A DATA-DRIVEN APPROACH TO PREDICTIVE MULTISCALE MATERIALS MODELLING

Nicholas Zabaras
Warwick Centre for Predictive Modelling
The University of Warwick
http://www2.warwick.ac.uk/wcpm/
and
Institute for Advanced Study, Technische Universität Munchen
http://www.tum-ias.de/

http://www.zabaras.com/
nzabaras@gmail.com

EPSRC
Engineering and Physical Sciences Research Council

DARPA

IAS TUM Institute for Advanced Study

European Commission

NIST Workshop, Quantification of Uncertainties in Materials Science, Gaithersburg, MD, January 14-15, 2016

Outline

- *Exchange-Correlation Functional in DFT*: Uncertainty Quantification and Propagation
- Quantifying *uncertainties when using surrogate* (DFT) *models* trained with a few ab initio simulations
- *Design of Experiments:* What are the most informative simulations needed for predicting a given QoI?
- A few words on *stochastic coarse-graining*: A Bayesian generative approach
- *Scalable UQ for Multiscale Problems*: A Graph-Theoretic Approach (EP, Belief propagation, etc.)

DFT: Exhange Correlation Energy

- DFT energy functional (as a function of charge density):
  - $E[n] = U^{ext}[n] + T^{0}[n] + U^{ee}[n] + E^{xc}[n]$
    - $U^{ext}[n]$: Interaction with external potential
    - $T^{0}[n]$: non-interacting kinetic energy
    - $U^{ee}[n]$: classical electro-electron-interaction
    - $E^{xc}[n]$: exchange-correlation energy, unknown

R. Jones et al. (1989)
R. Jones (2015)
W. Kohn and L. Sham (1965)

- Different levels of approximations for $E^{xc}[n] = \int n\varepsilon^{xc}(n;\mathbf{r})\,d\mathbf{r}$
  - Local density approx. (LDA) $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r})]$
  - Generalized gradient approx. (GGA) $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r}), \nabla n(\mathbf{r})]$
  - Meta-Generalized gradient approx. (meta-GGA) (used here)
    $\varepsilon^{xc}(n;\mathbf{r}) = \varepsilon^{xc}[n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r})] \quad \left(\tau(\mathbf{r}) = 2\sum_i{}' \tfrac{1}{2}|\nabla\psi_i(\mathbf{r})|^2\right)$
  - Hybrid (including exact exchange)

J. P. Perdew et al. (2001)

Linear Model for the Exchange Energy

- We work within the meta-GGA framework, it provides good cost-accuracy trade-off.

- Furthermore, *we will focus on the exchange energy only:*

$$E^x[n] = \int n\varepsilon^x(n, \nabla n, \tau)d\mathbf{r} = \int n\varepsilon^x_{UEG}(n)F^x(n, \nabla n, \tau)d\mathbf{r}$$

- We look for the *exchange enhancement factor* $F^x(s, \alpha)$ as a linear regression model using Legendre polynomials:

$$F^x(s,\alpha) = \sum_i \xi^x_i \phi_i(s,\alpha) = (\xi^x)^T \phi(s,\alpha)$$

$$= \sum_i^{M_s} \sum_j^{M_\alpha} \xi^x_{ij} P_i(t_s(s)) P_j(t_\alpha(\alpha))$$

- Rescale s and α on [-1,1] based on earlier non-empirical functionals: PBEsol, MS, etc.

J. Wellendor et al. (2014)

Linear Model for the Exchange Energy

- The model for the exchange energy functional, $E^x[n]$, is:

$$E^x[n;\boldsymbol{\xi}^x] = \int n\varepsilon^x_{UEG}(n)\sum_{i=1}^{M}\xi^x_i\phi_i(s,\alpha)d\mathbf{r}$$

Use PBE functional

$$= \sum_{i=1}^{M}\xi^x_i\int n\varepsilon^x_{UEG}(n)\phi_i(s,\alpha)d\mathbf{r} = \sum_{i=0}^{M-1}\xi^x_i E^x[n;\hat{\mathbf{e}}_i] = (\boldsymbol{\xi}^x)^T\mathbf{E}^x[n;\hat{\mathbf{e}}]$$

- We use a data driven Bayesian framework to find the expansion coefficients
  - Coefficients are random variables
  - The exchange energy accounts for uncertainty:
    - model error (meta-GGA framework)
    - limited data (used in the training of the model)

M. Aldegunde, J. Kermode, N. Zabaras, JCP (submitted, 2015)

Bayesian Linear Regression

- Bayesian linear regression, we need:
  - Likelihood: assume Gaussian with noise precision β

$$\mathcal{L}(\mathbf{t} \mid \mathbf{n}, \boldsymbol{\xi}, \beta) = \prod_{i=1}^{N} \mathcal{N}(t_i \mid \boldsymbol{\xi}^T \mathbf{E}^x[n_i; \hat{\mathbf{e}}], \beta^{-1})$$

  - Priors on parameters:
    - Gaussian for coefficients ξ $\quad p(\boldsymbol{\xi} \mid \beta, \mathbf{m}_0, \mathbf{S}_0) = \mathcal{N}(\boldsymbol{\xi} \mid \mathbf{m}_0, \beta^{-1}\mathbf{S}_0)$
    - Gamma for noise precision β $\quad p(\beta \mid a_0, b_0) = \mathcal{G}(\beta \mid a_0, b_0)$

- The resulting posterior is:

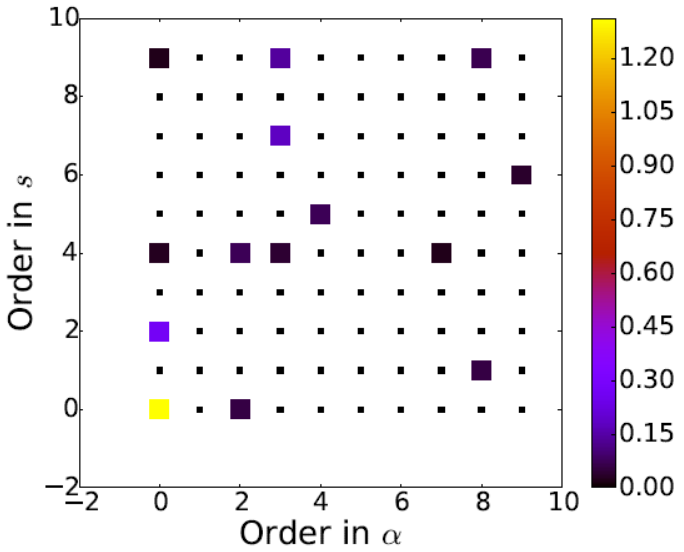
$$p(\boldsymbol{\xi}, \beta \mid \mathbf{t}) = \frac{\mathcal{L}(\mathbf{t} \mid \mathbf{n}, \boldsymbol{\xi}, \beta)\, p(\boldsymbol{\xi}, \beta)}{\int \mathcal{L}(\mathbf{t} \mid \mathbf{n}, \boldsymbol{\xi}, \beta)\, p(\boldsymbol{\xi}, \beta) d\boldsymbol{\xi} d\beta} = \mathcal{N}(\boldsymbol{\xi} \mid \mathbf{m}_N, \beta^{-1}\mathbf{S}_N)\mathcal{G}(\beta \mid a_N, b_N)$$

- The predictive distribution is a Student t-distribution:

$$p(\tilde{t} \mid \tilde{n}, \mathbf{t}) = \int p(\tilde{t} \mid \tilde{n}, \boldsymbol{\xi}, \beta)\, p(\boldsymbol{\xi}, \beta \mid \mathbf{t}) d\boldsymbol{\xi} d\beta = St(\tilde{t} \mid \mu, \lambda, \upsilon)$$

Relevance Vector Machine

- Hyperparameters?
  - a_0, b_0, $\mathbf{m}_0$, $\mathbf{S}_0$
  - Sparsity inducing prior: relevance vector machine
    - $\mathbf{m}_0=0$; $\mathbf{S}_0=\text{diag}(\alpha^{-1}_0, \ldots, \alpha^{-1}_{M-1})$
  - Use of evidence approximation: *maximize the log of the marginal likelihood*:



$$p(\mathbf{t}\,|\,\mathbf{m}_0,\mathbf{S}_0,a_0,b_0)=\int p(\mathbf{t}\,|\,\xi,\beta)\,p(\xi,\beta\,|\,\mathbf{m}_0,\mathbf{S}_0,a_0,b_0)\,d\xi\,d\beta$$

$$\mathcal{E}(\mathbf{m}_0,\mathbf{S}_0,a_0,b_0)=\log p(\mathbf{t}\,|\,\mathbf{m}_0,\mathbf{S}_0,a_0,b_0)$$

$$\mathcal{E}(\mathbf{m}_0,\mathbf{S}_0,a_0,b_0)=\frac{1}{2}\log\frac{|\mathbf{S}_N|}{|\mathbf{S}_0|}-\frac{N}{2}\log(2\pi)$$

$$+\log\frac{\Gamma(a_N)}{\Gamma(a_0)}+a_0\log(b_0)-a_N\log(b_N)$$

M. E. Tipping (2001), C. Bishop (2006), Bilionis & Zabaras (2014)

Training Data

- Train the model with *atomization energies* for molecules and solids.
  - For a material $M=A_{n_A}B_{n_B}\ldots$

$$E_{at} = \frac{1}{N}\left(\sum_I n_I E_I - E_M\right), I = A, B, ..$$

- In solids, we can add *indirect measurements* using an equation of state (adds extra uncertainty):
  - Given the atomization energy (E_0), equilibrium volume (V_0), bulk modulus and pressure derivative (B_0, B_1) we can obtain the energy of the strained material:

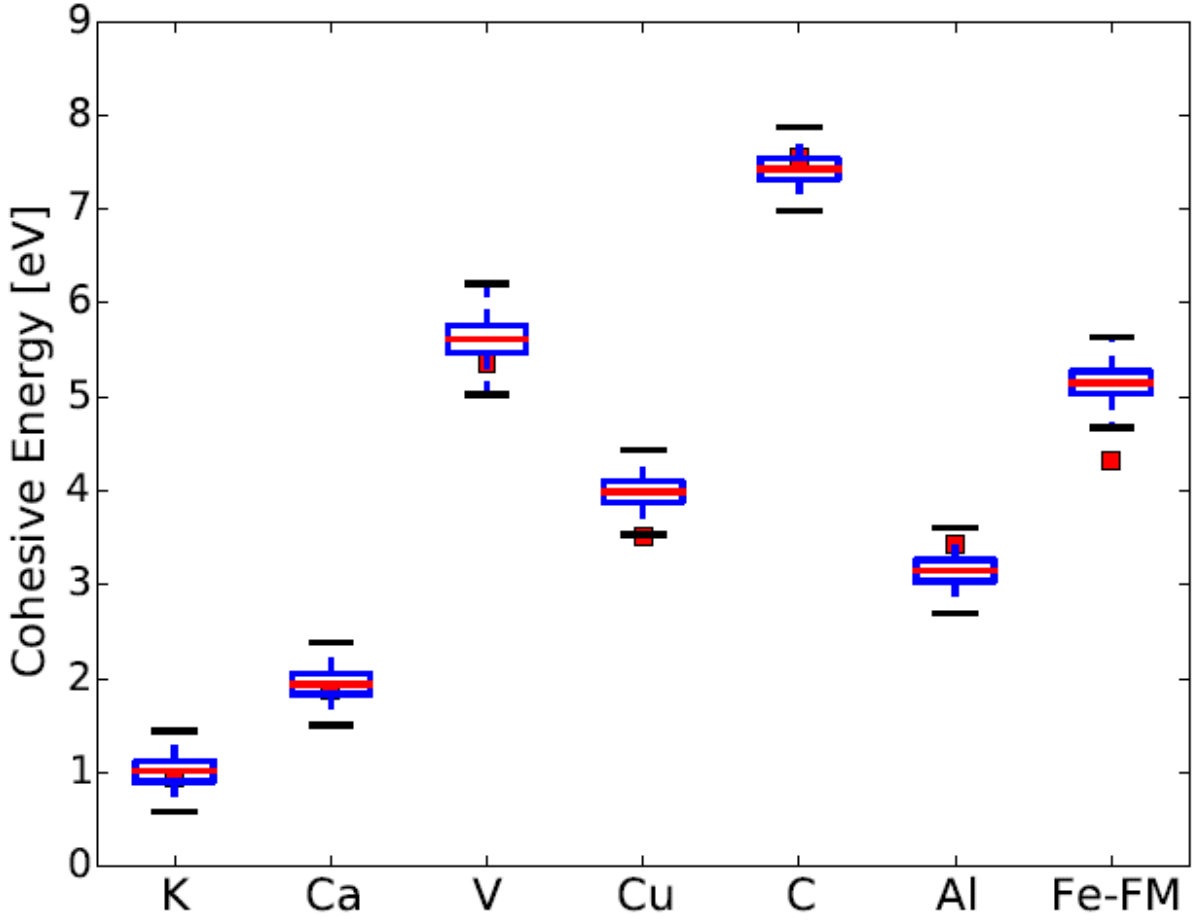
$$E(V) = a + b\frac{V_0^{1/3}}{V^{1/3}} + c\frac{V_0^{2/3}}{V^{2/3}} + d\frac{V_0}{V} = \boldsymbol{\gamma}^T \boldsymbol{\phi}(V)$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 3 & 2 & 1 & 0 \\ 18 & 10 & 4 & 0 \\ 108 & 50 & 16 & 0 \end{pmatrix} \boldsymbol{\gamma} = \begin{pmatrix} -E_0 \\ 0 \\ 9V_0B_0 \\ 27V_0B_0B_1 \end{pmatrix}$$

A. B. Alchagirov et al. (2001)

Atomization Energies

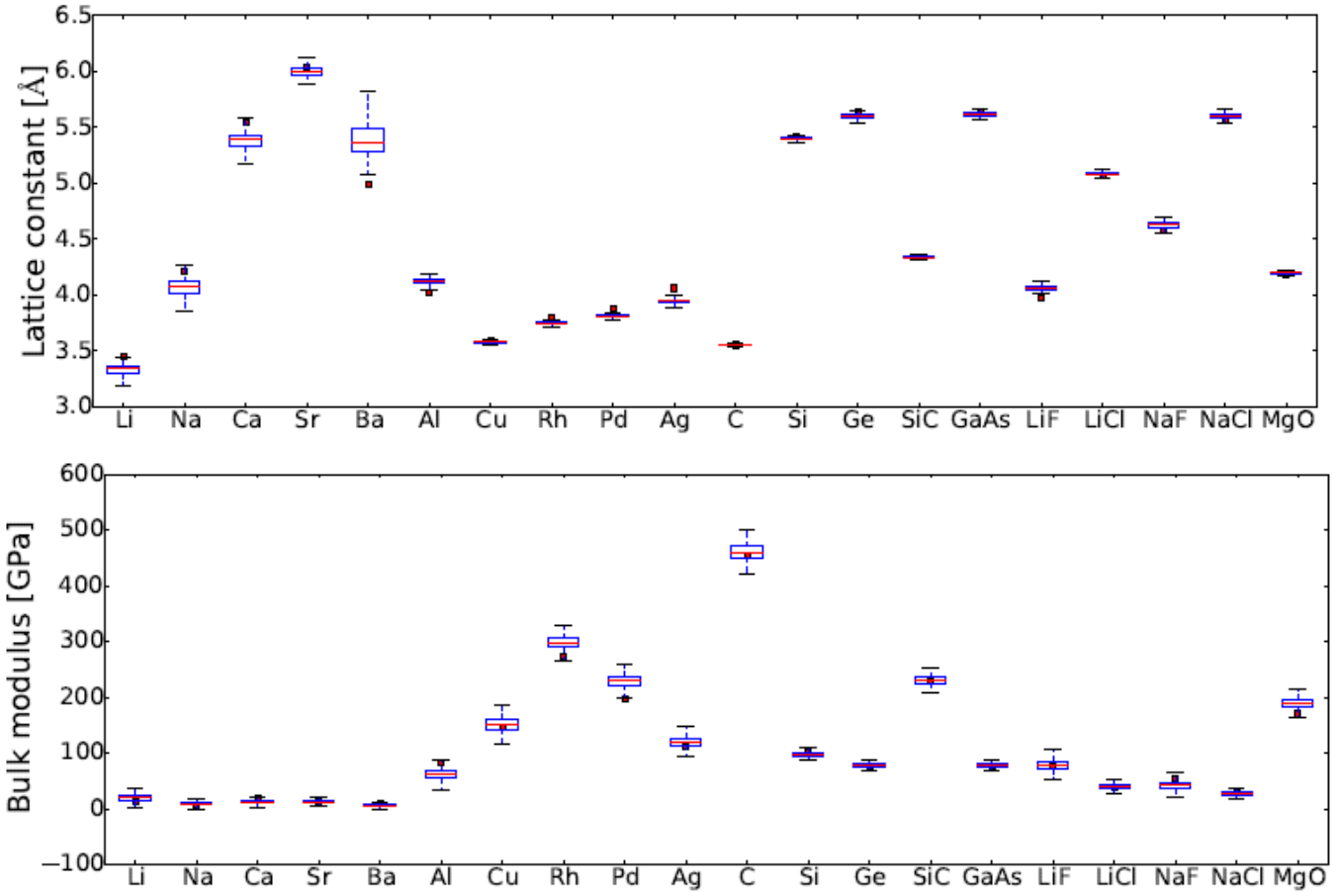
- Predicting cohesive (atomization) energies:
  - Explicit expression (Student t-distribution)



M. Aldegunde, J. Kermode, N. Zabaras,, JCP (2016)

Lattice Constants and Bulk Moduli

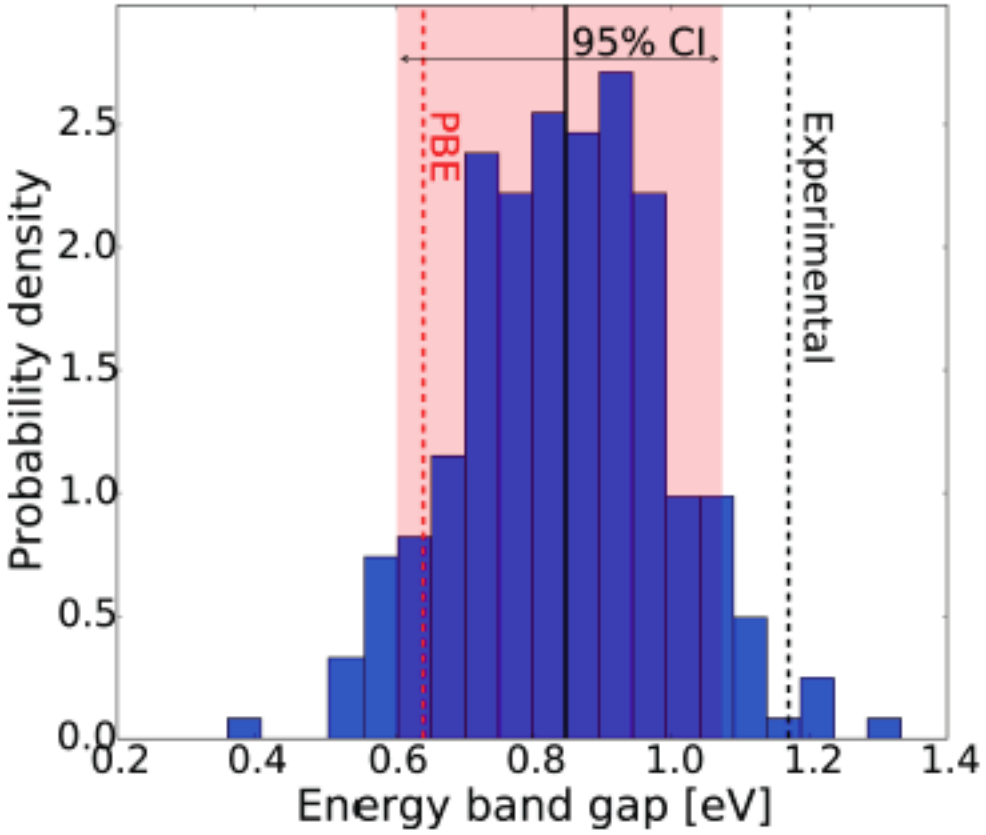
- Run self-consistently for five different strains to compute n_i^* (use m_N as the average model)
- Sample from $p(\xi, \beta|t)$ and run non-self consistent simulation using n_i^*
- Fit the EOS to the values from this X energy (Bayesian fit)
- Sample coefficients from the fitting $p(\gamma|E)$ to calculate V_0, B_0, B_1



Uncertainty in band structure: DFT

- We run a self-consistent calculation of the solid of interest (PBE)
- Calculate band structure with fixed density
- For each sample of the XC coefficients different band structure

- Histogram of the band gap of Silicon obtained from sampling realizations of the XC functional.
- Black vertical line: average XC functional.
- Red dotted vertical line: PBE functional.
- Black dotted vertical line: experimental value.

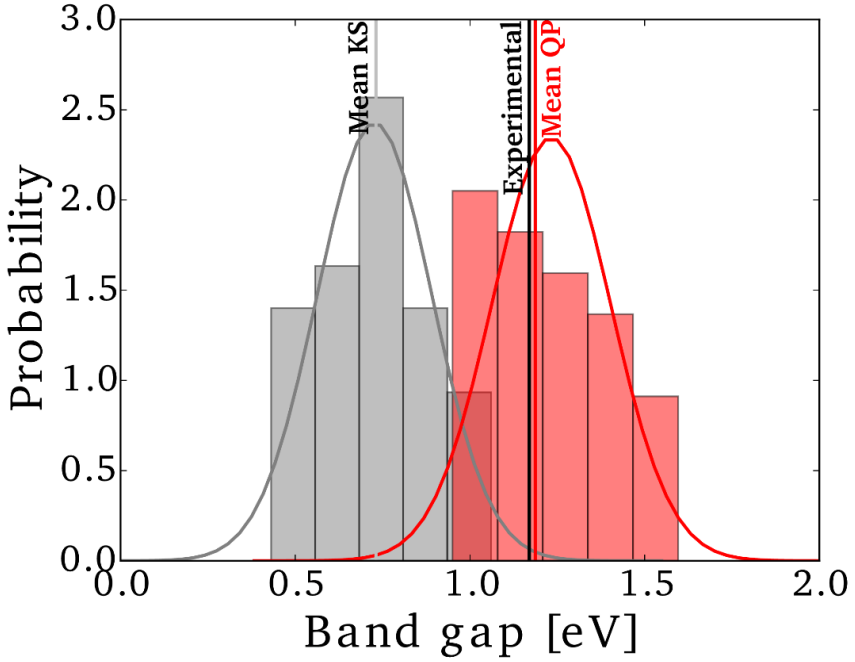


Semilocal XC functionals cannot reproduce the experimental data

UQ in Band Structure: Quasi-Particle Approximation

- Use GW approximation of the self-energy $\Sigma = iGW^*$ of many-body system of electrons. Modelling the interaction of charged particles with the polarization they induce in the surrounding medium (QP).
- Approximate the quasi-particle energies non-self consistently as a correction to the KS energies (GPAW) - G_0W_0 quasi-particle approximation.
- The results will depend on the XC coefficients sample and this will give rise to a probability model for the band structure

Example: Band gap of bulk Si from Kohn-Sham (KS) energies (**grey**) and from the G_0W_0 approximation (QP) (**red**). The experimental (**black**) and mean values are shown as vertical lines for reference.



* G=Green's function of the single particle, W=Screened Coulomb interaction

Why Surrogates work in alloy modeling?

- Replace VASP response surface with surrogate model
  - What should the surrogate model be?
- Quantum Mechanical energy invariant under space group operations of the lattice
  - Surrogate must account for this
- Possible (and popular) surrogate in alloy modeling:
  - The Cluster Expansion*

Expansion coefficients (ECI)

$$E^{(i)} = \sum_{F} J_F \left\langle \Pi_F \right\rangle_{\sigma^{(i)}}$$

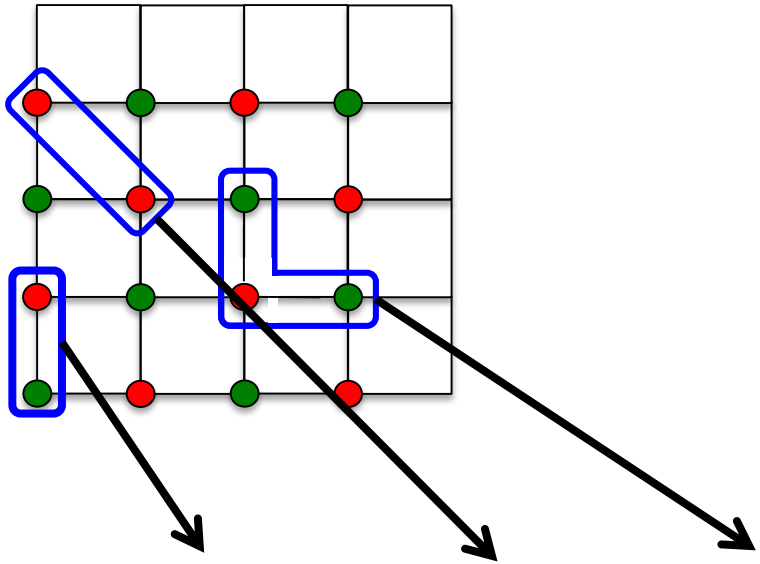
Correlation functions (Basis)

Accounts for symmetries

J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A 128, 334 (1984).

The Cluster Expansion

- Corresponds to: Multidimensional discrete Fourier transform*
- Generalized Ising model (map atoms to integers)
- For binary alloys correlation functions reduce to simple products of atoms σ on atomic sites:



$$E^{(t)} = J_0 + J_1 \sum_i \sigma_i^{(t)} + J_{2,1} \sum_{i,j} \sigma_i^{(t)} \sigma_j^{(t)} + \cdots + J_{3,1} \sum_{i,j,k} \sigma_i^{(t)} \sigma_j^{(t)} \sigma_k^{(t)} + \cdots$$

(Not general in this form)

*Sanchez, J. M. *Physical Review B* 81.22 (2010)

PARAMETRIZATION

- Infinite series: True response surface is part of parametrization class
- Truncation required:
  - How do we choose truncated parametrization?

$$E^{(t)} = J_0 + J_1 \sum_i \sigma_i^{(t)} + J_{2,1} \sum_{i,j} \sigma_i^{(t)} \sigma_j^{(t)} + \cdots + J_{3,1} \sum_{i,j,k} \sigma_i^{(t)} \sigma_j^{(t)} \sigma_k^{(t)} + \cdots$$

$$\gamma = (1, 1, 0, \cdots, 1, \cdots)$$

$$\beta_\gamma = (J_0, J_1, 0, \cdots, J_{3,1}, \cdots)$$

$$\theta = (\gamma, \beta_\gamma)$$

$$E^{(i)} = \sum_F J_F \langle \Pi_F \rangle_{\sigma^{(i)}} \longrightarrow \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon}$$

Uncertainty on the QoI

- Observe data set *D* and form the posterior

Posterior on QoI

Determined by QoI

We need to model this

$$p(I|D,\cdot) = \int \mathrm{d}\theta \delta(\hat{I}[f(\cdot;\theta)] - I) p(\theta|D,\cdot)$$

- Our choice, using Laplace-inspired priors on the ECI:

$$p(\theta|D) \propto \Gamma(k) B(k, p-k+1) ||\boldsymbol{J}||_1^{-k} ||y - X\boldsymbol{J}||_2^{-n}$$

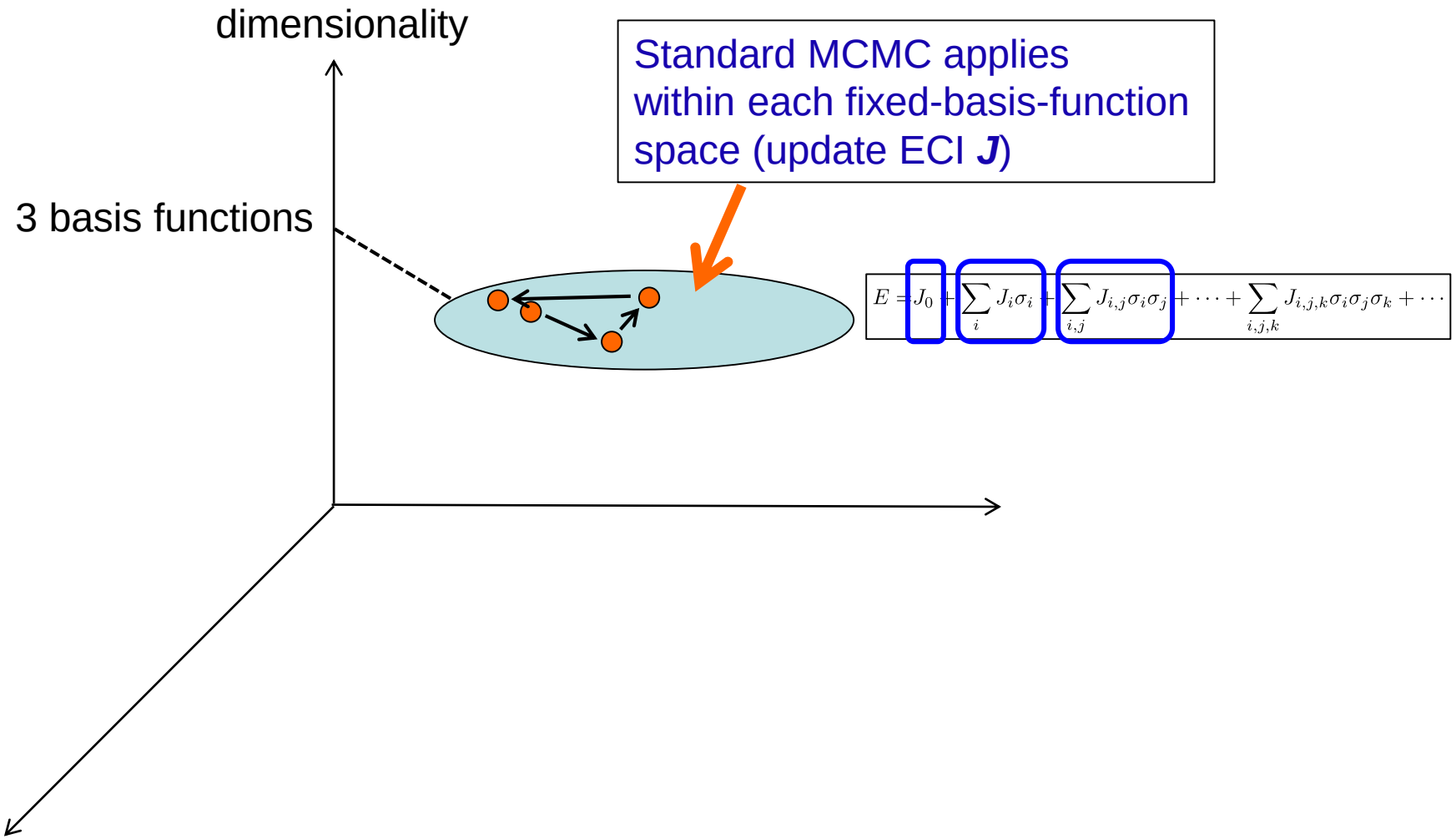
k = model complexity (# of clusters)

lasso

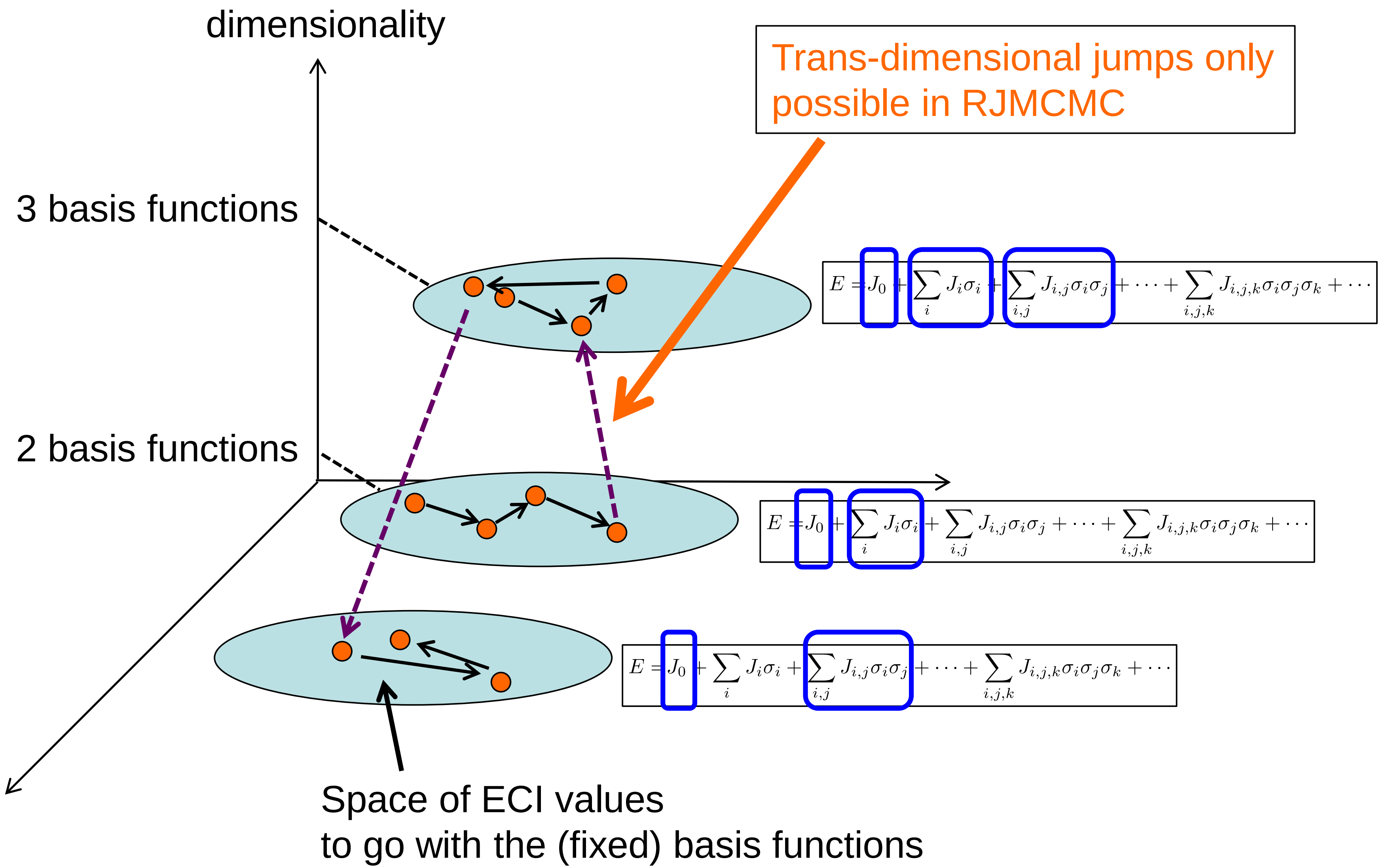
- Expectations wrt. *p*(•) are not in closed form!
- Solution: We need to sample from it
  - Standard Markov chain Monte Carlo (MCMC) does not suffice
  - **Solution:** Use RJMCMC*

*Green, Peter J. Biometrika 82.4 (1995)

Visualizing RJMCMC



Visualizing RJMCMC



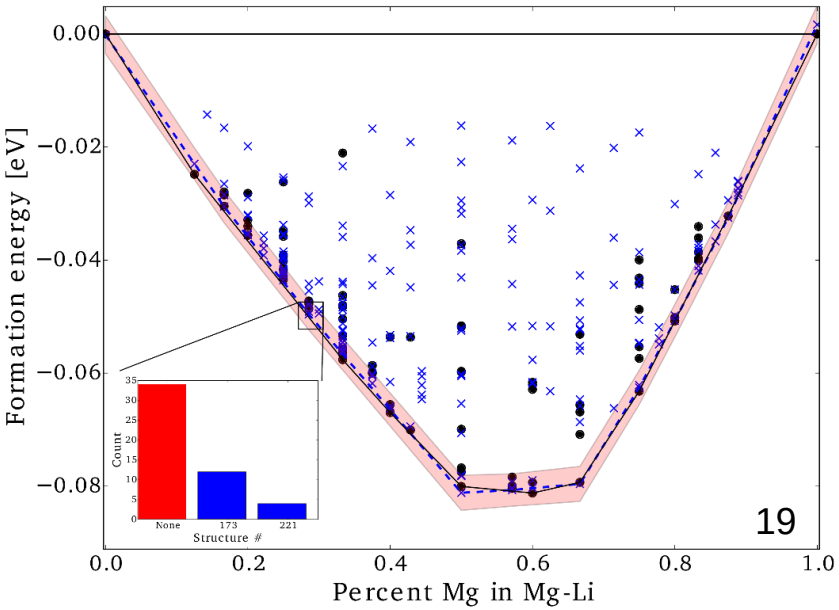
Ground State Predictions

- Fit ECI to formation energies (from DFT)

$$\Delta E(\boldsymbol{\sigma}|\gamma) = \sum_{i=0}^{M} \gamma_i \phi_i(\boldsymbol{\sigma})$$

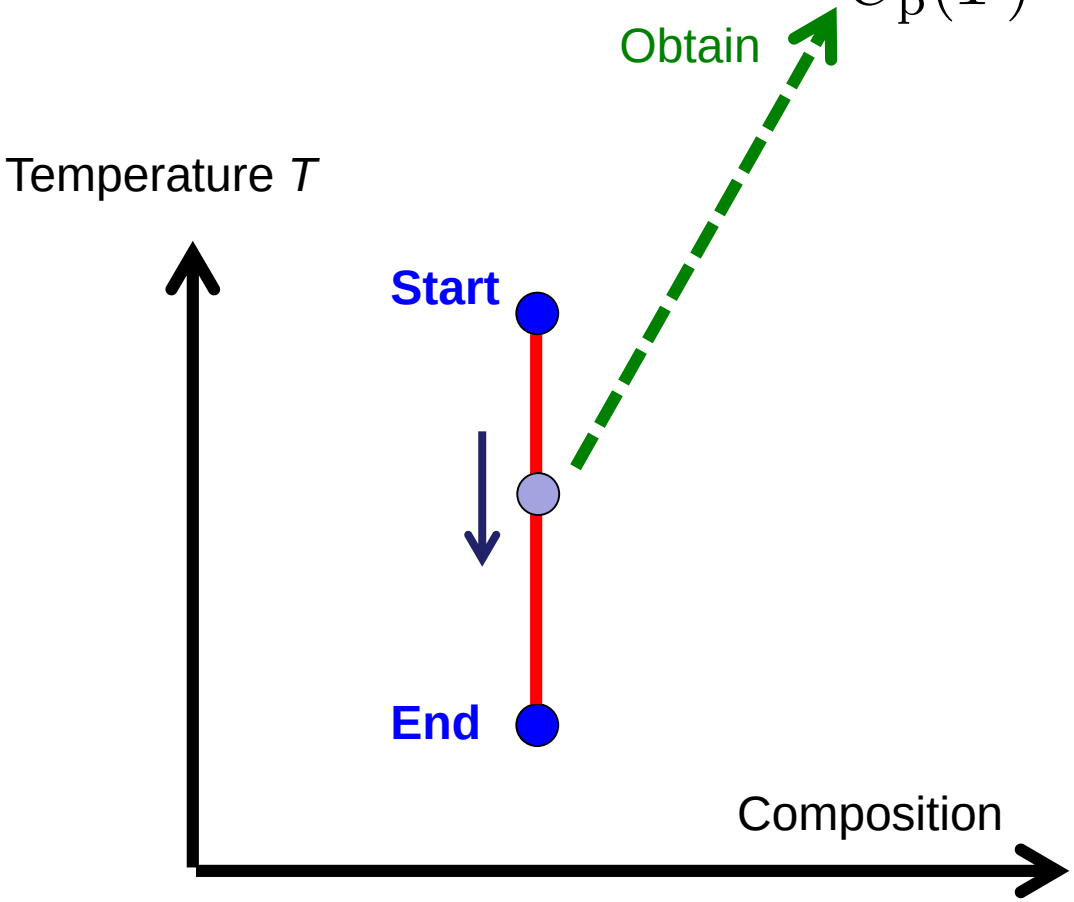
- Training values for the formation energies obtained from DFT simulations
- Training configurations generated by the *maps* tool in ATAT
- *Uncertainty in predicted energy and also structures forming* ***the GSL***

Example: Formation energies of MgLi from DFT (black circles) and CE (blue crosses). Also shown the ground state line from DFT (black) and from CE (**dashed blue** with red confidence interval). Results using up to 5 point clusters. The inset shows uncertainty in the structures forming the GSL. Red bar indicates the DFT prediction



Computing Thermodynamic Quantities

$$C_{\mathrm{p}}(T) = k_{\mathrm{B}}\beta^2\left(\langle E^2\rangle - \langle E\rangle^2\right)$$



To perform average:
Sample a bunch of states with probability:

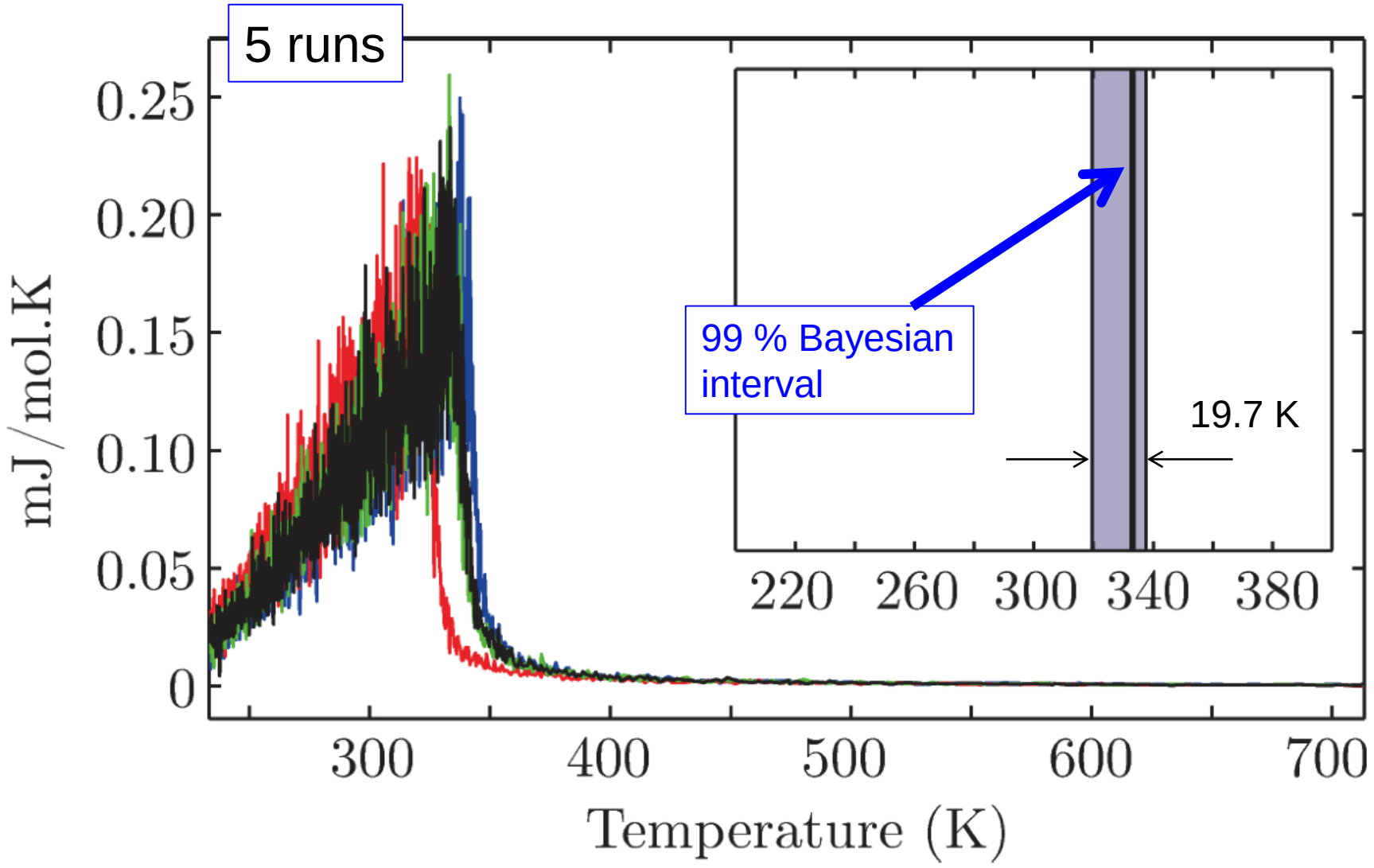
$$\propto \exp(-\beta \hat{E}(\boldsymbol{\sigma};\boldsymbol{\gamma}))$$

Traditional solution:
Traditional Markov Chain Monte Carlo (MCMC).
➔Go to each *T* and thermalize+take *M* samples.

$$\langle Q(T)\rangle = \frac{1}{M}\sum_{j=1}^{M} Q(\boldsymbol{\sigma}^{(j)})$$

A peak (divergence in the limit of an infinite lattice) in the specific heat signals the phase transition.

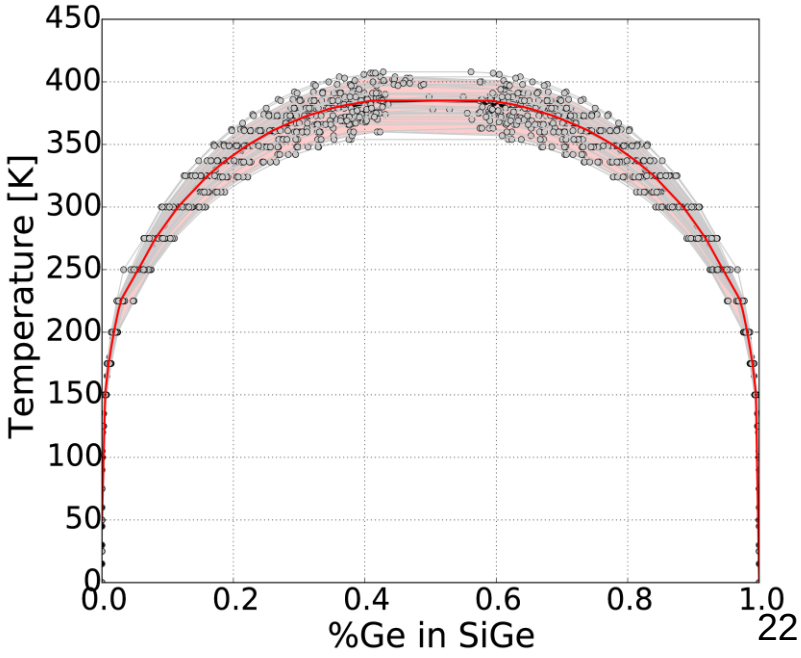
"Error Bars" on Phase Transitions



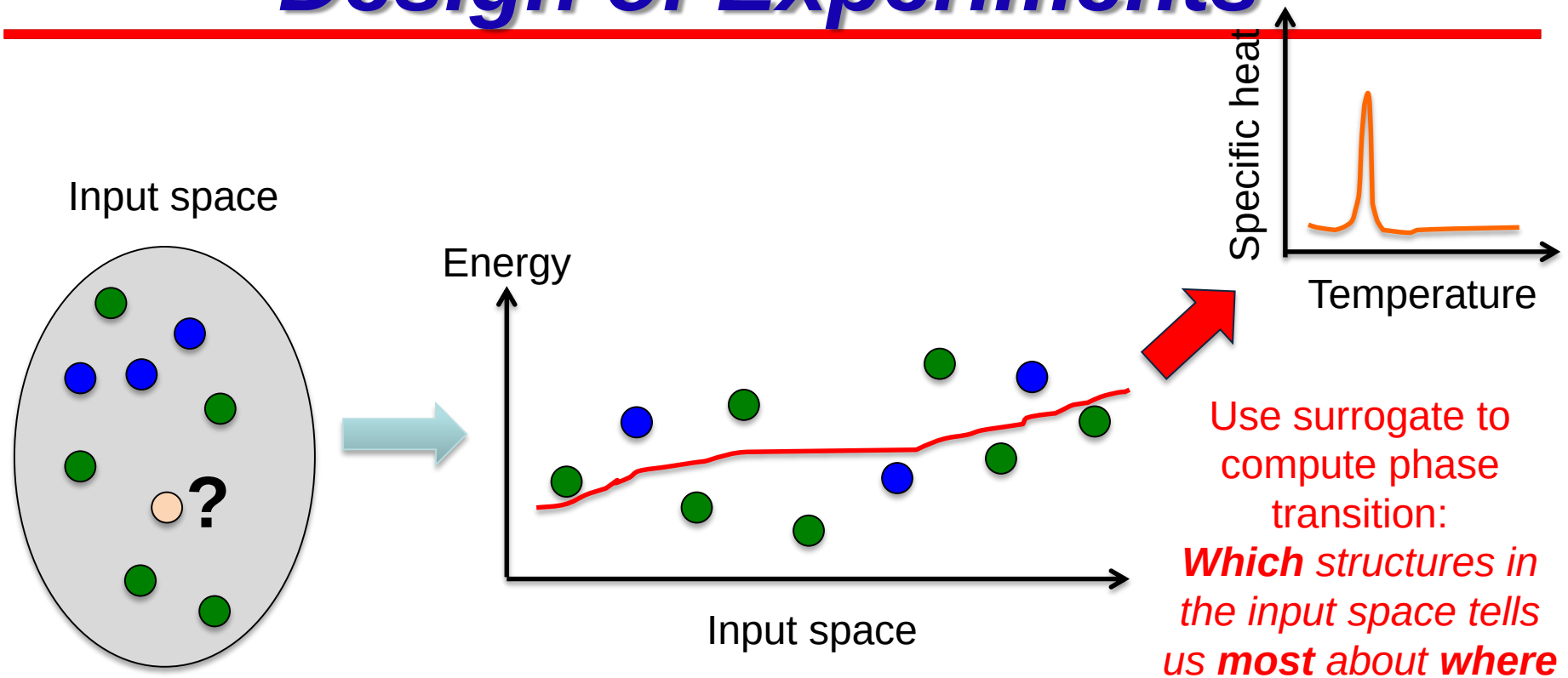
Phase Diagram: SiGe

- ATAT also calculates phase boundaries with using "*phb*"
- ECI sampling → different realizations of the phase diagram
- Uncertainty in the phase diagram using a Monte Carlo approach

Example: Phase diagram of SiGe. **Red line**: median of the simulations with the 95% confidence interval. **Black line** (below red line): result with ATAT's ECI. **Grey lines**: realisations with different ECI



Design of Experiments



Use surrogate to compute phase transition:
***Which** structures in the input space tells us **most** about **where** the transition happens?*

How do we find structures that are maximally informative about the thermodynamic quantity of interest (under a budget)?

Bayesian Design of Experiments

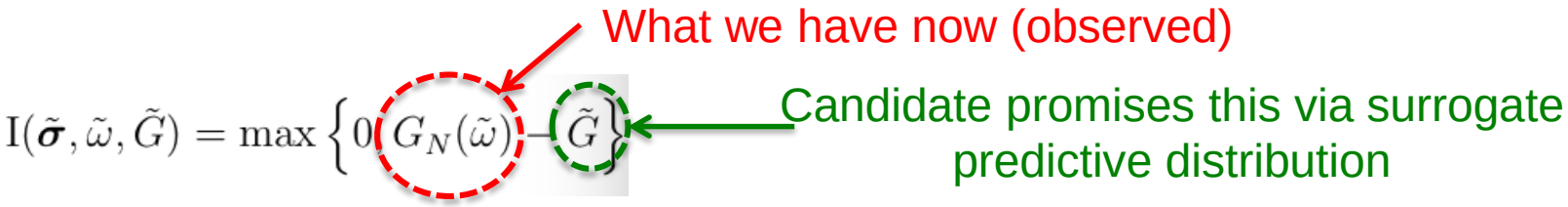
- Denote a set of thermodynamic parameters ω
  - Temperature, pressure, concentration, etc.
- We consider quantities of interest in the form

$$\boldsymbol{\sigma}^*(\omega) = \arg\min_{\boldsymbol{\sigma}} G(\boldsymbol{\sigma}, \omega)$$

  where the arg min must satisfy any ω constraints
- G is the thermodynamic potential whose minimization gives the thermodynamically stable structures at ω
- For ground state structures of binary alloy A_xB_{1-x} (phase diagram at T=0), $\omega = \{x\}$ where x is the concentration

Bayesian Design of Experiments

- Consider *the improvement* offered by a candidate structure:



- But *G*-tilde is hypothetical: it is predicted by the CE surrogate (thus function of γ), so we don't know which *G*-tilde to use
  - Therefore, take the expectation of *I*(…) to integrate out this lack of knowledge:

$$\mathrm{EI}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega}) = \mathbb{E}\left[\mathrm{I}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega}, \tilde{G}) | \tilde{\boldsymbol{\sigma}}, \tilde{\omega}, \mathcal{D}_N\right]$$

Expectation is over all possible values of the surrogate $\tilde{G}$ when predicting on structure $\tilde{\sigma}$

Bayesian Design of Experiments

- The expectation* is computed analytically based on the Gaussianity of the Bayesian predictive distribution:

$$\mathrm{EI}(\tilde{\boldsymbol{\sigma}},\tilde{\omega}) = [G_N(\tilde{\omega}) - \mu_{G,N}(\tilde{\boldsymbol{\sigma}},\tilde{\omega})]\,\Psi\left(\frac{G_N(\tilde{\omega})-\mu_{G,N}(\tilde{\boldsymbol{\sigma}},\tilde{\omega})}{\nu_{G,N}(\tilde{\boldsymbol{\sigma}},\tilde{\omega})}\right) + \nu_{G,N}(\tilde{\boldsymbol{\sigma}},\tilde{\omega})\,\psi\left(\frac{G_N(\tilde{\omega})-\mu_{G,N}(\tilde{\boldsymbol{\sigma}},\tilde{\omega})}{\nu_{G,N}(\tilde{\boldsymbol{\sigma}},\tilde{\omega})}\right),$$

  - Maximization is carried out on a pre-computed grid in this work (so we select the grid point with largest EI).
  - Generally one can use a 1D maximization routine (e.g. MCMC)

- Then choose the next structure $\boldsymbol{\sigma}^{(N+1)}$ which satisfies

$$\left(\boldsymbol{\sigma}^{(N+1)},\omega^{(N+1)}\right) = \underset{\tilde{\boldsymbol{\sigma}},\tilde{\omega}}{\mathrm{argmax}}\,\mathrm{EI}(\tilde{\boldsymbol{\sigma}},\tilde{\omega}).$$

… and run the expensive simulator on that structure

**Schonlau, Matthias, William J. Welch, and D. R. Jones (Phys. & Eng. Sciences, Am. Stat. Assoc., 1996)*

From surrogate model at iteration *N*
(surrogate changes each iteration)

Algorithm 1: EI structure acquisition strategy for learning the thermodynamic potential*

Require:

$\mathcal{D}_{N_0}$ (an initial pool of N_0 observed $\boldsymbol{\sigma}$-ω-G triples),

$N_{\max}$ (maximum number of observations that can be afforded),

ϵ (maximum EI tolerance),

$\mathcal{S}_{N_{\text{pool}}}$ (large pool of $\boldsymbol{\sigma}$-ω pairs to select simulations from)

$N \leftarrow N_0$

$\mathcal{D}_N = \mathcal{D}_{N_0}$

repeat

  Find:

$$\left(\boldsymbol{\sigma}^{(N+1)}, \omega^{(N+1)}\right) = \underset{(\tilde{\boldsymbol{\sigma}},\tilde{\omega}) \in \mathcal{S}_{N_{\text{pool}}}}{\text{argmax}} \ \text{EI}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega})$$

  if $\text{EI}\left(\boldsymbol{\sigma}^{(N+1)}, \omega^{(N+1)}\right) < \epsilon$ **then**

    Break loop

  end

  $G^{(N+1)} \leftarrow G\left(\boldsymbol{\sigma}^{(N+1)}, \omega^{(N+1)}\right)$

  $\mathcal{D}_{N+1} \leftarrow \mathcal{D}_N \cup \left\{\left(\boldsymbol{\sigma}^{(N+1)}, \omega^{(N+1)}, G^{(N+1)}\right)\right\}$

  $N \leftarrow N + 1$

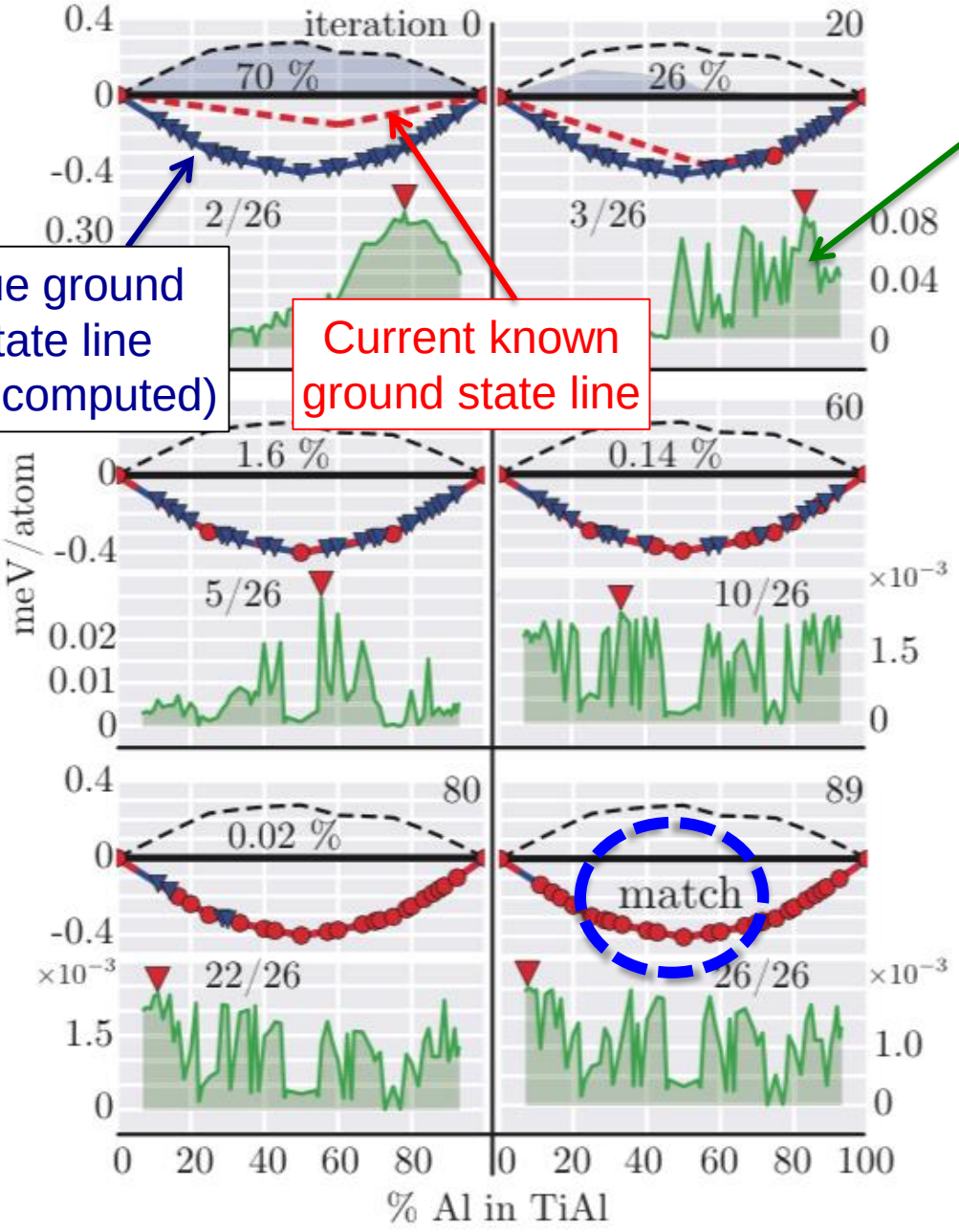
until $N >= N_{max}$;

$$\text{EI}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega}) = \left[G_N(\tilde{\omega}) - \mu_{G,N}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega})\right] \Psi\left(\frac{G_N(\tilde{\omega}) - \mu_{G,N}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega})}{\nu_{G,N}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega})}\right) + \nu_{G,N}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega}) \psi\left(\frac{G_N(\tilde{\omega}) - \mu_{G,N}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega})}{\nu_{G,N}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega})}\right),$$

- We need knowledge of *G* because we need to know the current lowest estimate of the thermodynamic potential function.
- G_N is lowest observed value of thermodynamic potential function at iteration *N*

- Keep selecting structures (expanding the design) based on thermodynamic expected improvement until termination conditions are met

**Kristensen, J., Bilionis, I., Zabaras, N. (JCP, submitted, 2015)*



True ground state line (pre-computed)

Current known ground state line

Thermodynamic expected improvement

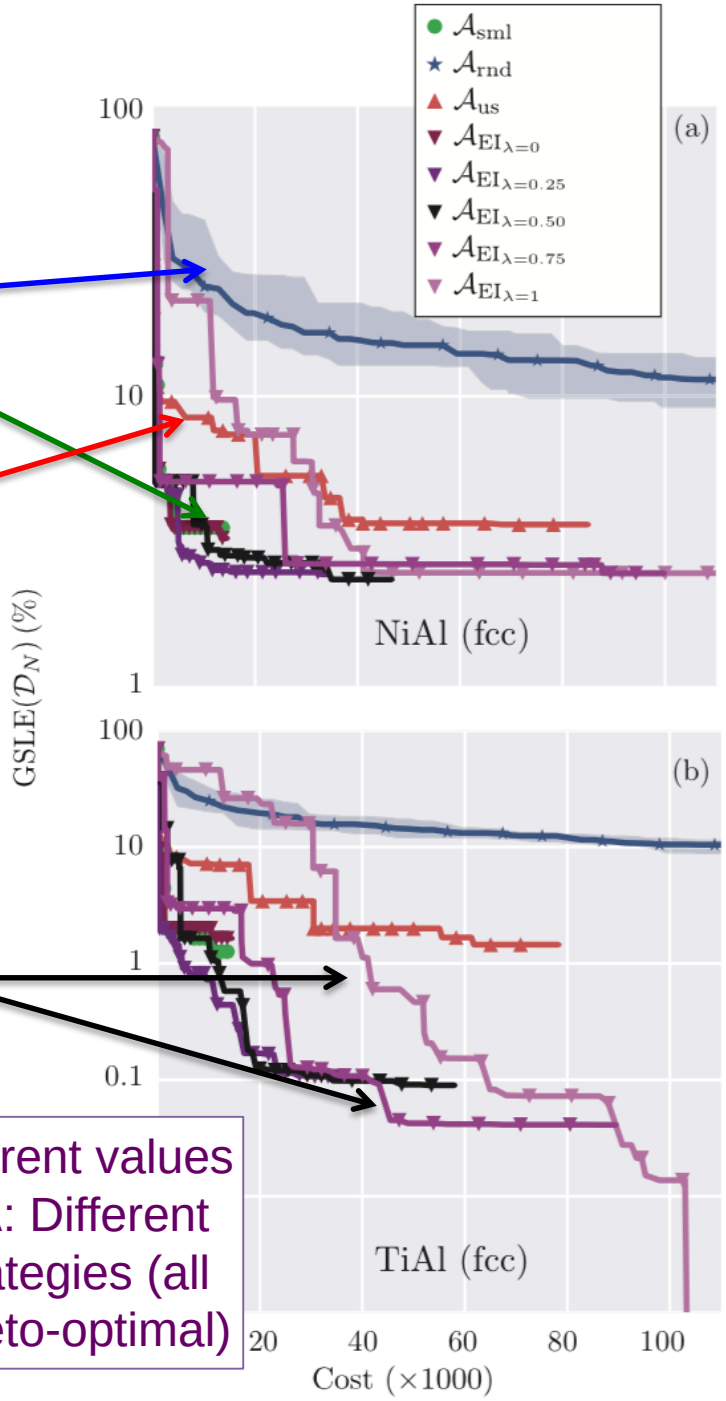
EAM potential

TiAl: Find the ground state line among ~35,000 structures

- We start with 4 structures
- We find the **exact** ground state line with only 89 total input structures!

J. Kristensen, I. Bilionis, N. Zabaras, JCP, 2016

1. Compare to other structure acquisition strategies:
+ Random selection (rnd)
+ Choose smallest structures (sml)
+ Choose the structure which has the largest predictive variance (uncertainty sampling: us)



2. We further develop our method to account for the cost of computing structures (important for DFT, e.g.):

$$\mathrm{EI}_\lambda \equiv \lambda \mathrm{EI}(\tilde{\boldsymbol{\sigma}}, \tilde{\omega}) - (1-\lambda)\mathrm{C}_{G(\tilde{\boldsymbol{\sigma}},\tilde{\omega})}$$

Let $N_{j,\mathrm{at}}$ be the number of atoms per configurational unit cell of structure j, $\boldsymbol{\sigma}^{(j)}$, then the cost is:

$$C_{G(\tilde{\boldsymbol{\sigma}}^{(j)},\tilde{\omega})} = N_{j,\mathrm{at}}^3$$

(same cost exponent as DFT)

Different values of λ: Different strategies (all Pareto-optimal)

Stochastic Coarse Graining

Models in statistical equilibrium following a PDF in fine- and coarse-scale:

Fine-scale (f) model

$$p_{\mathrm{f}}(\mathbf{x}|\beta,\mu) = \frac{\exp\{-\beta U(\mathbf{x},\mu)\}}{Z(\beta,\mu)}$$

Coarse-scale (c) model

$$p_{\mathrm{c}}(\mathbf{X}|\beta,\mu) = \frac{\exp\{-\beta U_{\mathrm{c}}(\mathbf{X},\boldsymbol{\theta}_{\mathrm{c}},\mu)\}}{Z_{\mathrm{c}}(\boldsymbol{\theta}_{\mathrm{c}},\beta,\mu)}$$

- Fine variables $\mathbf{x} \in \mathcal{M}$, $\mathcal{M} \subset \mathbb{R}^n$ with $n \gg 1$ degrees of freedom.
- Coarse variables $\mathbf{X} \in \mathcal{M}_{\mathrm{CG}}$, $\mathcal{M}_{\mathrm{CG}} \subset \mathbb{R}^{n_{\mathrm{c}}}$ with $n_{\mathrm{c}} \ll n$ degrees of freedom.
- Partition function $Z(\beta,\mu)$ and $Z_{\mathrm{c}}(\boldsymbol{\theta}_{\mathrm{c}},\beta,\mu)$ in fine- and coarse-scale.
- Potential $U(\mathbf{x},\mu)$ and $U_{\mathrm{c}}(\mathbf{X},\boldsymbol{\theta}_{\mathrm{c}},\mu)$ in fine- and coarse-scale. The dependency on the external field μ and β is dropped due to notational reasons.
- $\boldsymbol{\theta}_{\mathrm{c}}$ parametrization of coarse-grained potential U_{c}.
- $\beta \propto 1/T$, with temperature T.

Stochastic Coarse Graining

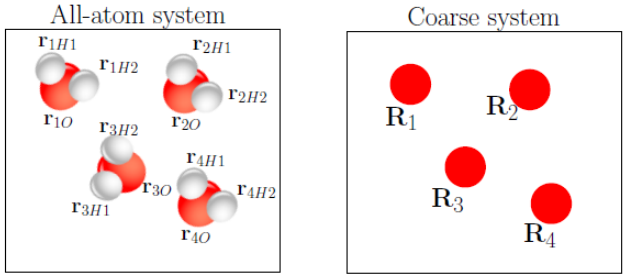
- Minimize information loss of approximating $p_f(\mathbf{x})$ by $p_c(\mathbf{X}|\theta_c)$.
- Using relative entropy to quantify the distance between distributions.
- Which fine variables $\mathbf{x}$ correspond to one coarse variable $\mathbf{X}$?
  - Deterministic mapping function $\xi(\mathbf{x})$ $(\xi : \mathcal{M} \longrightarrow \mathcal{M}_{CG})$

$$\mathbf{X} = \xi(\mathbf{x})$$

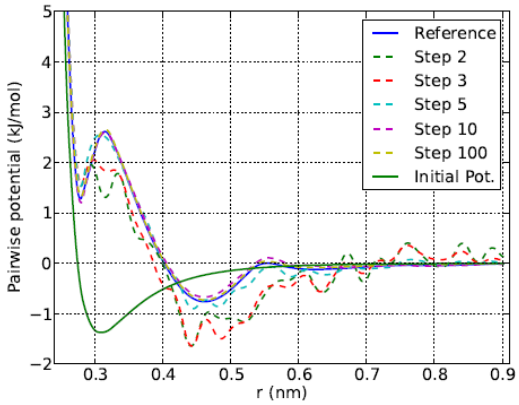
Relative Entropy

$$\mathcal{S}_{rel}[\theta_c] := \int_{\mathcal{M}} p_f(\mathbf{x}|\beta) \log\left(\frac{p_f(\mathbf{x}|\beta)}{p_c(\xi(\mathbf{x})|\theta_c, \beta)}\right) d\mathbf{x} + \mathcal{S}_{map}$$

- $\mathcal{S}_{map} = \int_{\mathcal{M}} \delta(\xi(\mathbf{x}) - \mathbf{X}) d\mathbf{x}$



$$\mathbf{R}_i = \frac{m_H \mathbf{r}_{iH1} + m_H \mathbf{r}_{iH2} + m_O \mathbf{r}_{iO}}{2m_H + m_O}$$



Hard Optimization

- Fine Simulations
- CG Simulations
- Noisy gradients

- *I. Bilionis & N. Zabaras, Journal of Chemical Physics, 138, 044313, 2013.*
- *Markus Schöberl, P.S. Koutsourelakis, N. Zabaras, JCP, 2016.*

Stochastic Coarse Graining

Framework following **generative probabilistic models** and builds upon two parts:

- Coarse-scale PDF $p_c(\mathbf{X}|\theta_c)$
  - parametrized by θ_c
- Probabilistic mapping from *coarse to fine* $p_{cf}(\mathbf{x}|\mathbf{X}, \theta_{cf})$
  - conditional PDF of $\mathbf{x}$ given coarse variable $\mathbf{X}$
  - CG variables $\mathbf{X}$ do not uniquely define the fine variables $\mathbf{x} \rightarrow$ probabilistic relation necessary
  - parametrization of the probabilistic mapping by θ_{cf}

- Possible to add prior information expressed by $p(\theta_c)$ and $p(\theta_{cf})$.

Simulation of Fine Scale System

Training the model by observations $\mathbf{x}^{(1:N)}$ one obtains optimal model parameters θ_c^* and θ_{cf}^* for the given data set $\mathbf{x}^{(1:N)}$.
The posterior distribution leads to the predictive distribution for fine-scale configurations $\mathbf{x}$:

Fine-scale predictive PDF $\tilde{p}$

$$\tilde{p}(\mathbf{x}|\theta_c^*, \theta_{cf}^*) = \int_{\mathcal{M}_{CG}} p_{cf}(\mathbf{x}|\mathbf{X}, \theta_{cf}^*) p_c(\mathbf{X}|\theta_c^*) d\mathbf{X}$$

Simulate fine-scale system

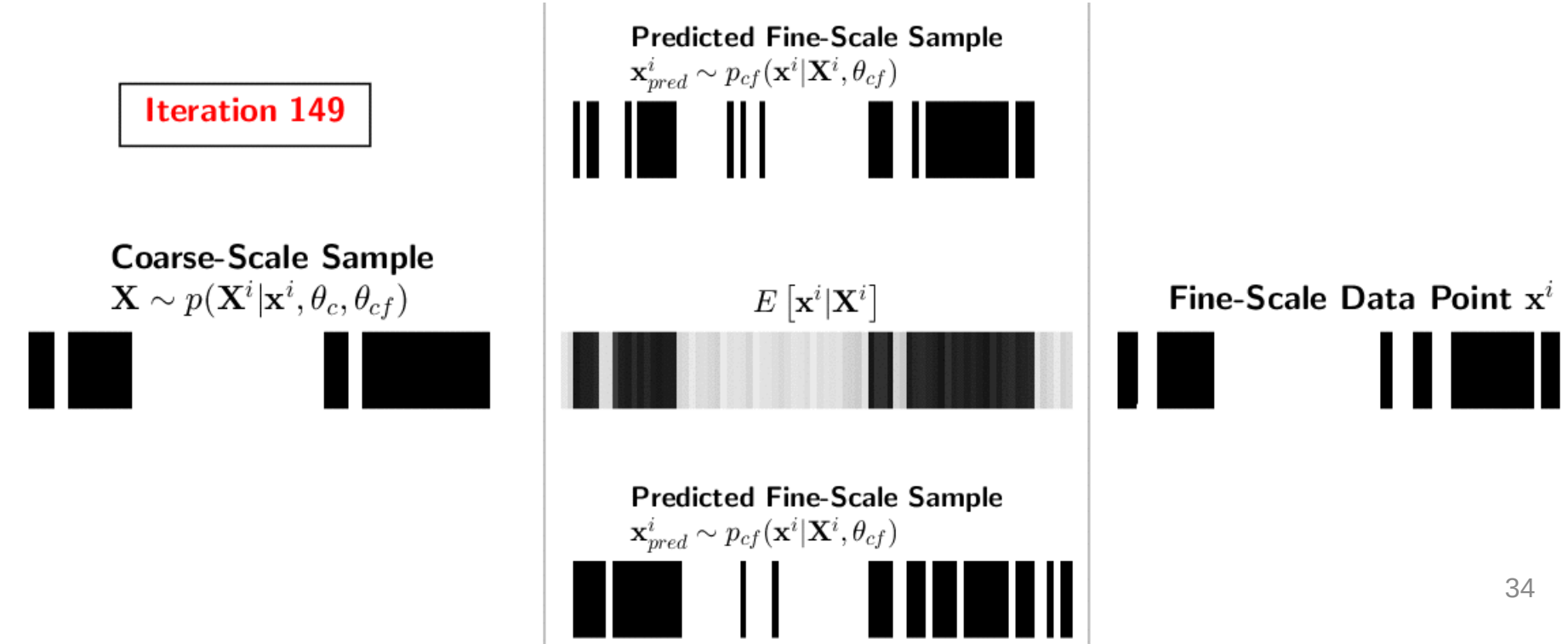
1. θ_c^* and θ_{cf}^* optimal for given data $\mathbf{x}^{(1:N)}$
2. Draw sample of coarse-scale description $\mathbf{X}^* \sim p_c(\mathbf{X}|\theta_c^*)$
3. Draw sample of fine-scale description $\mathbf{x}^* \sim p_{cf}(\mathbf{x}^*|\mathbf{X}^*, \theta_{cf}^*)$

Ensemble Averages

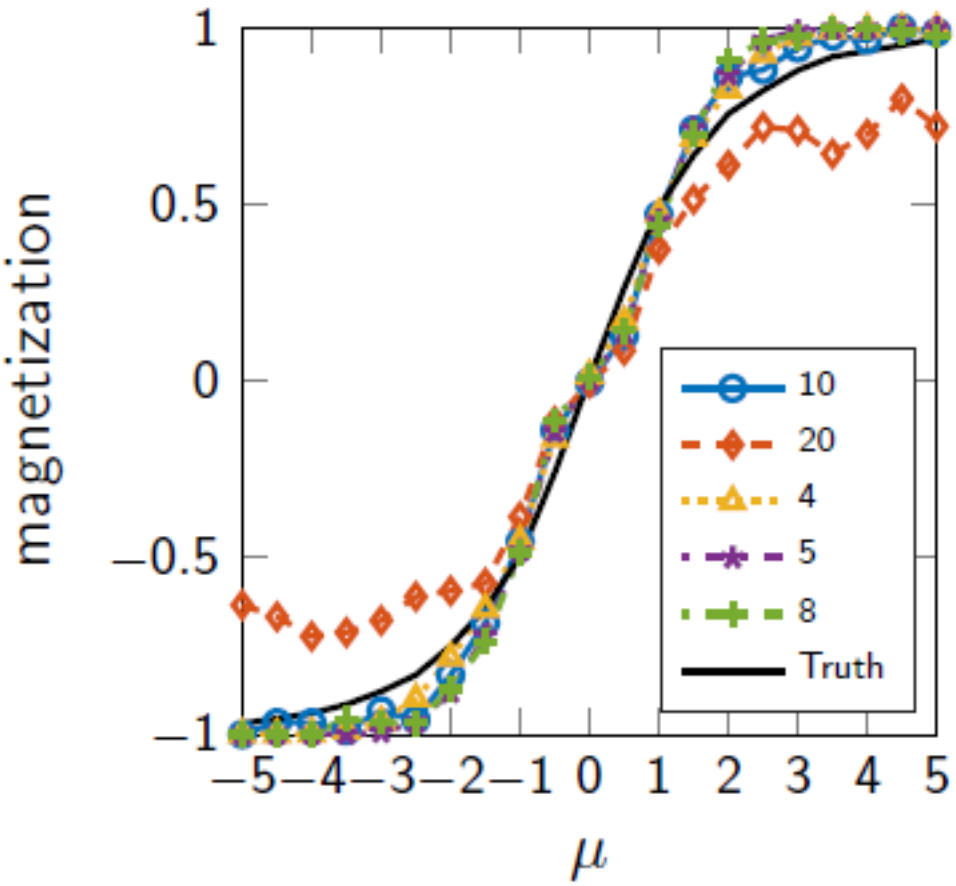
$$\langle f(\mathbf{x})\rangle = \int f(\mathbf{x})p_{\mathrm{f}}(\mathbf{x})d\mathbf{x}$$

$$\approx \int f(\mathbf{x})\tilde{p}(\mathbf{x}|\boldsymbol{\theta}^*_{\mathrm{c}}, \boldsymbol{\theta}^*_{\mathrm{cf}})d\mathbf{x}$$

$$= \int\int f(\mathbf{x})p_{\mathrm{cf}}(\mathbf{x}|\mathbf{X}, \boldsymbol{\theta}^*_{\mathrm{cf}})p_{\mathrm{c}}(\mathbf{X}|\boldsymbol{\theta}^*_{\mathrm{c}})d\mathbf{X}d\mathbf{x}$$



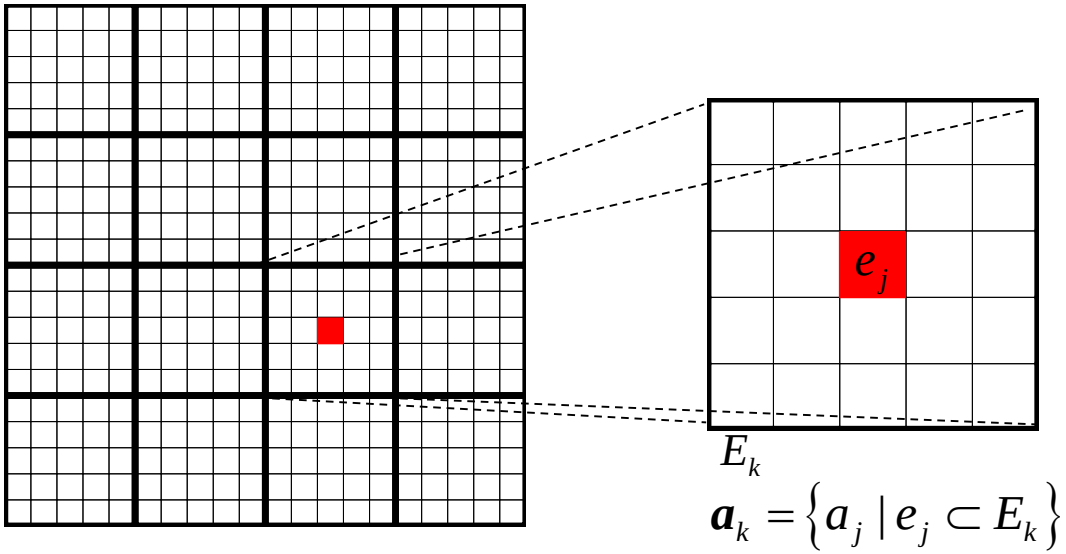
Ensemble Averages



- *Markus Schöberl, P.S. Koutsourelakis, N. Zabaras, JCP, 2016.*

Scalable UQ for Multiscale Problems

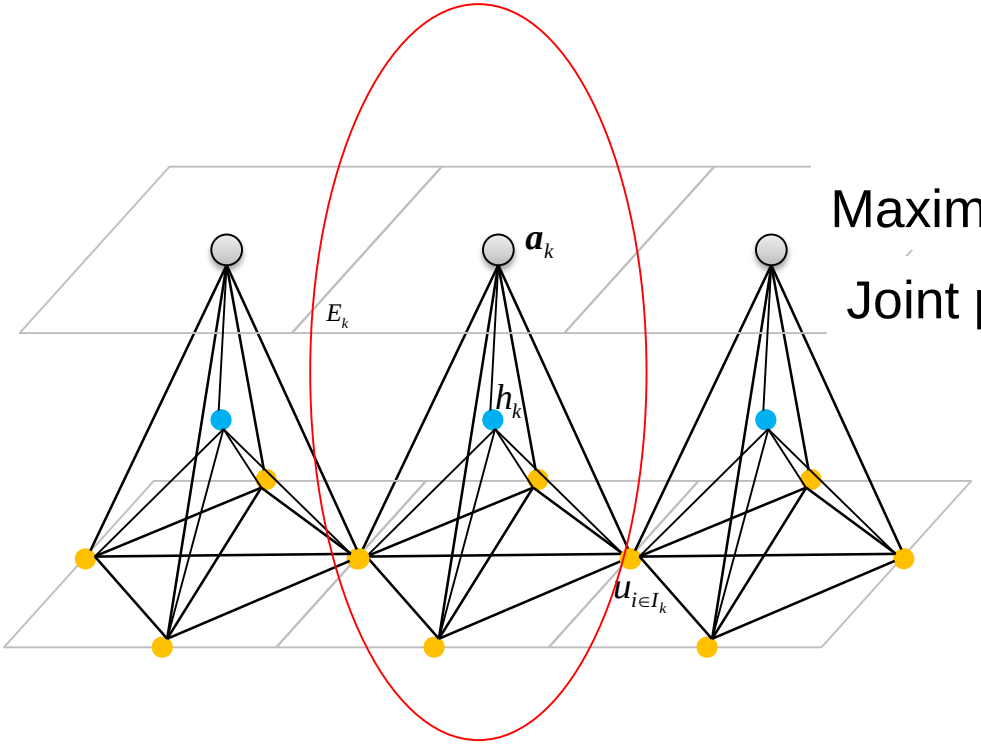
- Due to the multiscale features, the spatial domain *D* is discretized into
  - fine grid $T_h = \bigcup_{i=1}^{N_h} e_i$ where N_h is the number of fine elements.
  - coarse grid $T_c = \bigcup_{i=1}^{N_c} E_i$ where N_c is the number of coarse elements.
- Stochastic input
  - assume a constant property on each fine element, i.e. $\boldsymbol{a} = \left(a_1 \cdots\cdots a_{N_h}\right)$
  - local property on coarse element E_k: $\boldsymbol{a}_k = \left\{a_j \mid e_j \subset E_k\right\}$ such that $\boldsymbol{a}_k \subset \boldsymbol{a}$



Probabilistic Modeling of SPDEs

Assumption

- Each response u_i , h_k is only correlated to its neighboring nodes (including local feature $\boldsymbol{a}_k$) within the same coarse element.
- Long distance interactions among variables are ignored.



Maximal clique : potential $q_k\left(u_{I_k}, h_k; \boldsymbol{a}_k\right)$

Joint probability $p\left(\boldsymbol{u}, \boldsymbol{h} \mid \boldsymbol{a}\right) \propto \prod_k q_k\left(u_{I_k}, h_k; \boldsymbol{a}_k\right)$

$$q_k\left(u_{I_k}, h_k; \boldsymbol{a}_k\right) = \exp\left(-\mathcal{E}_k\left(u_{I_k}, h_k; \boldsymbol{a}_k\right)\right)$$

$$p\left(\boldsymbol{u}, \boldsymbol{h} \mid \boldsymbol{a}\right) \propto \exp\left(-\sum_k \mathcal{E}_k\left(u_{I_k}, h_k; \boldsymbol{a}_k\right)\right)$$

Probabilistic Modeling of SPDEs

- Apply the definition of energy function in these sub-problems, the local energy functions are expressed by

$$\mathcal{E}_k\left(u_{I_k}, h_k; \boldsymbol{a}_k\right) \approx \sum_{i\in I_k} \phi_{k,i}\left(u_i, \boldsymbol{a}_k\right) + \sum_{(i,j)\in I_k\times I_k, i\neq j} \phi_{k,ij}\left(u_i, u_j, \boldsymbol{a}_k\right) + \phi_{k,0}\left(h_k, \boldsymbol{a}_k\right) + \sum_{i\in I_k} \phi_{k,i0}\left(u_i, h_k, \boldsymbol{a}_k\right)$$

where

$$\phi_{k,i}\left(u_i, \boldsymbol{a}_k\right) = f_{k,i}\left(\boldsymbol{a}_k\right)u_i + f_{k,ii}\left(\boldsymbol{a}_k\right)u_i^2 \;,\; \phi_{k,ij}\left(u_i, u_j, \boldsymbol{a}_k\right) = f_{k,ij}\left(\boldsymbol{a}_k\right)u_i u_j$$

$$\phi_{k,0}\left(h_k, \boldsymbol{a}_k\right) = f_{k,0}\left(\boldsymbol{a}_k\right)h_k + f_{k,00}\left(\boldsymbol{a}_k\right)h_k^2 \;,\; \phi_{k,i0}\left(u_i, h_k, \boldsymbol{a}_k\right) = f_{k,i0}\left(\boldsymbol{a}_k\right)u_i h_k$$

- Since the functions of local features $\boldsymbol{a}_k$ in the energy functions are unknown, a nonparametric model is adopted

$$f_{k,\cdot}\left(\boldsymbol{a}_k\right) \equiv f_{k,\cdot}\left(\boldsymbol{a}_k; \boldsymbol{\theta}_k\right) = \theta_{k,\cdot}^{(1)} + \sum_{t=2}^{r} \theta_{k,\cdot}^{(t)} \zeta_t\left(\boldsymbol{a}_k\right)$$

with unnormalized Gaussian kernels

$$\zeta_t\left(\boldsymbol{a}_k\right) = \exp\left(-\frac{\left\|\boldsymbol{a}_k - \bar{\boldsymbol{a}}_t\right\|^2}{\sigma_\zeta^2}\right)$$

Probabilistic Modeling of SPDEs

- The complete probabilistic model for $p(\boldsymbol{u},\boldsymbol{h}|\boldsymbol{a})$ is factorized as a product of potential functions measuring the interactions between random variables

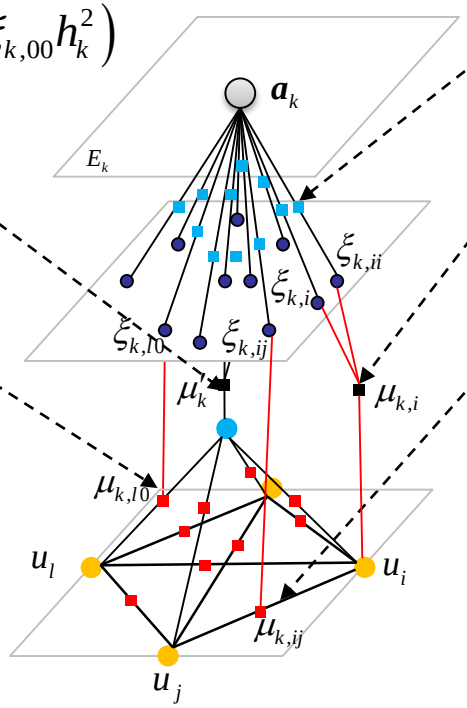
$$\mu_{k,0}\left(h_k,\xi_{k,0},\xi_{k,00}\right)=\exp\left(-\xi_{k,0}h_k-\xi_{k,00}h_k^2\right)$$

$$p\left(\xi_{k,ii}\,|\,\boldsymbol{a}_k\right)=\delta_{\xi}\left(f_{k,ii}\left(\boldsymbol{a}_k\right)\right)$$

$$\mu_{k,l}\left(h_k,u_l,\xi_{k,l0}\right)=\exp\left(-\xi_{k,l0}u_l h_k\right)$$

$$\mu_{k,i}\left(u_i,\xi_{k,i},\xi_{k,ii}\right)=\exp\left(-\xi_{k,i}u_i-\xi_{k,ii}u_i^2\right)$$

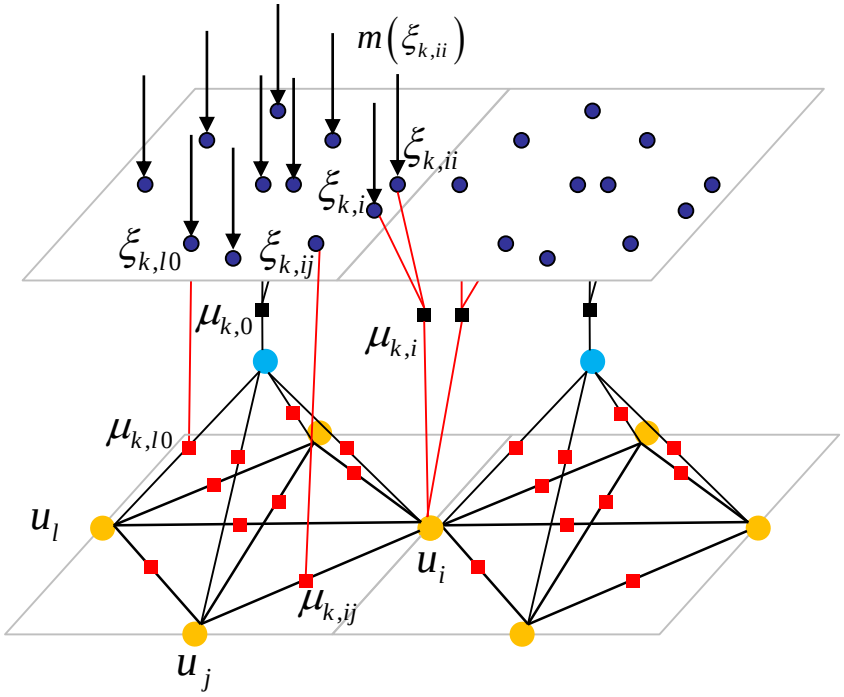
$$\mu_{k,ij}\left(u_i,u_j,\xi_{k,ij}\right)=\exp\left(-\xi_{k,ij}u_i u_j\right)$$



Inference in Graphs

- A challenge arises in the update of messages between hidden variables. Although analytic expressions of $p(\boldsymbol{a})$ and $p(\xi|\boldsymbol{a})$ are explicit, the joint distribution of hidden variables ξ could be complicated such that the links between them are implicit when stochastic input has been removed from the graph.

- To bypass the difficulties in passing messages between hidden variables, the graphical model is transformed as follows:



Messages in Belief Propagation

- For all messages except those between hidden variables, since there is no prior information, they are represented non parametrically (as weighted Gaussian mixtures)

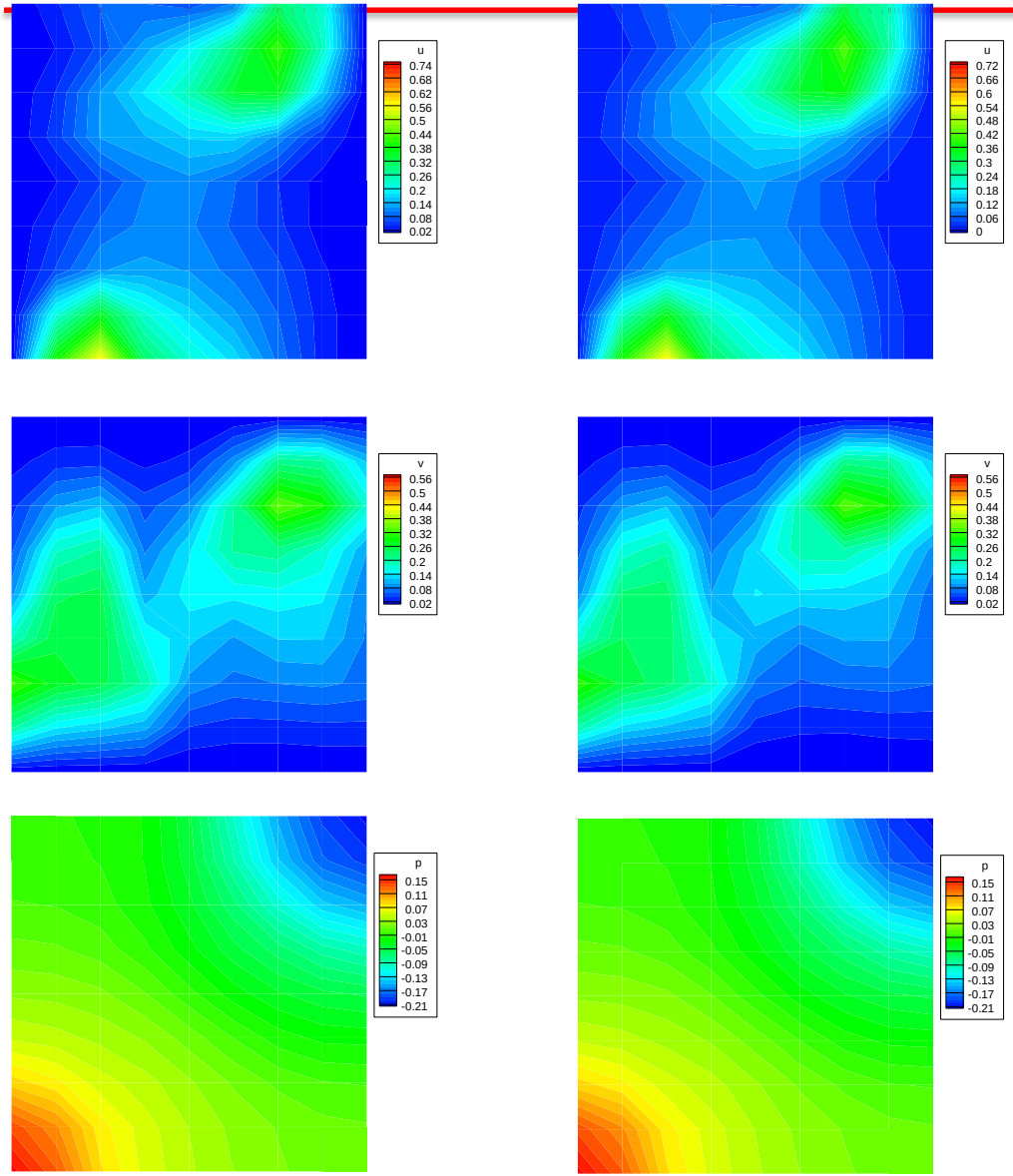
- Without loss of generality, consider the message from factor node $\mu_{k,ij}\left(u_i, u_j, \xi_{k,ij}\right)$ to variable node u_i

$$m_{\mu_{k,ij}\to u_i}\left(u_i\right) \approx \sum_{t=1}^{T} l_t \mathcal{N}\left(u_i; \bar{u}_i^t, \sigma_i^2\right)$$

- At iteration *n* of the BP algorithm, the messages between factor nodes and variables are updated by

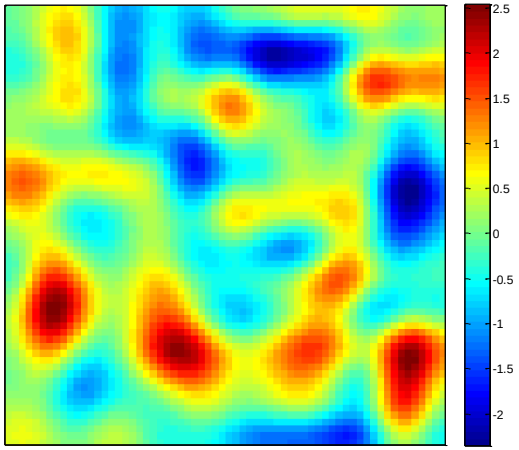
$$m_{\mu_{k,ij}\to u_i}^{(n)}\left(u_i\right) \leftarrow \int \mu_{k,ij}\left(u_i, u_j, \xi_{k,ij}\right) m_{u_j\to\mu_{k,ij}}^{(n)}\left(u_j\right) m_{\xi_{k,ij}\to\mu_{k,ij}}^{(n)}\left(\xi_{k,ij}\right) d\xi_{k,ij} du_j$$

Stochastic Multiscale Models: A Graph Theoretic Approach



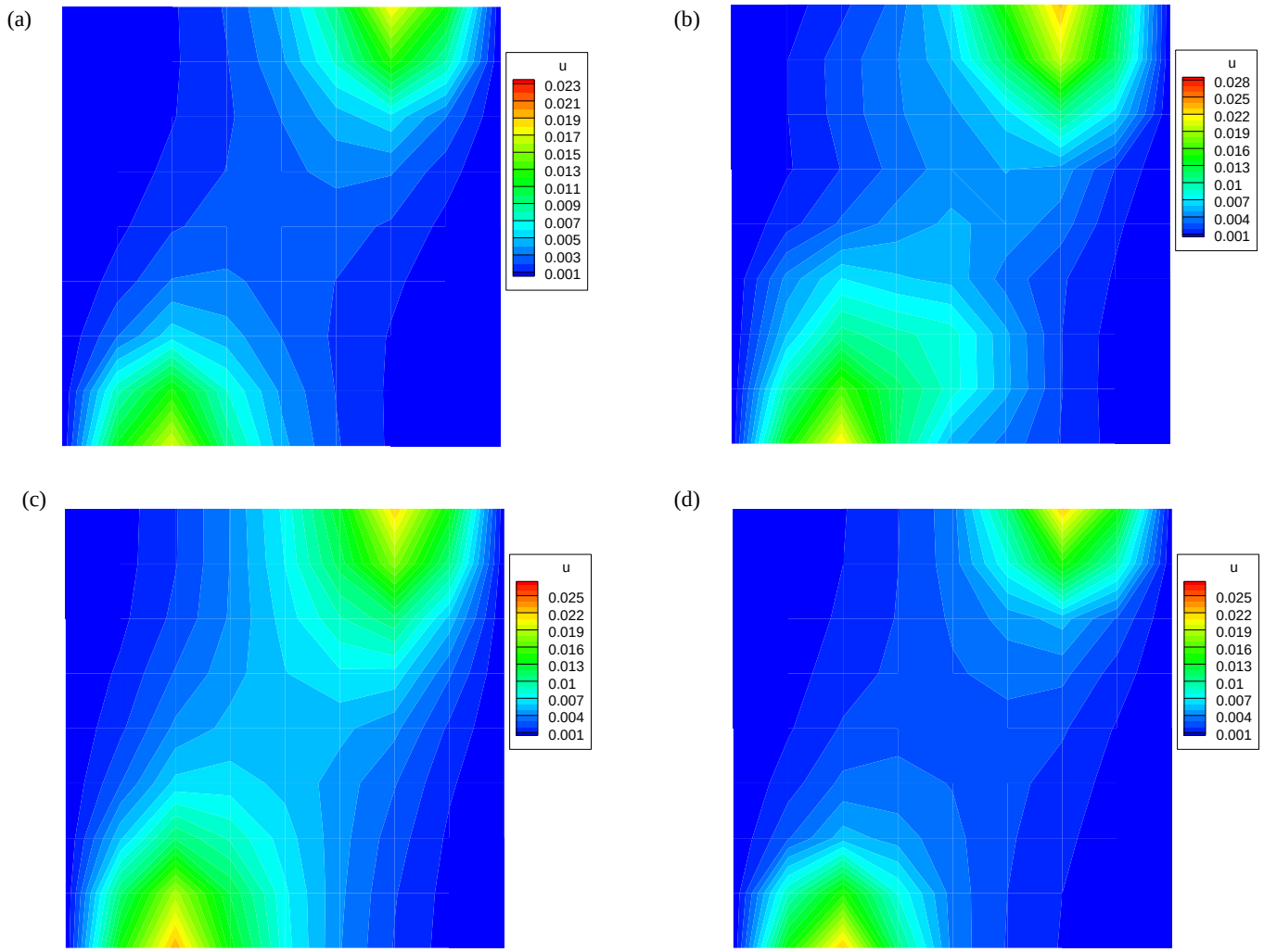
$$\mathrm{Cov}(\boldsymbol{x},\boldsymbol{y}) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|}{L_1} - \frac{|x_2 - y_2|}{L_2}\right)$$

σ=1.0, L_1=0.1, L_2=0.1



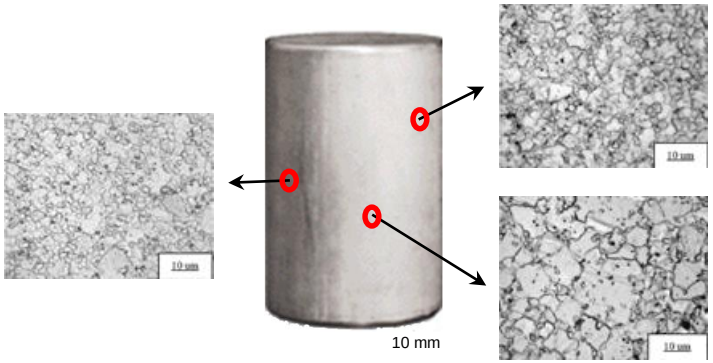
Predicted physical responses given a realization of stochastic input (a)-(c) x-velocity, y-velocity and pressure obtained from direct simulation, and (d)-(f) x-velocity, y-velocity and pressure predicted by the probabilistic graphical model (trained with 60 data points)

Predicted Variance



Predicted variance of x-velocity (a) MC simulation with 10^6 samples, and probabilistic model trained by (b) 20, (c) 40, (d) 60 data.

Multiscale Disk Forging: Microstructure Uncertainty

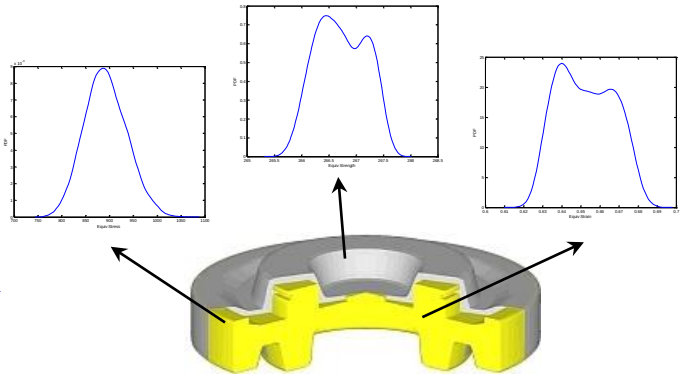


Location-specific microstructures

Mesoscale uncertainties

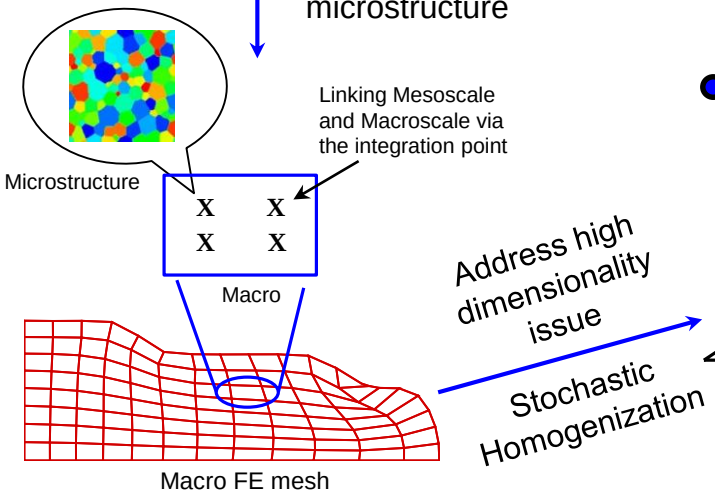
Multiscale Forging simulation

[Due to symmetry, 2D forging problem is actually solved]



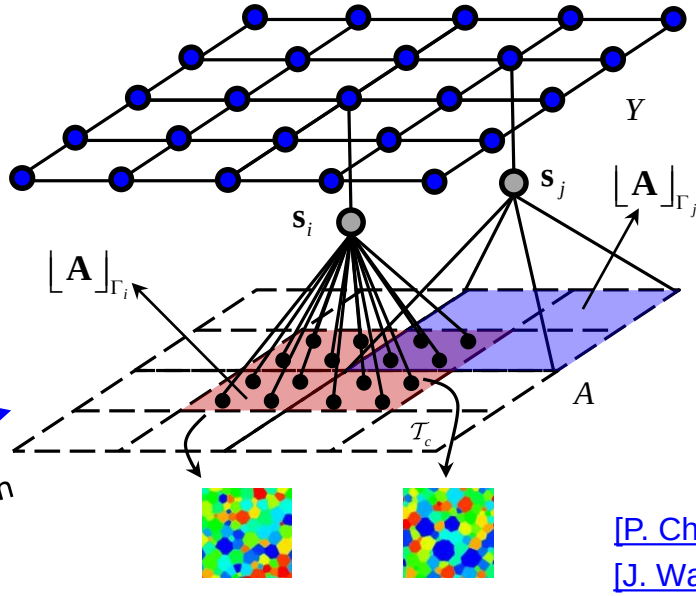
Macroscale uncertainties

Finite discretization of microstructure



Address high dimensionality issue

Stochastic Homogenization



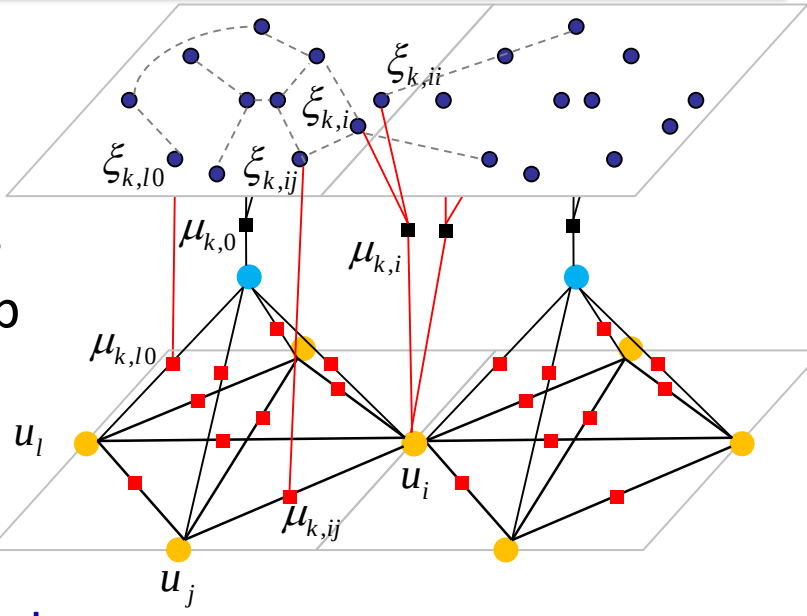
Uncertainty propagation problem addressed by solving an inference problem on the graph

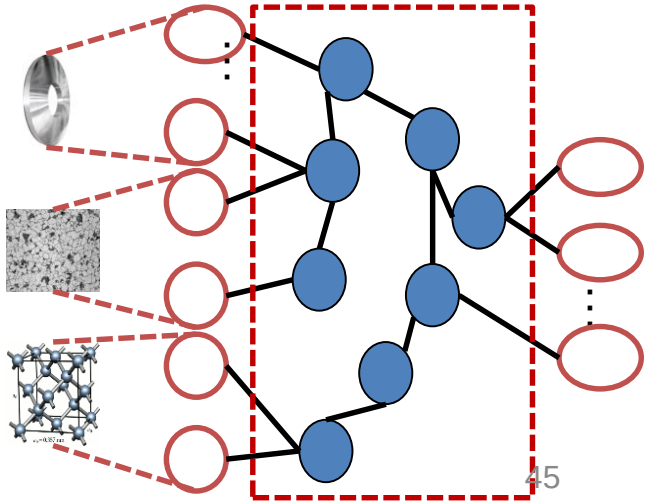
[P. Chen & N. Zabaras, JCP, 2014]

[J. Wan & N. Zabaras, JCP, 2014]

Stochastic Multiscale Models: A Graph Theoretic Approach

- Pose multiscale SPDEs in graphs
- Factorize conditional PDF of responses using `clique' potentials.
- Introduce hidden variables to account for coarse graining – naturally leads to a deep learning machine.
- Fully non-parametric approach
- All parameters are learned with local inference (EM, SMC, Variational,..)
- Conditional & marginal PDFs are computed with approximate inference (e.g. EP).
- The probabilistic graphical model can be used for multiple UQ tasks: E.g. as a surrogate model and for inverse problem solution!
- Data and models become one and the same!





- J. Wan and N. Zabaras, JCP, 2014
- P. Chen and N. Zabaras, JCP, 2014,2015

ITS ALL ABOUT DATA

- Experimental ***and*** simulation data used to design predictive models for quantities of interest
  - E.g. Quantifying *uncertainty in the Exchange Correlation functional and propagating it to QoI*
  - Designing predictive surrogate models for materials design: quantifying *epistemic uncertainty from limited training data*
  - Computing *the most informative simulations* to collect *online data* for training a Bayesian model for predicting a particular CoI.
- *Stochastic Multiscale Modelling* provides challenges that present UQ methods cannot handle
  - *Data-driven definition of coarse grained variables and their evolution* (from variational to fully Bayesian approaches), non-parametric modelling, capturing & propagating information loss during coarse graining
  - *Curse of Dimensionality:* Its all about data-driven (local) exploration of correlations within and across scales (*partitioning the space vs the data*)
  - Approximate inference (on graphs) a promising scalable UQ approach (*Deep Learning*)