)

RJMCMC CHAIN: ALGORITHM

Input: The number of iterations $T$. Random walk step size $\epsilon$.
Data: $X$ and $y$.
Output: $\{\theta^{(t)} = (\beta^{(t)}, \gamma^{(t)}) | t \in \{0, \ldots, T\}\}$.

begin
Initialisation: set $\theta^{(0)} = (\beta^{(0)}, \gamma^{(0)})$ and $t = 1$.
repeat
if $k^{(t-1)} = 1$ then
| $k^{(t)} \leftarrow k^{(t-1)} + U((0,1))$.
else if $k^{(t-1)} = p$ then
| $k^{(t)} \leftarrow k^{(t-1)} - U((0,1))$.
else
| $k^{(t)} \leftarrow k^{(t-1)} + U((-1,0,1))$.
end
Sample $s \sim N(0, \epsilon^2)$.
$K \leftarrow \gamma^{(t-1)}$ and $K^c \leftarrow \{1, \ldots, p\} \setminus K$.
if $k^{(t-1)} = k^{(t)}$ then
| Sample $j \sim U(K)$.
| Update $\beta_j^{(t)} \leftarrow \beta_j^{(t-1)} + s$ with an MH step, details in Section 3.2.
else if $k^{(t)} = k^{(t-1)} + 1$ then
| Sample $j \sim U(K^c)$.
| Perform a “birth” move and update $\beta_j^{(t-1)}$, details in Section 3.3.
else
| Sample $j \sim U(K)$.
| Perform a “death” move and update $\beta_j^{(t-1)}$, details in Section 3.3.
end
$t \leftarrow t + 1$.
until $t = T$.
end

Model selection of clusters

Model selection of the ECI

Standard Metropolis-Hastings*

* $N. Metropolis, et al. The journal of chemical physics 21.6 (1953)$
RESULTS ON REAL ALLOYS
Model Selection Results

\[ p(\theta|D) \propto \Gamma(k) B(k, p - k + 1) \|J\|_1^{-k} \|y - XJ\|_2^{-n} + \text{RJMCMC} \]

Any particular blue point in upper plot represents a cluster expansion truncation:
1) y-axis measures number of included clusters (but not which).
2) Actual values of ECI not shown.

Noise agrees with DFT
Bayesian uncertainty in ground state line with limited data

\[ p(I|D, \cdot) = \int d\theta \delta(I[f(\cdot; \theta)] - I)p(\theta|D, \cdot) \]

\( I = \text{ground state line} \)

Data set from VASP* (expensive energies)

Material: MgLi

Predictive variance of ground state line is around 12%.

We can implicitly conclude whether the data set is large enough!

This is the uncertainty induced in the quantity of interest from the uncertainty in the surrogate model

- If error bars too large: you need to increase/change your data set!

Bayesian uncertainty in phase transition from two-phase coexistence to disorder with limited data

\[ p(I|D, \cdot) = \int d\theta \delta(I[f(\cdot; \theta)] - I)p(\theta|D, \cdot) \]

Material: SiGe at 50 %

Predictive variance of phase transition is around 6 %

Transitions computed via ASMC

This is the uncertainty induced in the quantity of interest from the uncertainty in the surrogate model

- If error bars too large: you need to increase/change your data set!
Using Surrogates for Designing Materials

J. Kristensen and N. Zabaras. In review (2014)
We are now confident about the predictive capabilities of surrogate models. Can we also use surrogates for designing new structures with specified properties?

Materials by design

Application:
- Optimize thermal conductivity in nanowires*
  - Heat dissipation in nanochips
  - Thermoelectric materials
    - Solar cells
    - Refrigeration

But: Nanowires require a different way of using the cluster expansion
- by “the cluster expansion” we mean the standard bulk expansion implemented in, e.g., ATAT**
- We show shortly how we addressed this issue

Which Si-Ge Nanowire Configuration Minimizes the Thermal Conductivity?
Find the configuration with **lowest thermal conductivity**

Green-Kubo method:*  

\[ \kappa = \frac{1}{V k_B T^2} \lim_{\tau_m \to \infty} \int_0^{\tau_m} \langle J_x(\tau) J_x(0) \rangle d\tau \]

using microscopic heat current:

\[ \mathbf{J} = \sum_{i=1}^{N} \epsilon_i \mathbf{v}_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left( \frac{\partial V_{ij}}{\partial \mathbf{r}_j} \cdot \mathbf{v}_j \right) \mathbf{r}_{ij}, \]

and a Tersoff** potential energy b/w bonds:

\[ V_{ij} = f_C(\mathbf{r}_{ij}) [a_{ij} f_R(\mathbf{r}_{ij}) + b_{ij} f_A(\mathbf{r}_{ij})] \]

---

*R. Kubo Journal of the Physical Society of Japan 12.6 (1957)  
*M. Green The Journal of Chemical Physics 20.8 (1952)  
Alloy optimization problem

- Use cluster expansion surrogate

**Problem for nanowires:**

- Low-symmetry system
- ECI become layer-dependent close to surfaces
- Easily *thousands of unknowns!*

Energy is additive so we can write*:

\[
\Delta H_{f}^{\text{CE}} = \Delta H_{f}^{\text{vol}} + \Delta H_{f}^{\text{surf}}
\]

- Bulk part
- Surface part

---

*D. Lerch et al. Modelling and Simulation in Materials Science and Engineering 17.5 (2009)*
**New Cluster Expansion Approach**

- **Idea:** embed structure of *any* geometry in *sea of ghost sites*
- **Group clusters under bulk symmetries**
  - We get bulk contribution alone, but for *any* geometry!

\[
\Delta H_f^{CE} = \Delta H_f^{vol} + \Delta H_f^{surf}
\]

Could be any shape on any lattice: bcc sphere, fcc nanowire, sc 2D sheets, etc.

Ghost lattice method (GLM)
GLM on Nanowire Project

- Nanowire implementation with the GLM
  - Two different representations of the same wire (OVITO* used for visualization)

**ATAT representation**
- One periodic image of the wire
- “Image” Si
- “Image” Ge

**LAMMPS representation**
- End view
- Side view

Compute correlation functions with ATAT modified for GLM (i.e., modified to parse ghosts)

Compute thermal conductivity in LAMMPS

*A. Stukowski, Modelling and Simulation in Materials Science and Engineering 18.1 (2010)*
Bulk Si and Ge (easy case): Use method in Ref. [*] in LAMMPS**

We predict 170 W/m.K for Silicon. Experimental value = 150 W/m.K.

We predict 90 W/m.K for Germanium. Experimental value is 60 W/m.K.

Tersoff is known to overshoot. 
We obtain great agreement!

Verify Nanowire Implementation

- Compare data with Ceder’s group at MIT*

Annealed heating of the (relatively large) surface area was necessary.

<table>
<thead>
<tr>
<th>W/mK</th>
<th>Pure Si wire</th>
<th>PPG (defined later)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our work</td>
<td>4.1 +/- 0.4</td>
<td>0.12 +/- 0.03</td>
</tr>
<tr>
<td>(LAMMPS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceder group</td>
<td>4.1 +/- 0.3</td>
<td>0.23 +/- 0.05</td>
</tr>
<tr>
<td>(XMD)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Main sources of discrepancy
- thermalization techniques
- MD software
- thermalization times

140 wires each with random Si/Ge configuration
- This is the random nanowire dataset (RW)

Split RW into train and test sets
- Train CE with GLM on train

Additional data sets:
- Planes of pure Ge (PPG)
- Similar to PPG (SPPG)
  - Perturbed: atom(s) from plane swapped with atom (s) from non-plane region
Using the new cluster expansion surrogate approach to fit the nanowire data set.

Now that we have surrogate; find global minimum.
We find the PPG to have lowest thermal conductivity.

Very strong case for the GLM:

- Evidence that thermal conductivity of nanowires is well captured by first term.

\[ \Delta H_f^{\text{CE}} = \Delta H_f^{\text{vol}} + \Delta H_f^{\text{surf}} \approx 0 \]

SPPGs generally lower than RW train and test sets as expected.
From Ref. [*] on the same problem
(but using a slightly different surrogate model)

They found as well that the PPG wire has lowest $\kappa$

Concluding Remarks
- Quantify uncertainties in
  - Band gaps
  - Energies
  - Phase Diagrams
  - Thermal Conductivities
  - Any material property

- Use information theory to design materials

- Help improve how data is collected (and the resources spent in doing so) in general
  - Choosing the limited data set in most informed way

- What happens to uncertainty quantification across length and time scales?
  - How do uncertainties in microscopic properties affect macroscopic properties?