Uncertainty Quantification and Propagation in Multiscale Materials Modelling

James Kermode
Warwick Centre for Predictive Modelling
School of Engineering, University of Warwick

CECAM 'Error control in first principles modelling'
Lausanne – 21 June 2022

Chemomechanical Processes

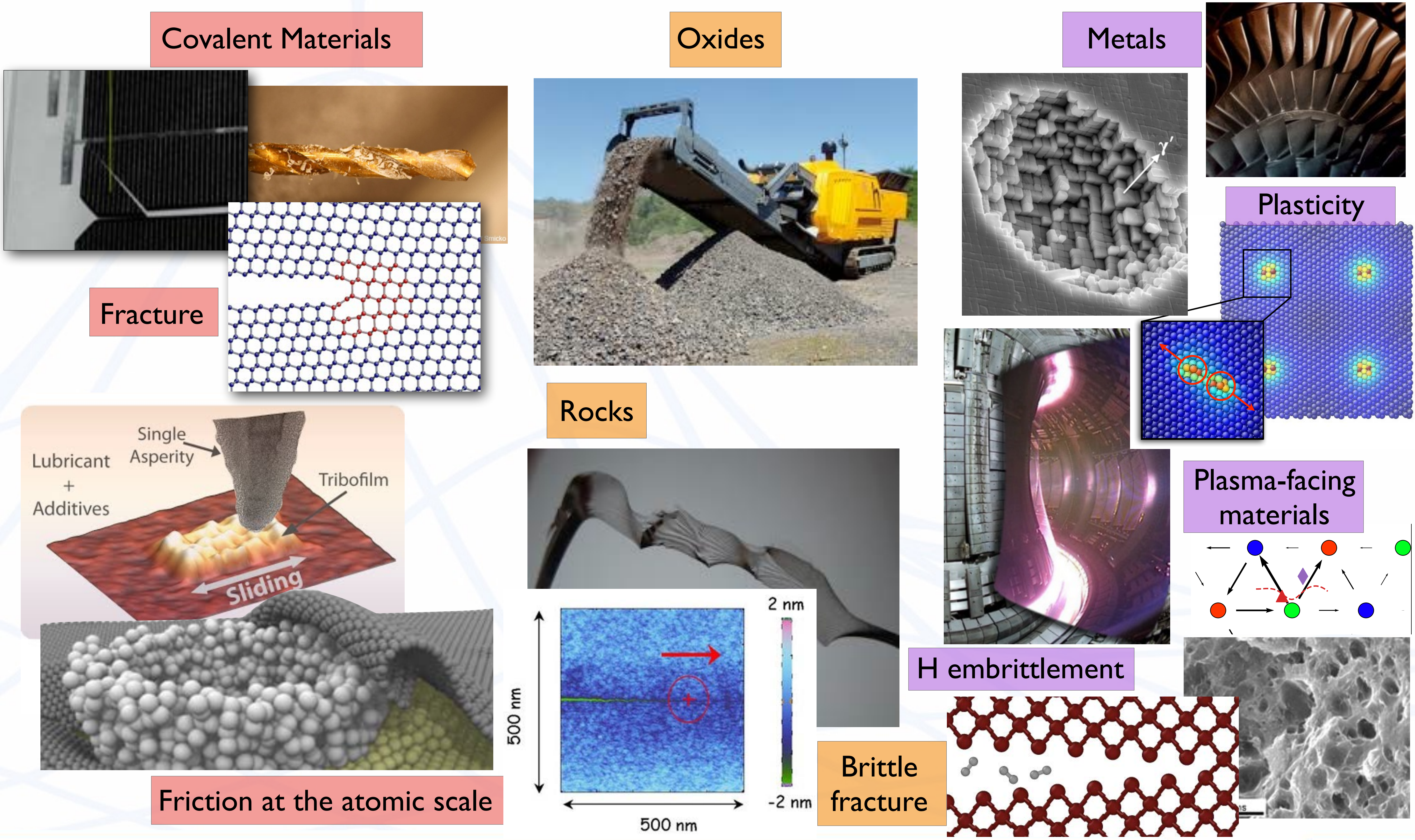
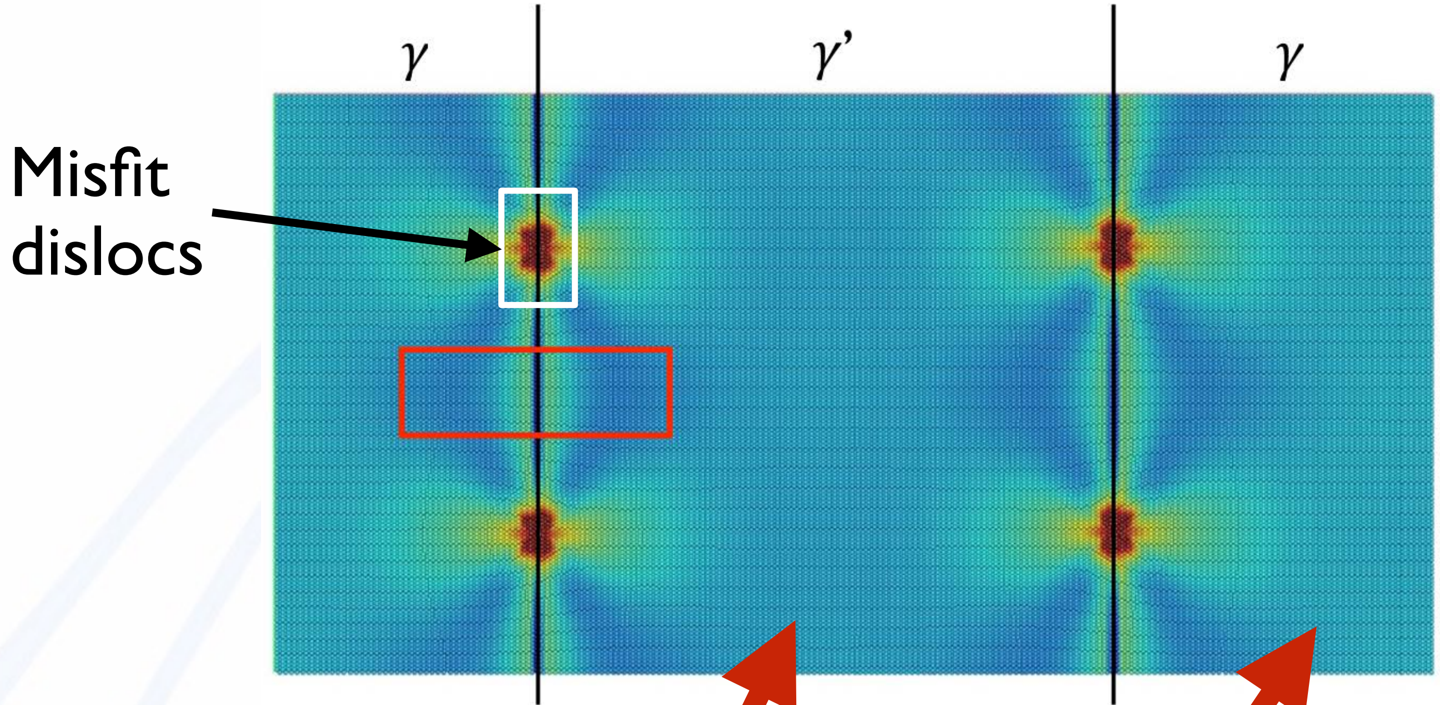


Image credits: R. Carpick, L. Pastewka, C. Marlière, A. Weck, B. Lawn, A. Piglione

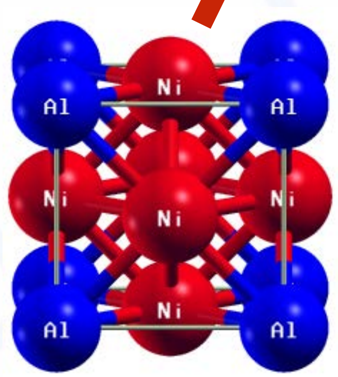
Review of atomistic aspects of fracture: E Bitzek, JRK, P Gumbsch, *Int. J. Fract.* (2015)

Example: dislocation glide in Ni-based superalloys

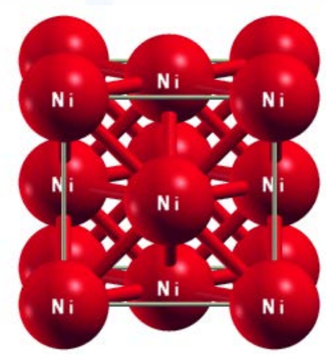


MD simulation of dislocations in γ phase Ni

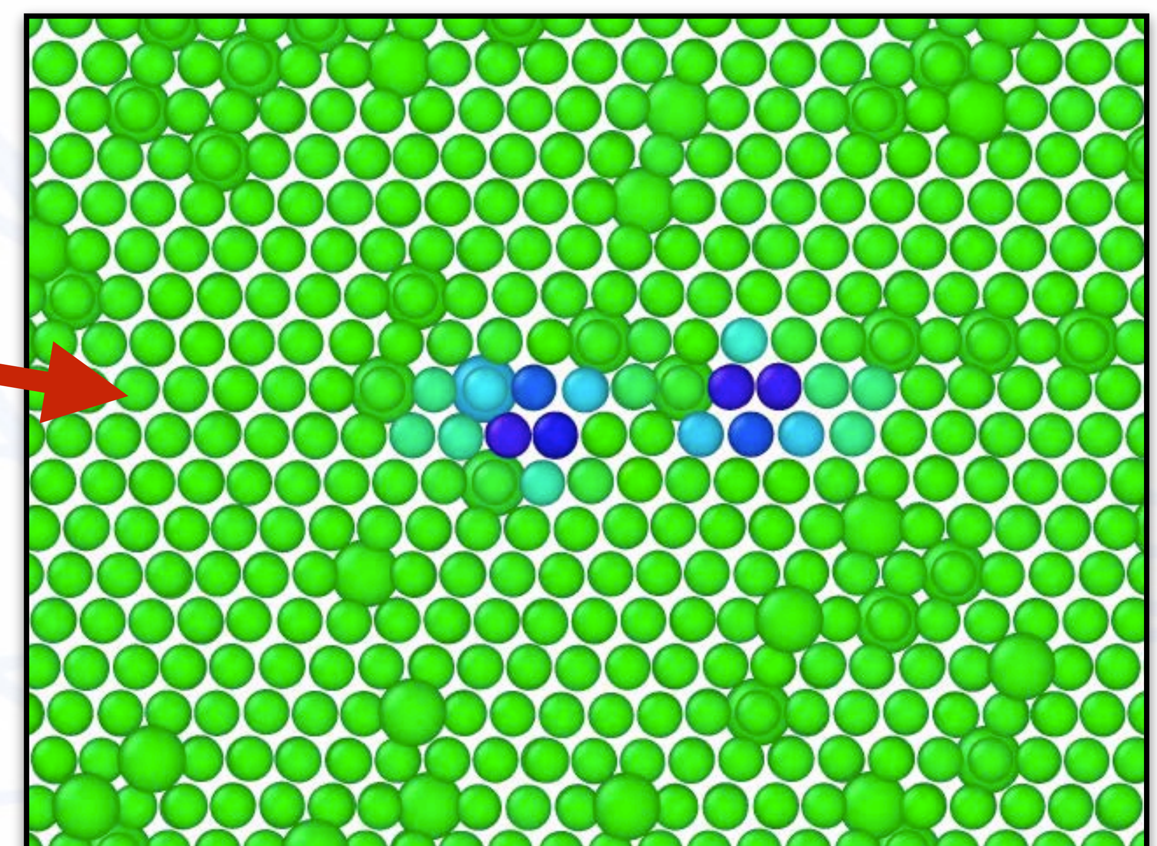
- Aleatoric and epistemic uncertainties
- Model uncertainty: how accurate is interatomic potential?
  - parameters and functional form
- Random microstructures, limited data
- Algorithmic uncertainty in solvers
- Limited transferability: can we model chemical complexity, e.g. impurities?



γ' phase



γ phase

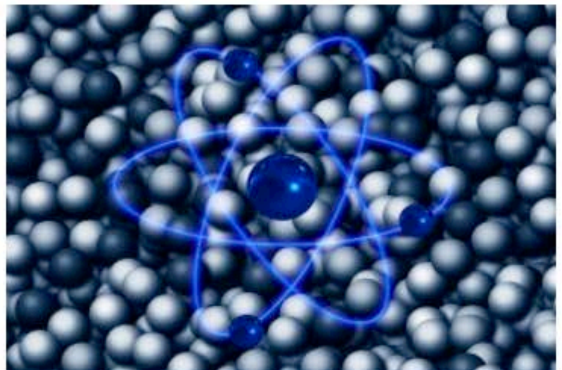


EAM, 5% Al, T = 300 K, 100 MPa shear stress

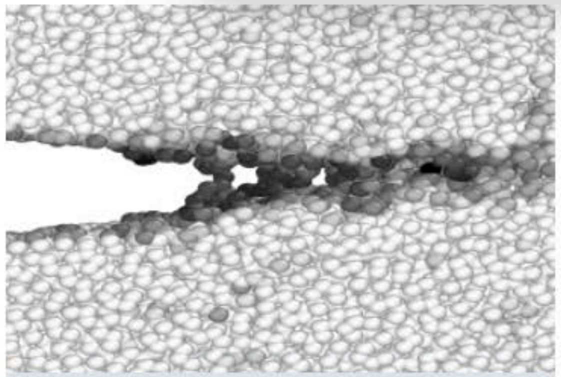
F Bianchini, JRK and A De Vita, *Modell. Simul. Mater. Sci. Eng.* **24** 045012 (2016)

F Bianchini, A Glielmo, JRK and A De Vita **3** 043605 *Phys. Rev. Mat.* (2019)

HetSys CDT and Warwick Centre for Predictive Modelling



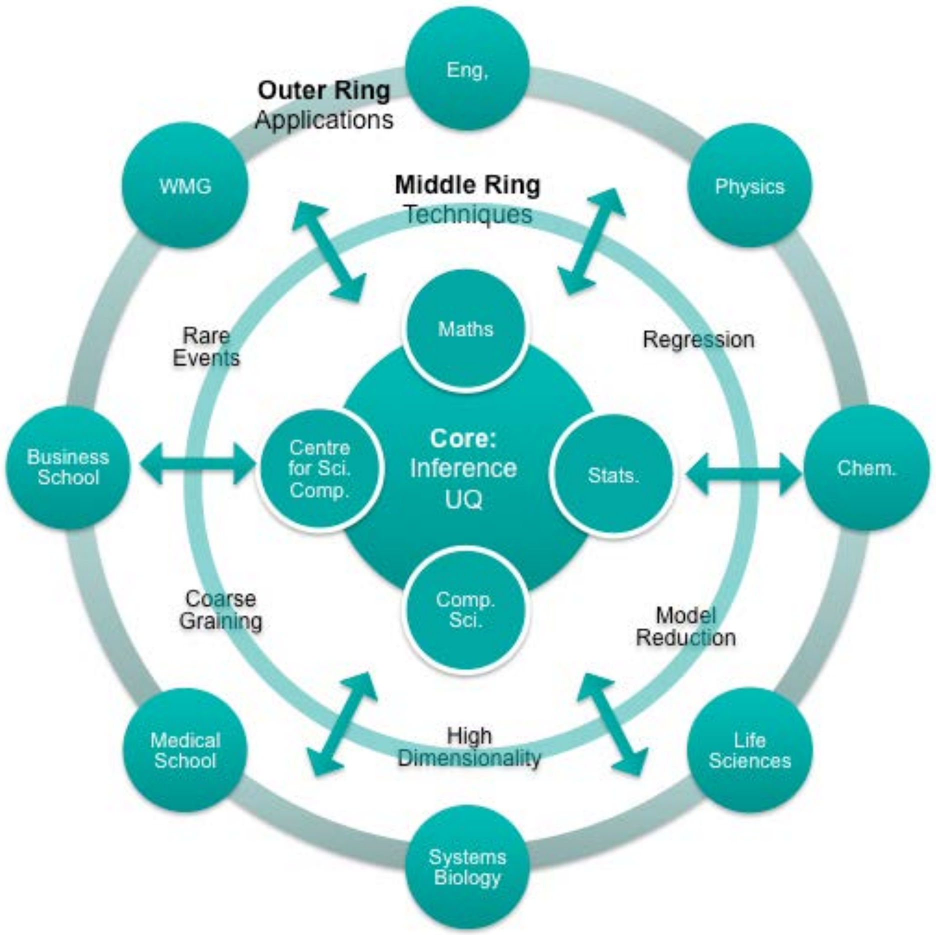
Quantum: Electrons, atoms and molecules for catalysis, medicines and devices



Atomistic: Materials structure, phases and defects for properties and applications

Continuum: New methods for fluids, plasma, porous media and composites for technological solutions

WCPM



The HetSys' training programme is designed to enable students to become high-quality computational scientists who are comfortable working in interdisciplinary environments, have excellent communication skills, and well prepared for a wide range of future careers in areas where there is demonstrable need.

The HetSys training programme will meet three key training needs:

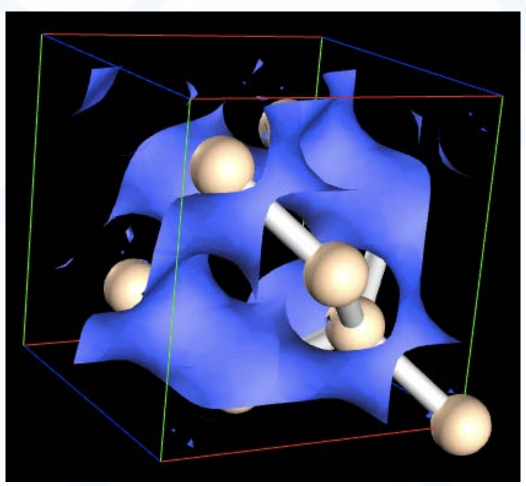
1. **Span disciplinary barriers.** The most challenging real-world heterogeneous systems are intrinsically multidisciplinary, requiring integration of a diverse range of modelling methods.
2. **Incorporate uncertainty in modelling.** Training in uncertainty quantification will enable students not only to perform simulations, but also to quantitatively assess their reliability.
3. **Promote robust Research Software Engineering (RSE).** Training in sustainable software development will enhance software usability and extend its lifetime.

www.warwick.ac.uk/wcpm
www.warwick.ac.uk/HetSys

Quantifying uncertainties across the scales

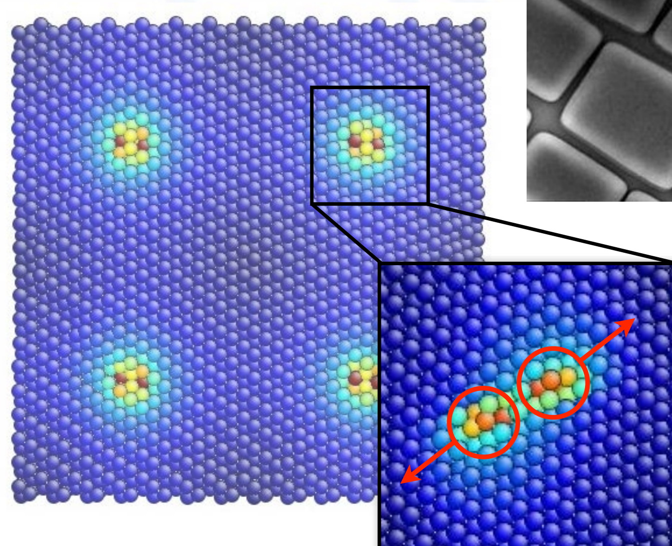
Stochastic coarse graining

Electronic structure



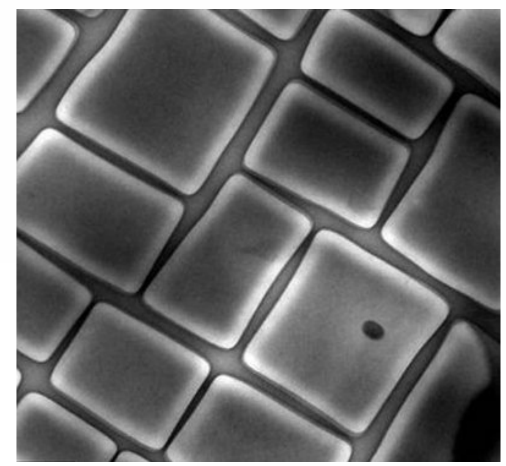
Quantifying uncertainty in exchange-correlation functionals

Atomistic/ nanoscale



Quantifying uncertainty in interatomic potentials

Microstructure
(e.g. superalloy)



Modelling rare events in heterogenous media

Macroscale
(e.g. turbine blade)

Materials design: error bars on properties

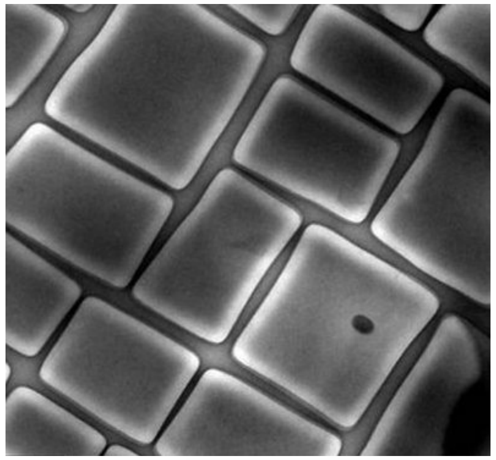
Quantifying uncertainties in electronic structure

Stochastic coarse graining

Macroscale
(e.g. turbine blade)

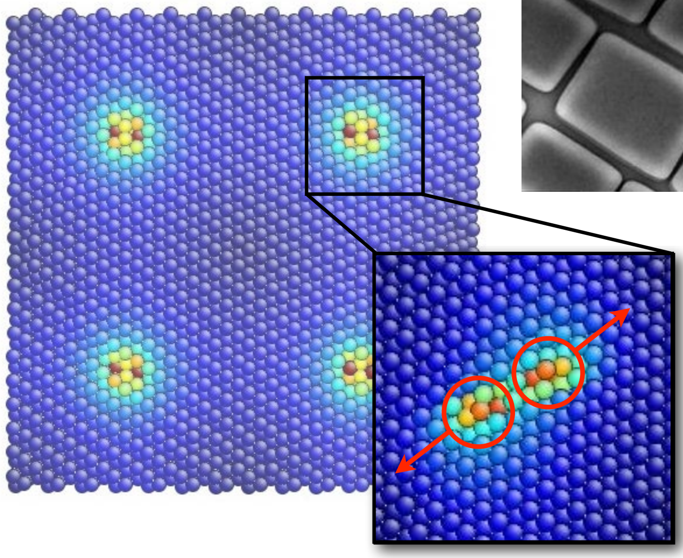
Materials design: error bars on properties

Microstructure
(e.g. superalloy)



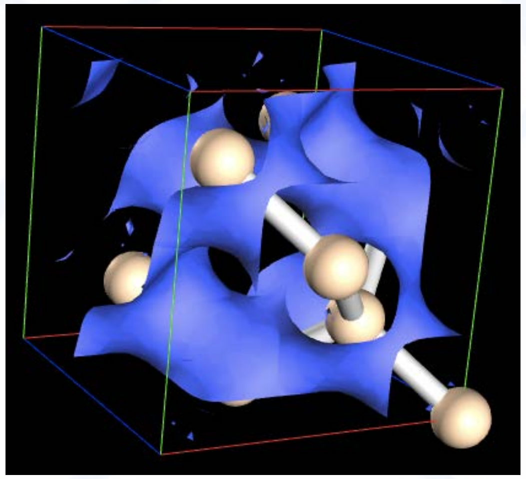
Modelling rare events in heterogenous media

Atomistic/ nanoscale



Quantifying uncertainty in interatomic potentials

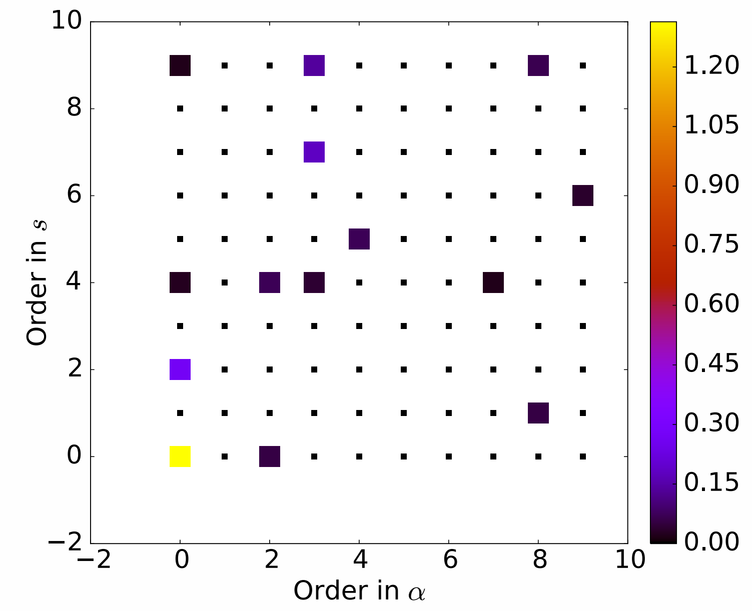
Electronic structure



Quantifying uncertainty in exchange-correlation functionals

UQ for DFT Exchange Functionals

- Linear model for meta-GGA exchange energy



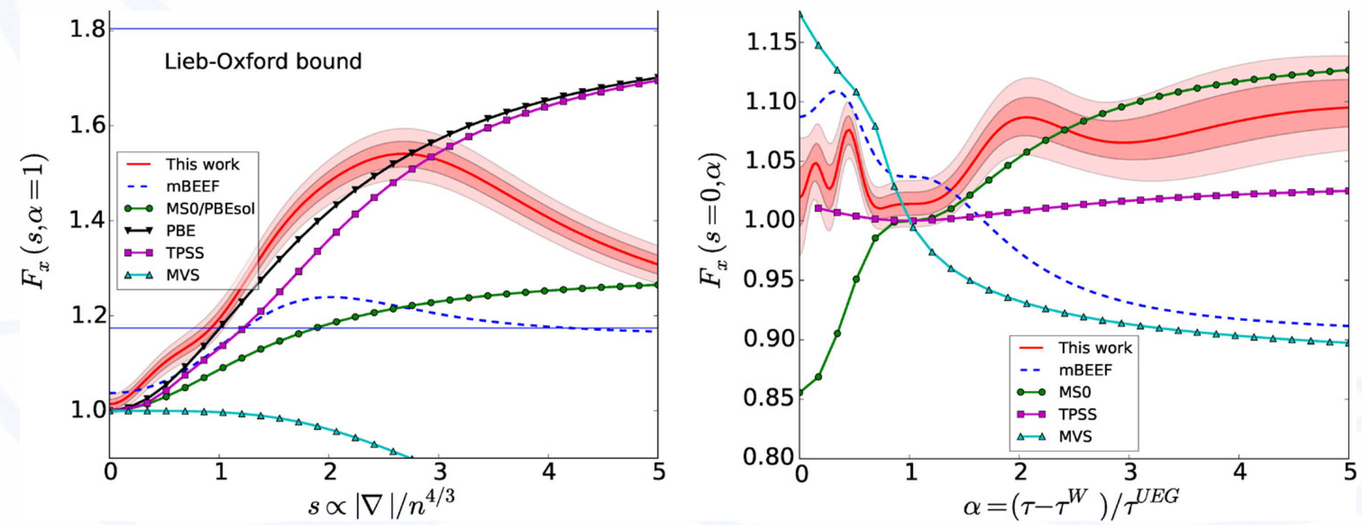
$$E^{xc}[n] = E^{c}[n] + \int n\varepsilon^{x}(n(\mathbf{r}), \nabla n(\mathbf{r}), \tau(\mathbf{r}))\, d\mathbf{r}$$

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

$$E^{x}[n; \boldsymbol{\xi}^{x}] = \sum_{i}^{M_s} \sum_{j}^{M_\alpha} \xi^{x}_{ij} \int n\varepsilon^{x}_{UEG}(n) P_i(t_s(s)) P_j(t_\alpha(\alpha))\, d\mathbf{r}$$

- Assume observed data (experimental atomisation energies, plus energy-volume data) **t** follows proposed model on average, with iid Gaussian observational noise

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



- Conjugate priors for parameters $\boldsymbol{\xi}$ and β

$$p(\boldsymbol{\xi} \mid \beta, \mathbf{m}_0, \mathbf{S}_0) = \mathcal{N}(\boldsymbol{\xi} \mid \mathbf{m}_0, \beta^{-1}\mathbf{S}_0)$$

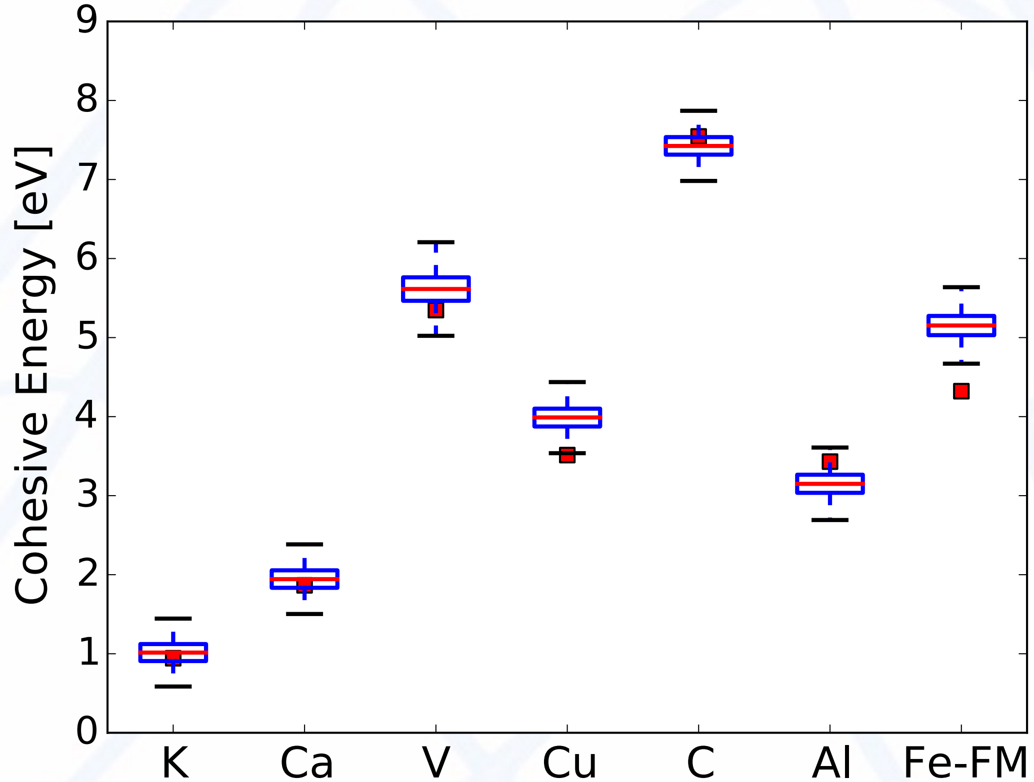
$$p(\beta \mid a_0, b_0) = \mathcal{G}(\beta \mid a_0, b_0)$$

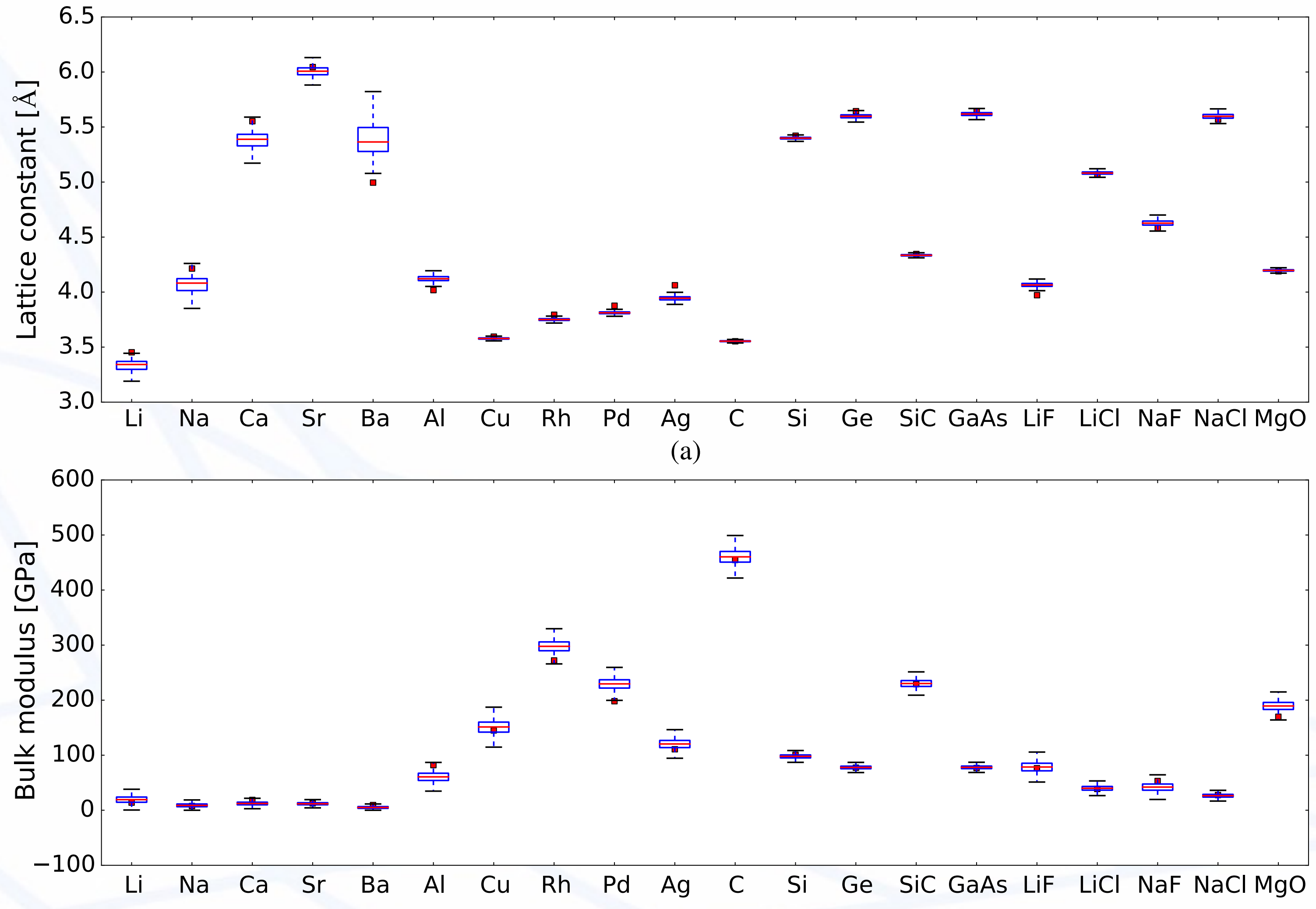
- Standard Bayesian linear regression gives analytic posterior predictive distrib for $E^{x}[n]$
  Use ARD with relevance vector machine to prevent overfitting.

M. Aldegunde, JRK, N. Zabaras, *J. Comp. Phys.* **311** 173-195 (2016)

Propagating uncertainties to bulk properties

- Nested Monte Carlo – sample model coefficients for E^x from posterior distribution, then fit eq. of state to yield distributions of B_0 & a_0
- Can also include numerical errors, e.g. Gaussian-distributed with std. dev. 10 meV
- Extensible to other QoIs: we also looked at band gaps at KS and G_0W_0 level

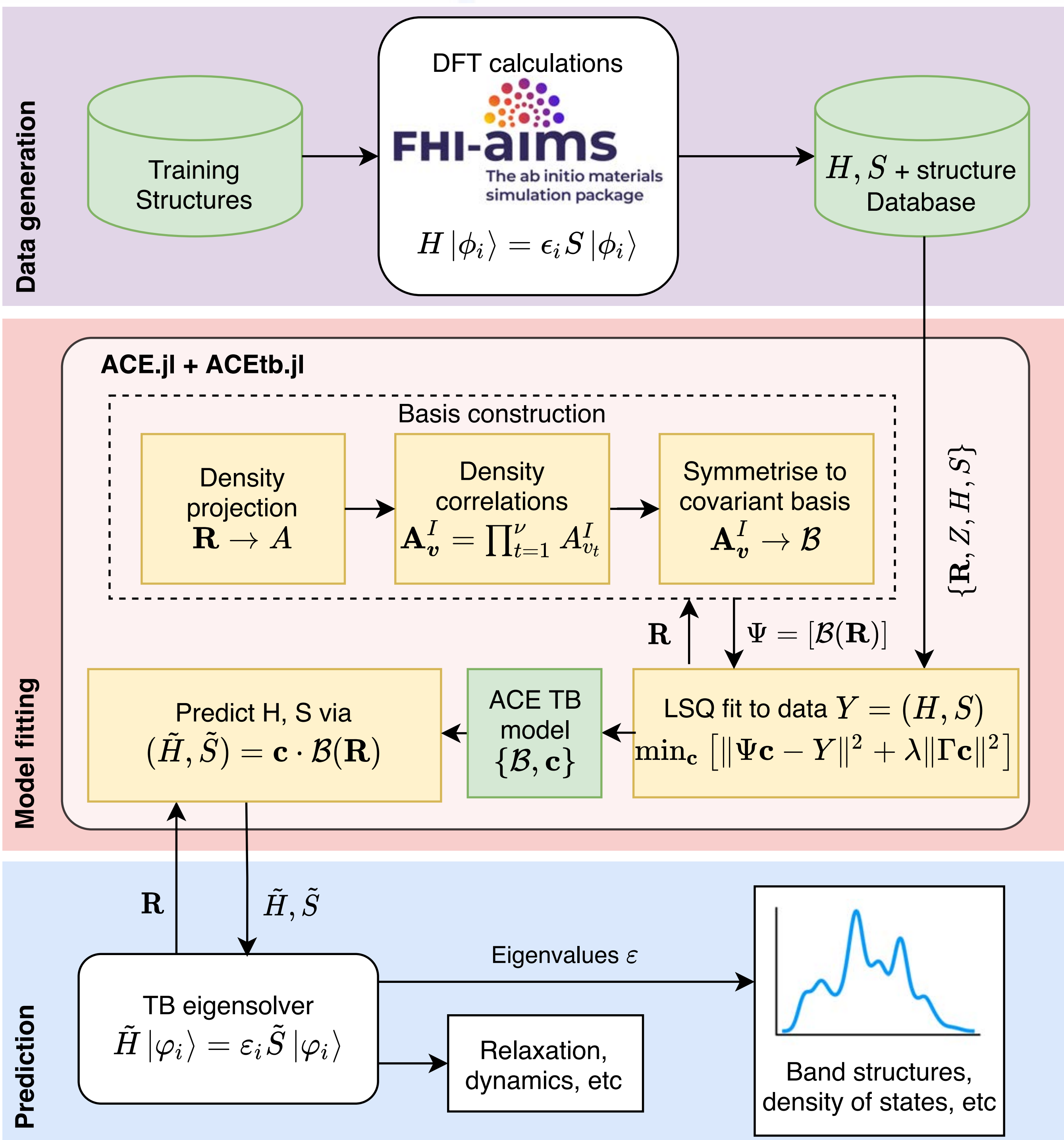


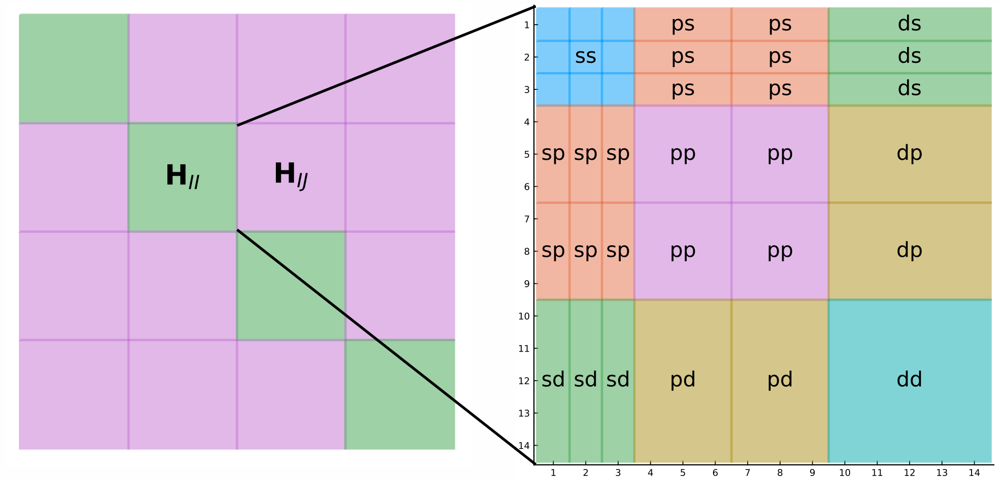


XC functional	Error	G2/97-test	G2/97	EL20-test	EL20
This work	MAE	0.116	0.103	0.243	0.0975
	MARE	3.27	1.46	8.56	5.62
PBE	MAE		0.703		0.238
	MARE		5.09		6.88

M. Aldegunde, JRK, N. Zabaras, *J. Comp. Phys.* **311** 173-195 (2016)

Parameterising Hamiltonians from DFT data





Blocks of H, S have equivariant structure:

$$\boldsymbol{H}_{\mathrm{on/off}}(Q\mathbf{R}) = D(Q)^* \boldsymbol{H}_{\mathrm{on/off}}(\mathbf{R}) D(Q).$$

Represent blocks using ACE basis

$$\boldsymbol{H}_{II} = \boldsymbol{H}_{\mathrm{on}}(\mathbf{R}_I) \approx \tilde{\boldsymbol{H}}^{\mathrm{PI}}_{\mathrm{on}}(\mathbf{R}_I) = \sum_{\boldsymbol{v}} C_{\boldsymbol{v}} \boldsymbol{A}^I_{\boldsymbol{v}},$$

$$\boldsymbol{H}_{IJ} = \boldsymbol{H}_{\mathrm{off}} \approx \tilde{\boldsymbol{H}}^{\mathrm{PI}}_{\mathrm{off}}(\boldsymbol{r}_{IJ}, \mathbf{R}_{IJ}) := \sum_{\boldsymbol{v}} C_{\boldsymbol{v}} \boldsymbol{A}^{IJ}_{\boldsymbol{v}}.$$

Symmetrising by integrating over *O(3)* gives *linear* models for each on/offsite block:

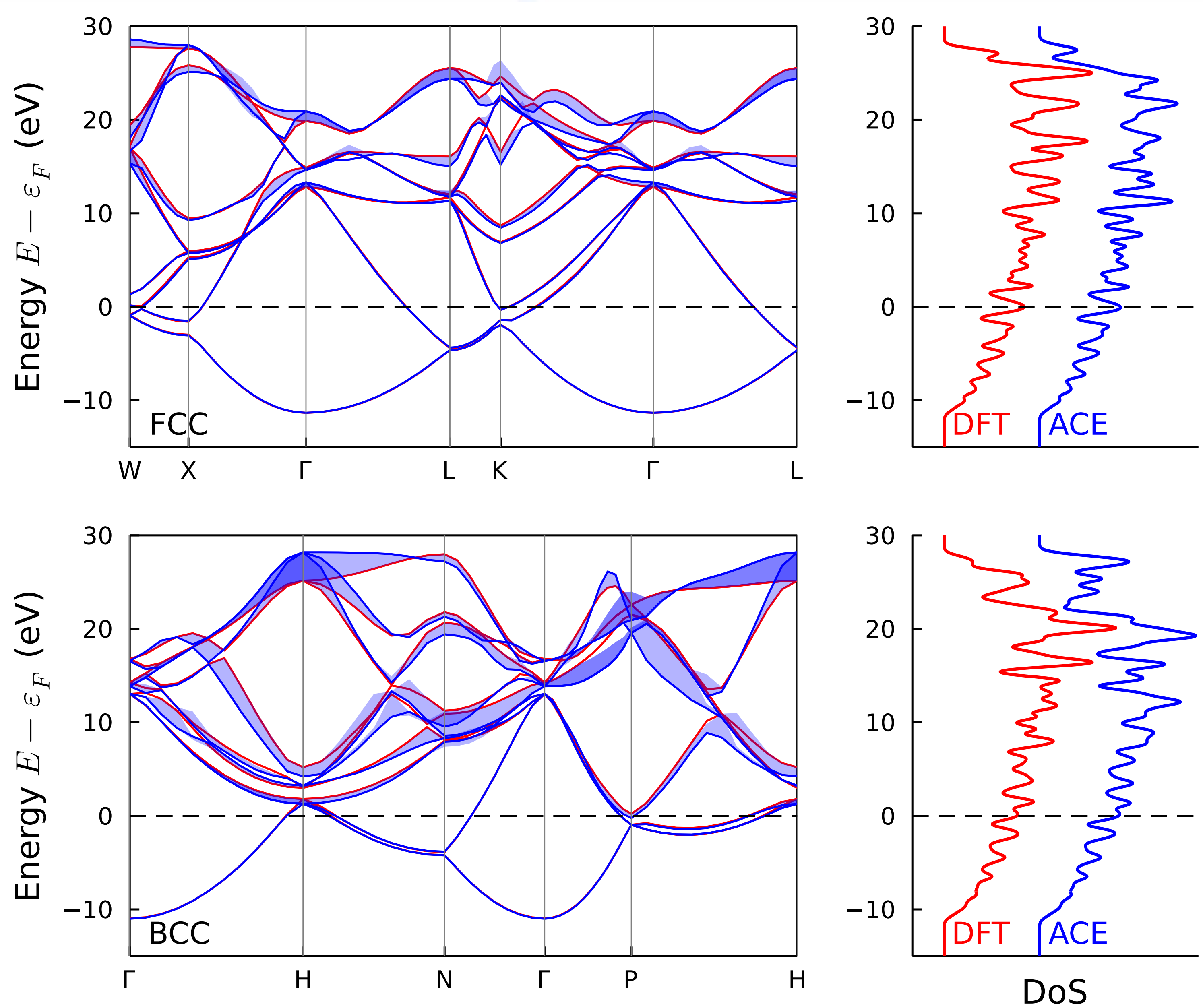
$$\tilde{H}_{\mathrm{on}} := \mathbf{c}^{\mathrm{on}} \cdot \mathcal{B}^{\mathrm{on}},$$

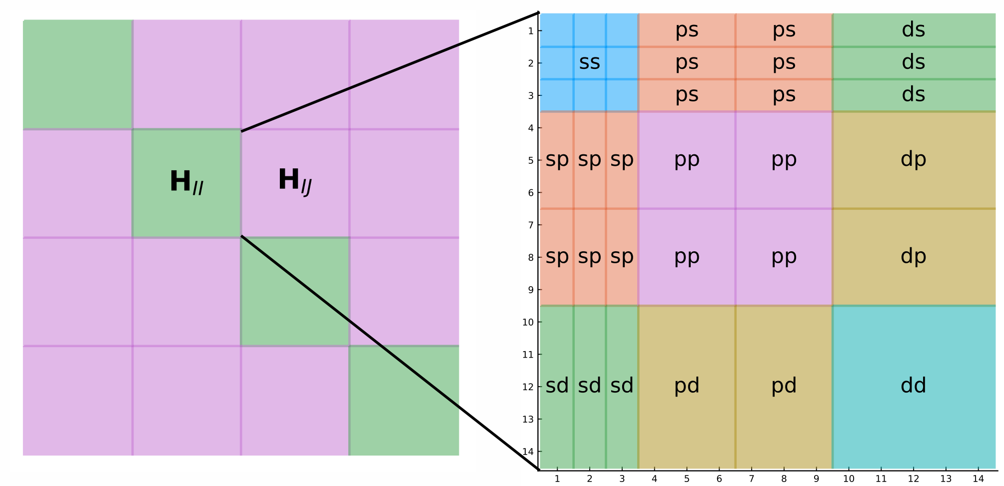
$$\tilde{H}_{\mathrm{off}} := \mathbf{c}^{\mathrm{off}} \cdot \mathcal{B}^{\mathrm{off}},$$

$$\tilde{S}_{\mathrm{off}} := \mathbf{c}^{\mathrm{S}} \cdot \mathcal{B}^{\mathrm{S}},$$

Atomic Cluster Expansion (ACE): R. Drautz, Phys. Rev. B. **99**, 014104 (2019)
Completeness: G. Dusson, M. Bachmayr, G. Csanyi, R. Drautz, S. Etter, C. van der Oord and C. Ortner, arXiv:1911.03550
L. Zhang, B. Onat, A. McSloy, G. Dusson, G. Anand, R.J. Maurer, C. Ortner and JRK, In press at npj Comput Mater (2022)

Parameterising Hamiltonians from DFT data





Blocks of H, S have equivariant structure:

$$\boldsymbol{H}_{\mathrm{on/off}}(Q\mathbf{R}) = D(Q)^* \boldsymbol{H}_{\mathrm{on/off}}(\mathbf{R}) D(Q).$$

Represent blocks using ACE basis

$$\boldsymbol{H}_{II} = \boldsymbol{H}_{\mathrm{on}}(\mathbf{R}_I) \approx \tilde{\boldsymbol{H}}^{\mathrm{PI}}_{\mathrm{on}}(\mathbf{R}_I) = \sum_{\boldsymbol{v}} C_{\boldsymbol{v}} \boldsymbol{A}^I_{\boldsymbol{v}},$$

$$\boldsymbol{H}_{IJ} = \boldsymbol{H}_{\mathrm{off}} \approx \tilde{\boldsymbol{H}}^{\mathrm{PI}}_{\mathrm{off}}(\boldsymbol{r}_{IJ}, \mathbf{R}_{IJ}) := \sum_{\boldsymbol{v}} C_{\boldsymbol{v}} \boldsymbol{A}^{IJ}_{\boldsymbol{v}}.$$

Symmetrising by integrating over *O(3)* gives *linear* models for each on/offsite block:

$$\tilde{\boldsymbol{H}}_{\mathrm{on}} := \mathbf{c}^{\mathrm{on}} \cdot \mathcal{B}^{\mathrm{on}},$$

$$\tilde{\boldsymbol{H}}_{\mathrm{off}} := \mathbf{c}^{\mathrm{off}} \cdot \mathcal{B}^{\mathrm{off}},$$

$$\tilde{\boldsymbol{S}}_{\mathrm{off}} := \mathbf{c}^{\mathrm{S}} \cdot \mathcal{B}^{\mathrm{S}},$$

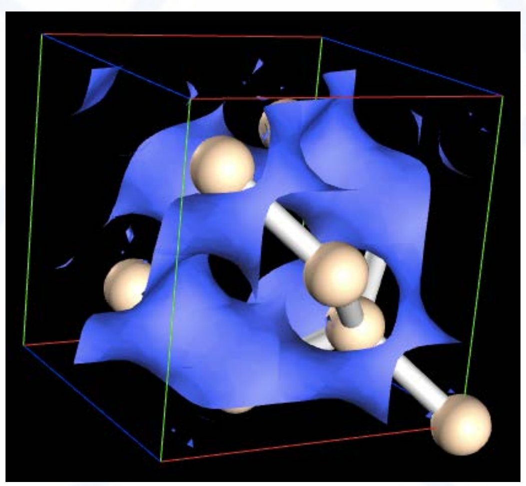
Application to Al: trained on 500 K MD for FCC and BCC.
Can also predict electronic structure along Bain path and near vacancies without expanding training set.

Atomic Cluster Expansion (ACE): R. Drautz, Phys. Rev. B. **99**, 014104 (2019)
Completeness: G. Dusson, M. Bachmayr, G. Csanyi, R. Drautz, S. Etter, C. van der Oord and C. Ortner, arXiv:1911.03550
L. Zhang, B. Onat, A. McSloy, G. Dusson, G. Anand, R.J. Maurer, C. Ortner and JRK, In press at npj Comput Mater (2022)

Quantifying uncertainties in atomistic simulations

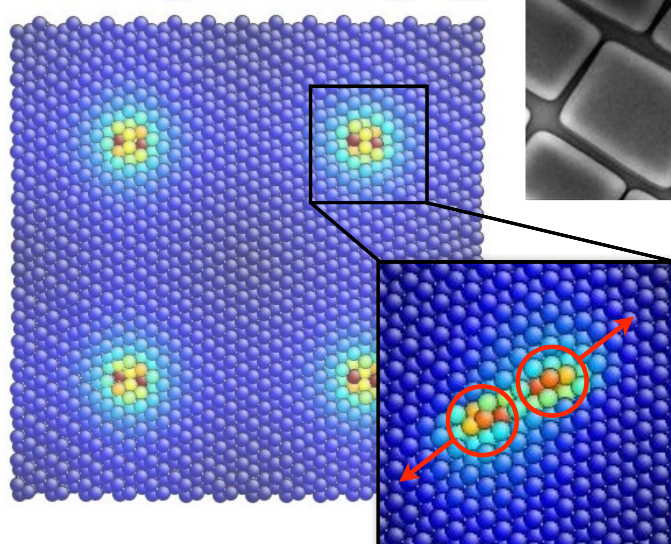
Stochastic coarse graining

Electronic structure



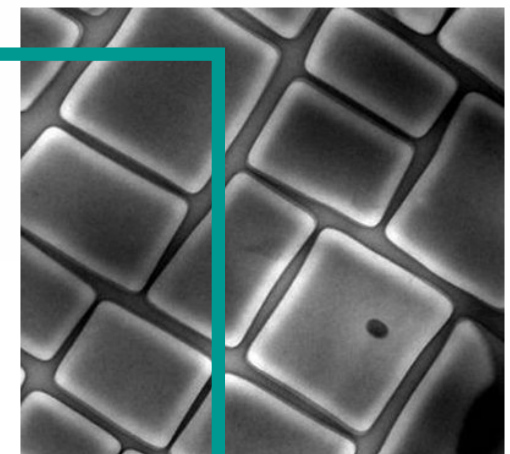
Quantifying uncertainty in exchange-correlation functionals

Atomistic/ nanoscale



Quantifying uncertainty in interatomic potentials

Microstructure
(e.g. superalloy)



Modelling rare events in heterogenous media

Macroscale
(e.g. turbine blade)

Materials design: error bars on properties

UQ for potentials with Bayesian linear regression

Uncertainty Quantification in Atomistic Simulations using Interatomic Potentials

Iain Best, Tim Sullivan, James Kermode
HetSys CDT, University of Warwick, Coventry, UK
iain.best@warwick.ac.uk

WARWICK
THE UNIVERSITY OF WARWICK

Introduction

- Interatomic potentials (IPs) are widely used in materials modelling and other disciplines to compute physical quantities of interest (QoIs).
- IP use offers vastly reduced simulation time/cost when compared with ab-initio methods like density functional theory (DFT), allowing access to otherwise impractical time- and length- scales.
- Since IP use also reduces accuracy and increases uncertainty in QoIs, we seek a method of calculating statistically meaningful error bars, by recasting model calibration as a Bayesian inverse problem.

1) Bayesian Inverse Problems

For a model V with coefficients $\mathbf{w}$, some inputs $\mathbf{x}$, targets $\mathbf{y}$ and precision β on said data, a basic Bayesian inverse problem can be broken into stages;

1. specify **prior** distribution for coefficients $\mathbb{P}(\mathbf{w})$,
2. calculate **likelihood** of our model given data
$$\mathbb{P}(\mathbf{y}|\mathbf{x},\mathbf{w},\beta) = \mathcal{N}(\mathbf{y}|V(\mathbf{w},\mathbf{x}),\beta^{-1}),$$
3. from these, form **posterior** distribution for weights $\mathbb{P}(\mathbf{w}|\mathbf{y})$.

Once we have $\mathbb{P}(\mathbf{w}|\mathbf{y})$, we form an ensemble of potentials $\{V_i\}$, which we push through simulations, giving a distribution in the desired QoI.

3) ACE with BLR

We now shift our attention to the Atomic Cluster Expansion (ACE) potential [3]; which we view as a linear model

$$V(\{\mathbf{R}\}) = \sum_i w_i \phi_i(\{\mathbf{R}\}).$$

Taking advantage of Bayesian Linear Regression (BLR) and choosing a conjugate prior to our Gaussian likelihood, can write down our posterior distribution analytically

$$\mathbb{P}(\mathbf{w}|\mathbf{y}) = \mathcal{N}(\mathbf{w}|\boldsymbol{\mu},\mathbf{S}),$$

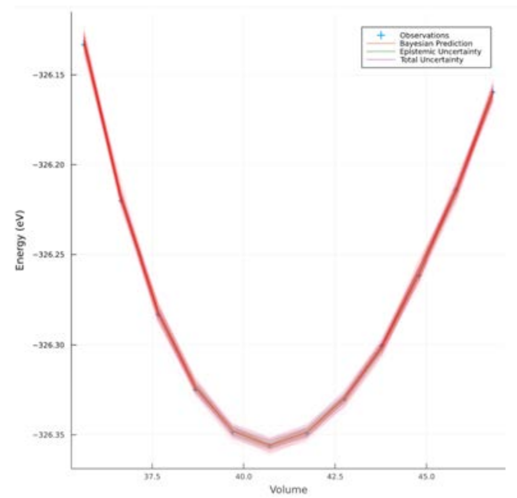


Figure 3: Representative samples from posterior shown on E-V curve for Si.

where covariance matrix $\mathbf{S} = (\alpha I + \beta\Phi^T\Phi)^{-1}$ and mean vector $\boldsymbol{\mu} = \beta\mathbf{S}\Phi^T\mathbf{y}$ are given in terms of a design matrix $\Phi_{N\times M}$, with Φ_{ij} giving the value of the j^{th} basis function on the i^{th} data point.

The α, β precision hyperparameters, on the weights and data respectively, are optimised to maximise the (log-) evidence

Iain Best, Tim Sullivan and JRK, Poster (2022)
plus discussions with Ralf Drautz, Yury Lysogorskiy, Ryan Elliot and Mark Transtrum

Gaussian Process regression – GPR

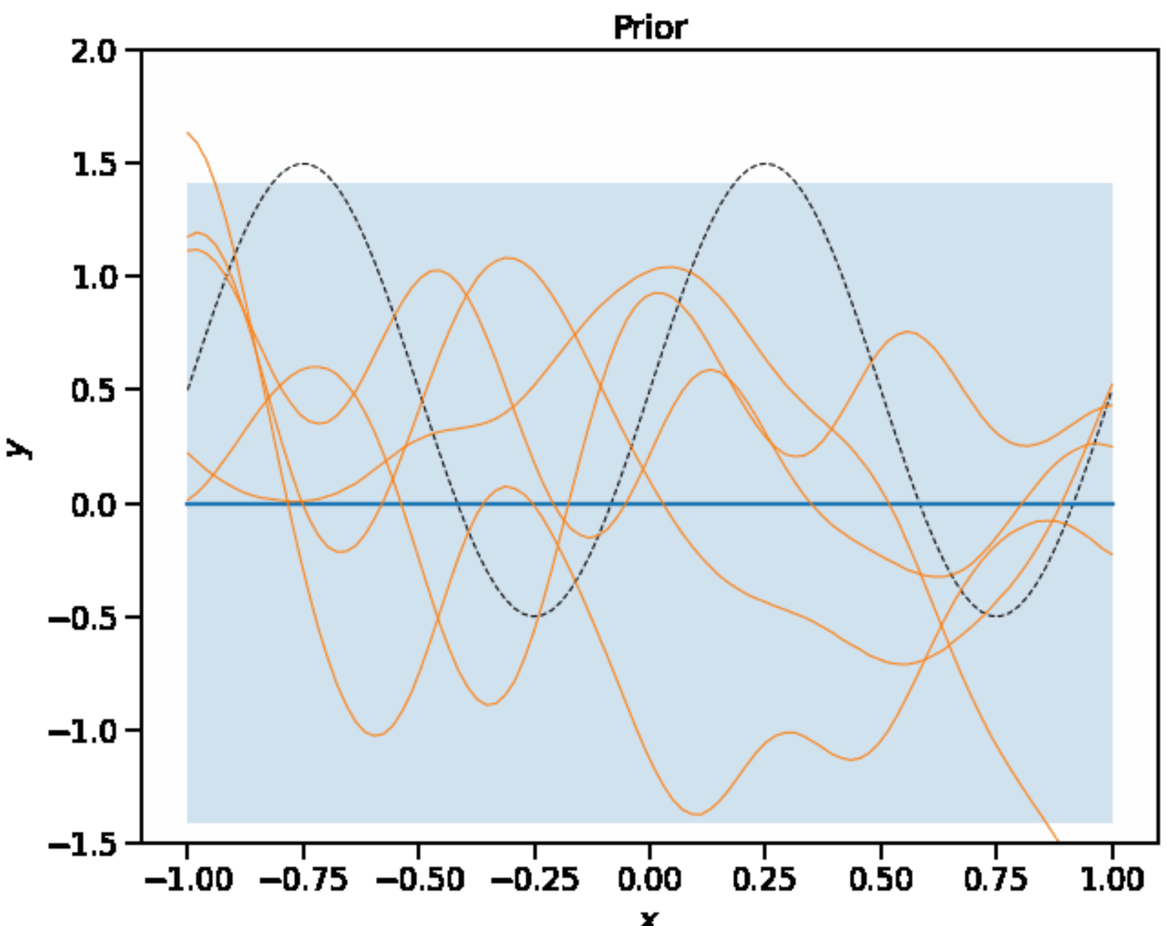
Infer most likely function values given data and prior covariance assumptions (typically smoothness)

Posterior: p(parameters|data) ∝ L(data|parameters) p(parameters)

Likelihood — L(data|parameters); *Prior* — p(parameters)

Prior - distribution over "smooth" functions

$$f \sim \mathcal{GP} \implies \mathbf{f} \sim \mathcal{N}\left(\mathbf{0}, K\right)$$



Gaussian Likelihood, i.e. observations are

$$y_i = f(\mathbf{x}_i) + \epsilon \text{ where } \epsilon \sim \mathcal{N}(0, \sigma_n^2)$$

Black: true function f(x)

Noisy observations *condition* (update) GP.

Crosses: noisy observations

Posterior also GP, with mean and variance at new point $\mathbf{x}^*$

Blue: GP mean, 95% confidence interval (2 std. devs)

$$\mathbb{E}[f^*] = \mathbf{k}_*^\top \left[K + \sigma_n^2 I\right]^{-1} \mathbf{y}$$

Orange: samples from prior/posterior

$$= \sum_{i=1}^{N} \alpha_i k(\mathbf{x}_i, \mathbf{x}_*) \text{ where } \boldsymbol{\alpha} = (K + \sigma_n^2 I)^{-1}\mathbf{y}$$

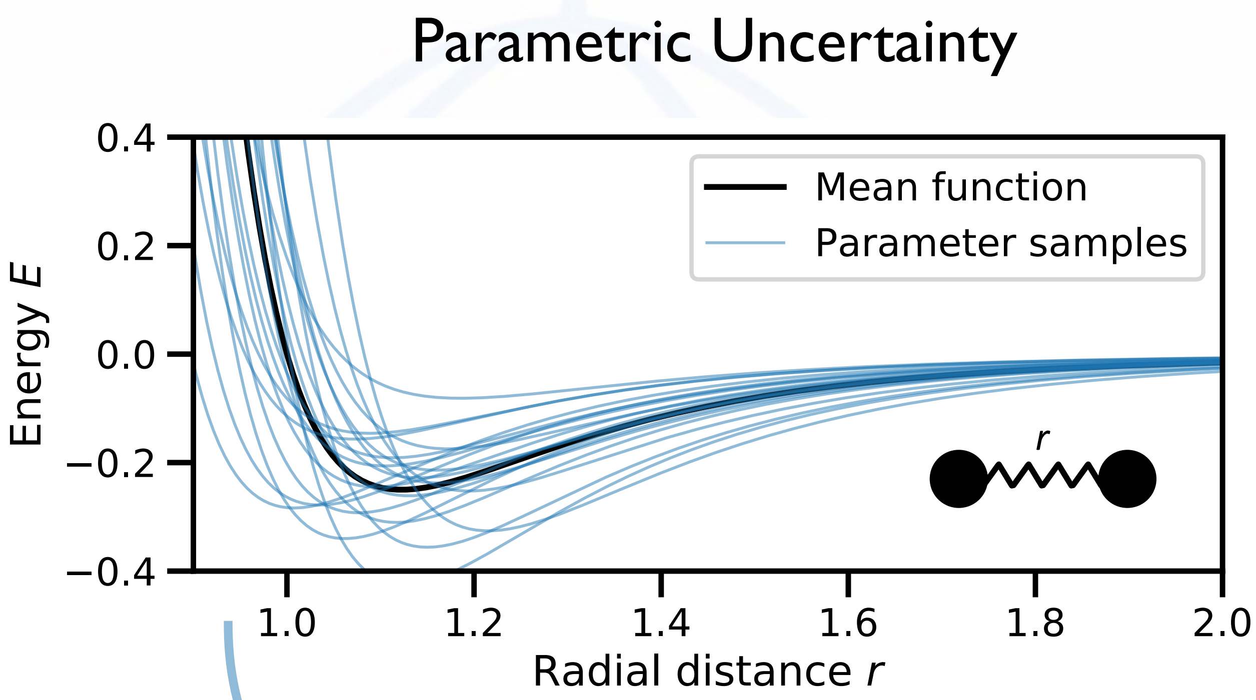
$$\mathrm{Var}[f^*] = K^* - \mathbf{k}_*^\top \left[K + \sigma_n^2 I\right]^{-1} \mathbf{k}_*$$

where $\mathbf{k}_* = K(\mathbf{x}^*, X)$ and $K_* = K(\mathbf{x}^*, \mathbf{x}^*)$

Smoothness set by kernel, e.g.

$$k(\mathbf{x}, \mathbf{x}') = v \exp\left[-\frac{1}{2}\sum_{i=1}^{d} \frac{(x_i - x_i')^2}{\ell_i^2}\right]$$

Quantifying Parametric Uncertainty

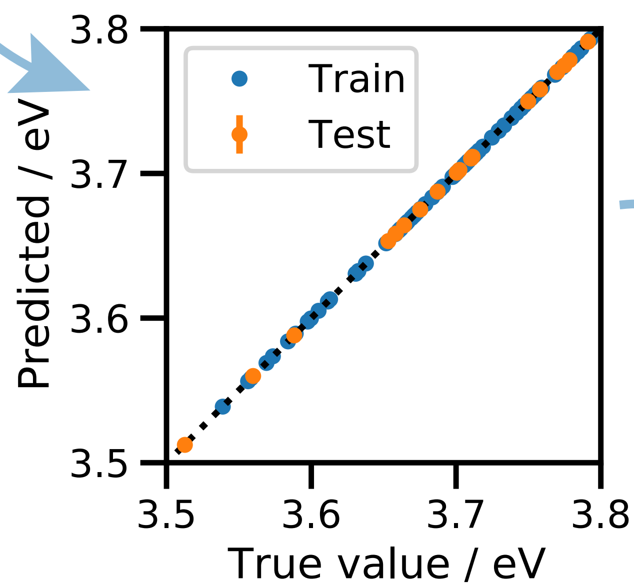


Example: vacancy formation energy using Tersoff model for silicon

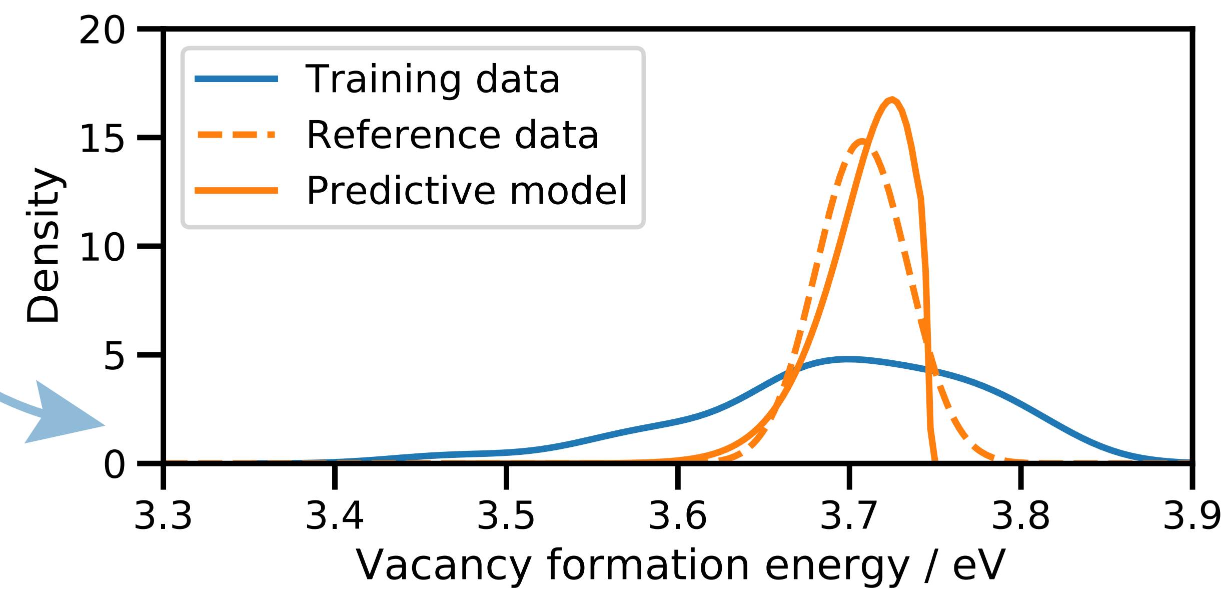
GP surrogate for parameters → QoI map trained with LHC design on ~100 points

Calibration wrt noisy experimental data as a Bayesian inverse problems solved via MCMC Allows 'sloppy' parameters to be identified.

Train surrogates using small ensembles of potentials (~10^2 samp)



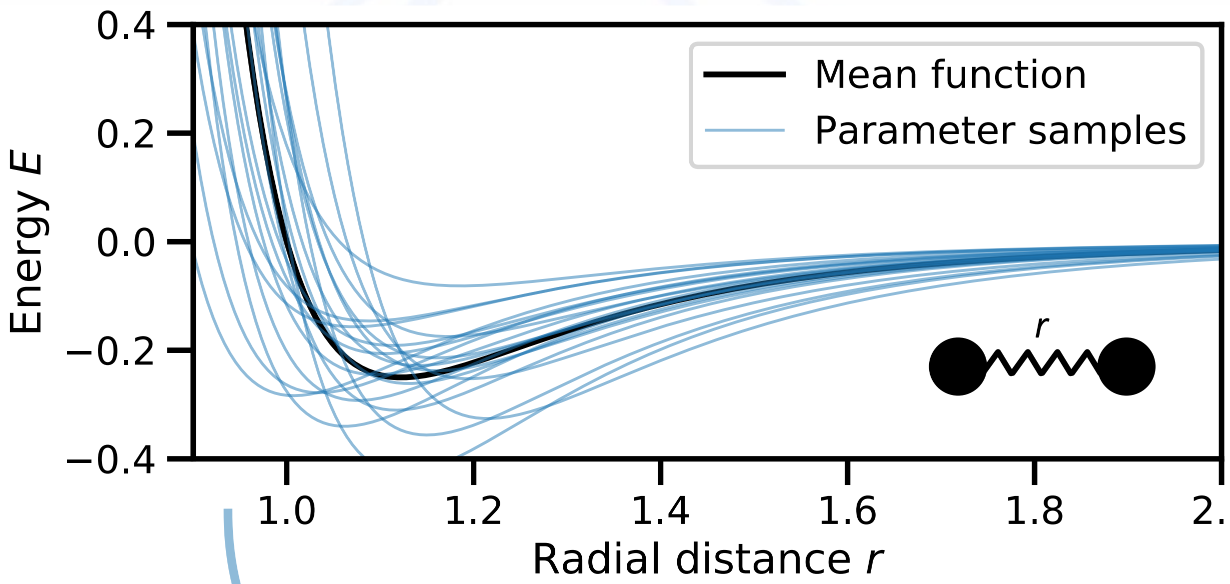
Propagate uncertainty to QoIs (~10^6 samp)



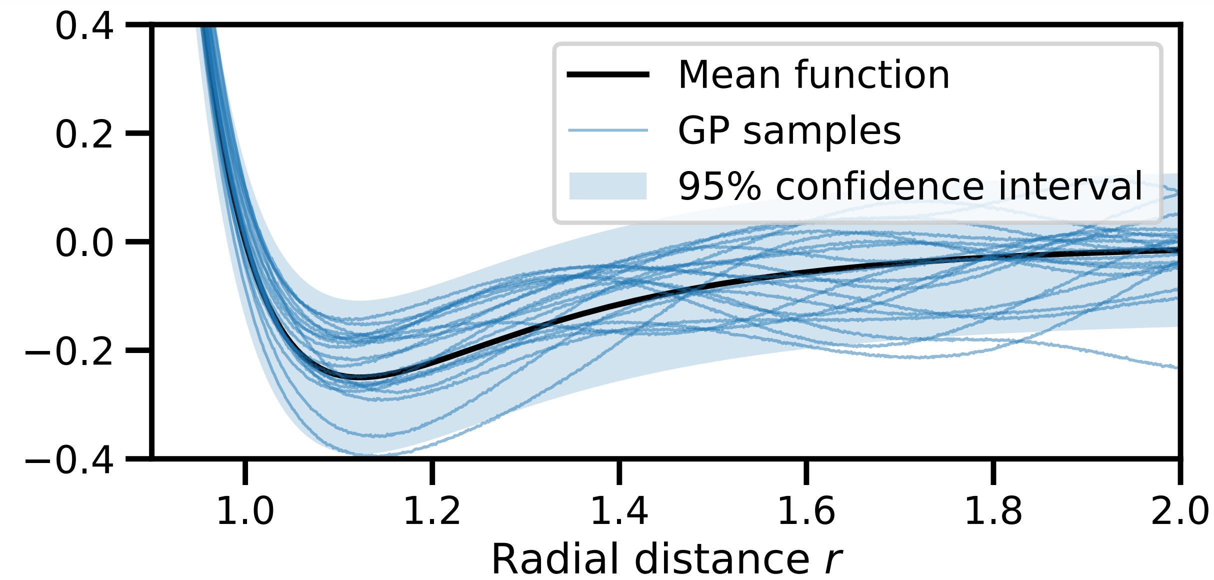
More details on Iain's poster - Iain Best, Tim Sullivan and JRK (2022)

Quantifying Model Form Error

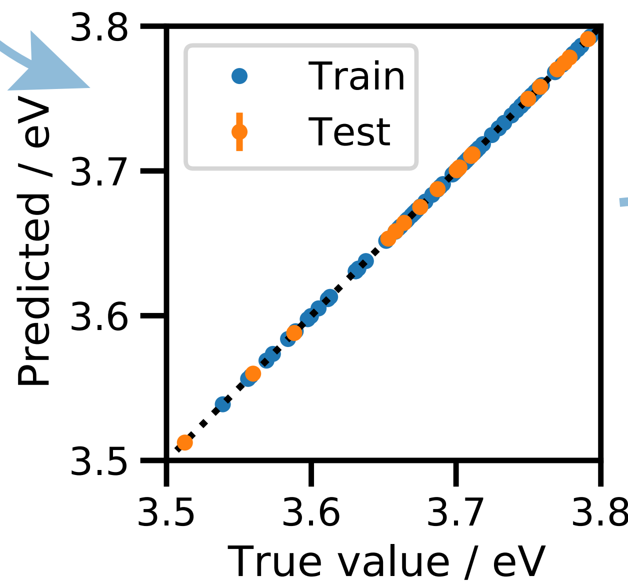
Parametric Uncertainty



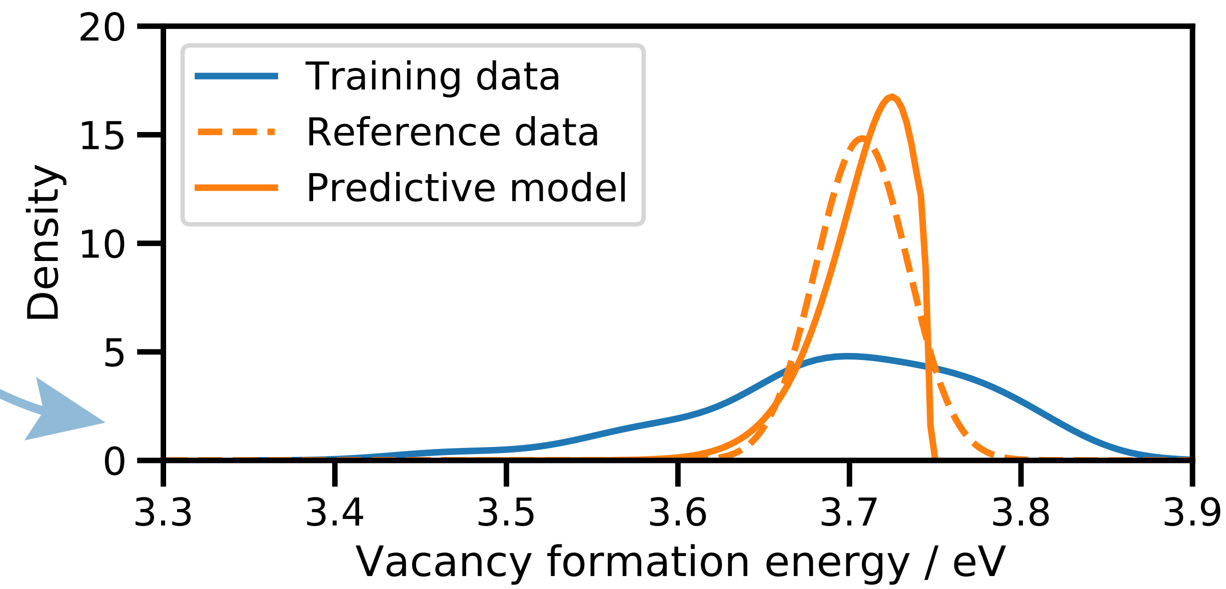
Model form error



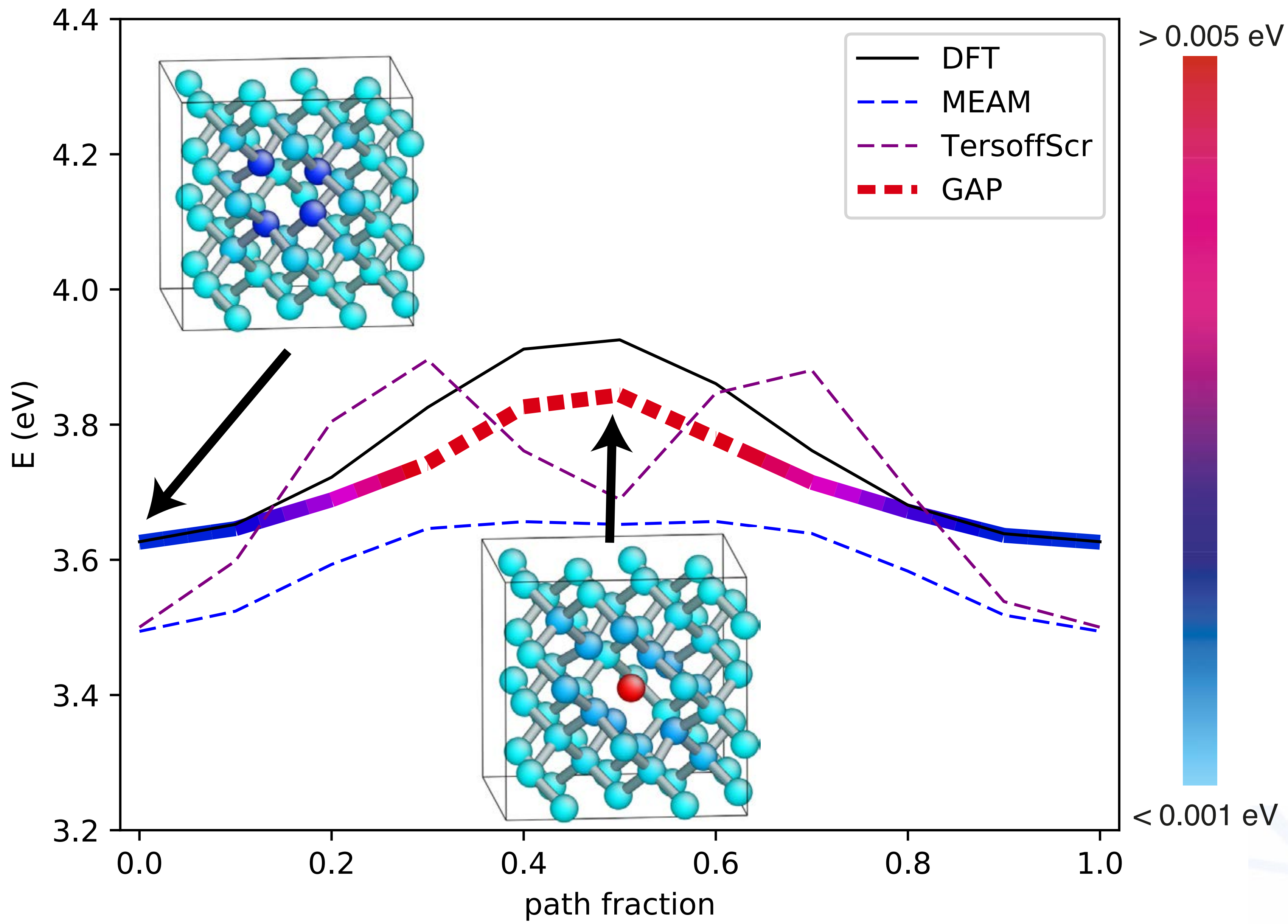
Train surrogates using small ensembles of potentials (~10² samp)

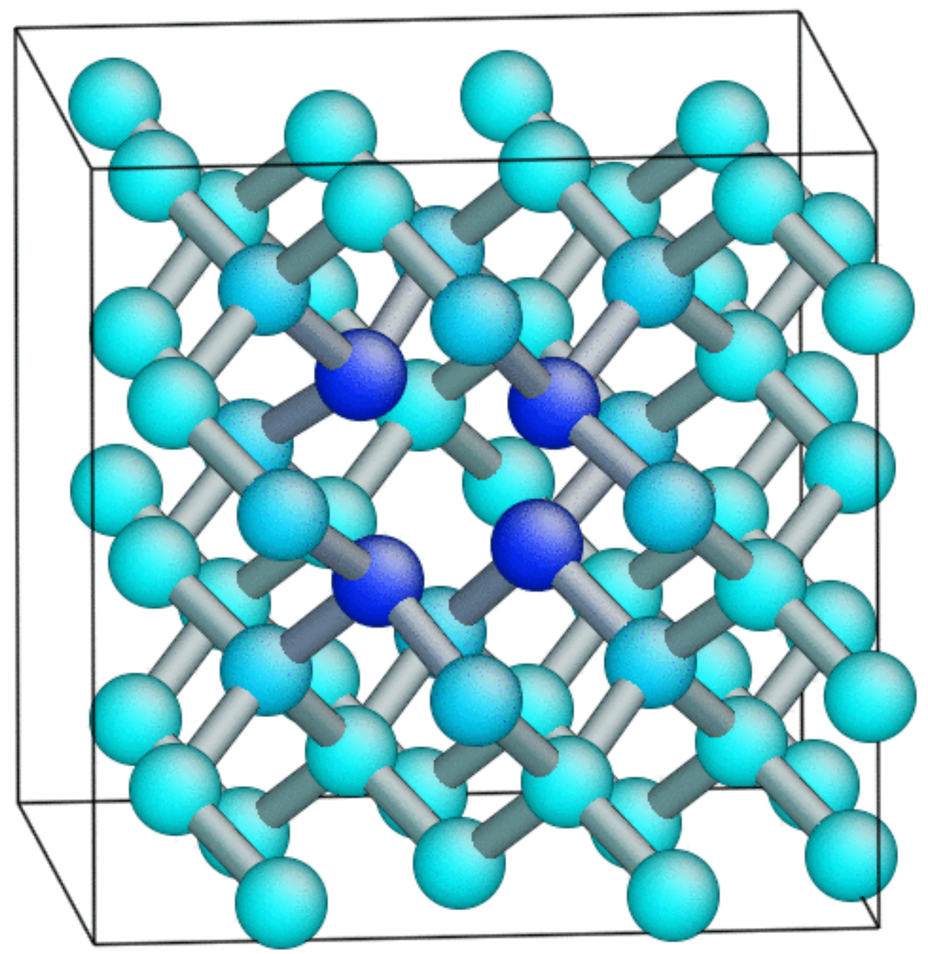


Propagate uncertainty to QoIs (~10⁶ samp)



GAP predictive variance – vacancy migration





GP predictive variance $V_i = \sigma_i^2 = K(\mathscr{R}_i, \mathscr{R}_i) - \mathbf{k}^T(\mathbf{K}_{MM} + \sigma_e^2\mathbf{I})^{-1}\mathbf{k}$ Regularisation $\sigma_e \approx 1$ meV/atom

(based on energies at sparse points only) $[\mathbf{k}]_s = K(\mathscr{R}_i, \mathscr{R}_s)$

A. P. Bartok, JRK, N. Bernstein and G. Csanyi, PRX **8**, 041048 (2018)

GAP uncertainty propagation & posterior samples

Interested in size of error in (100) surface energy in Fe predicted by a 2-body + SOAP GAP model

$$E_{\text{bulk}} = \sum_{i \in \text{bulk}} \epsilon_i \qquad E_{\text{surf}} = \sum_{i \in \text{surf}} \epsilon_i$$

$$\gamma = \frac{E_{\text{surf}} - E_{\text{bulk}}}{2A}$$

$$V_i = K(\mathscr{R}_i, \mathscr{R}_i) - \mathbf{k}^T(\mathbf{K}_{MM} + \sigma_e^2 \mathbf{I})^{-1}\mathbf{k}$$

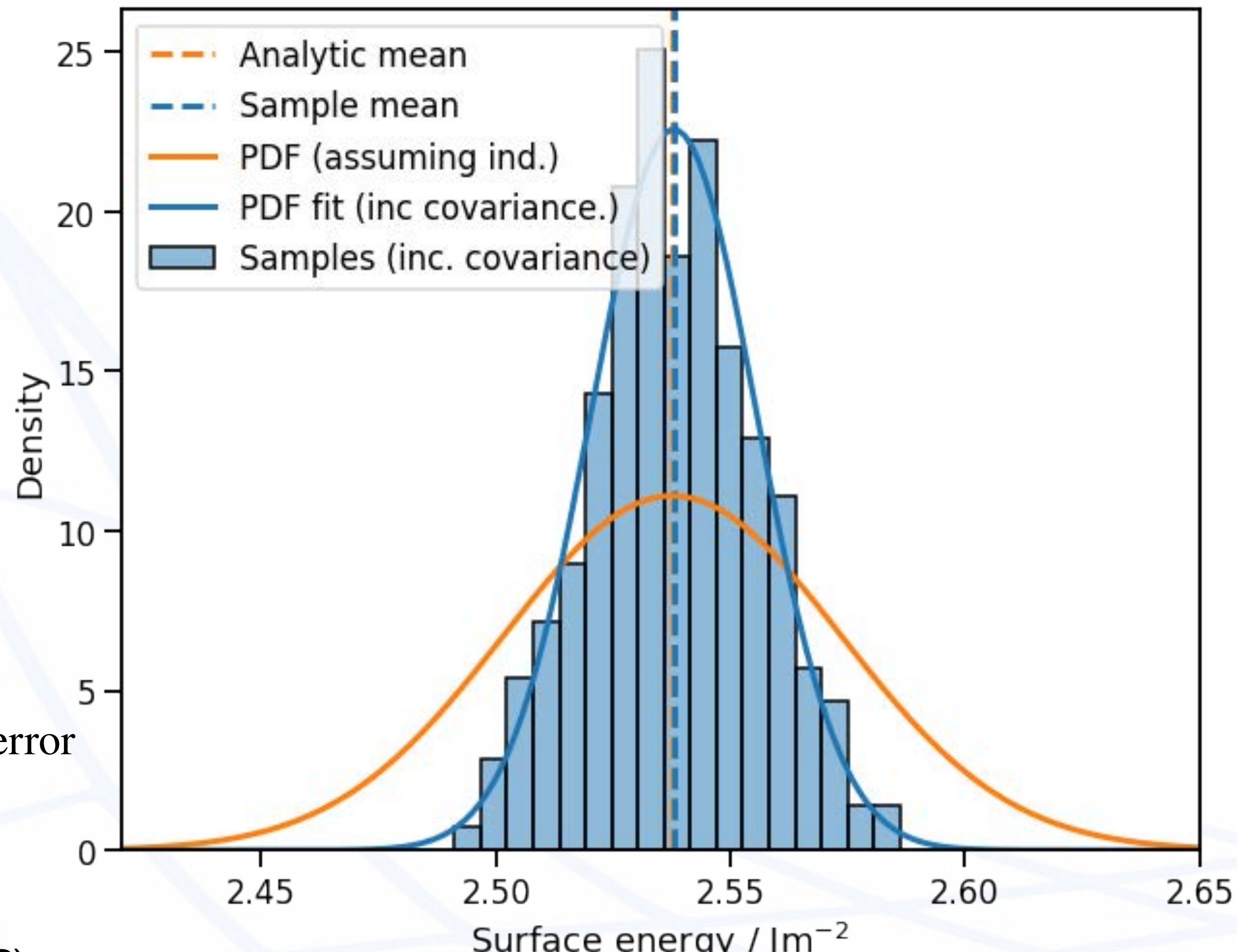
Attempt 1: Assume independence:

$$V_{\text{bulk}} = \sum_{i \in \text{bulk}} V_i \qquad V_{\text{surf}} = \sum_{i \in \text{surf}} V_i$$

$$V_\gamma = \frac{V_{\text{surf}} + V_{\text{bulk}}}{2A}$$

$$\gamma_{\text{GAP}} = 2.538 \pm 0.036 \text{ J/m}^2 \approx 1.4\,\% \text{ error}$$

$$\gamma_{\text{DFT}} = 2.543 \text{ J/m}^2 \implies 0.005 \text{ J/m}^2 \approx 0.2\,\% \text{ error}$$



Attempt 2: Assess effects of correlation by sampling models from GP posterior distribution (500 samples, 2-body+SOAP)

$$\gamma_{\text{GAP}} = 2.538 \pm 0.018 \text{ J/m}^2 \approx 0.7\,\% \text{ error}$$

L. Shenoy, A. P. Bartok and JRK (2021) – https://github.com/lakshenoy/PX915_UQ_Lakshmi
Based on Fe GAP database: D. Dragoni, T. Daff, G. Csányi and N. Marzari, Phys Rev Materials (2018)

Simplified GPR potential setup: Ar trimers

Dataset: 1921 CCSD(T) Ar dimers and trimers
Low-data limit: 1% train, 99% test

$$E(\mathbf{r}) = E_0 + \sum_{i<j} V_2(r_{ij}) + \sum_{i<j<k} V_3(r_{ij}, r_{ik}, r_{jk})$$

Include explicit basis set contrib in 2-body

$$\begin{aligned} V_2(r) &= f_2(\mathbf{r}) + \mathbf{h}^T(r)\boldsymbol{\beta} \\ f_2(r) &\sim \mathcal{GP}\left(0, k(r, r')\right) \\ V_3(\mathbf{r}) &\sim \mathcal{GP}\left(0, k_3(\mathbf{r}, \mathbf{r}')\right) \end{aligned}$$

Prior for coefficients $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{b}, B)$

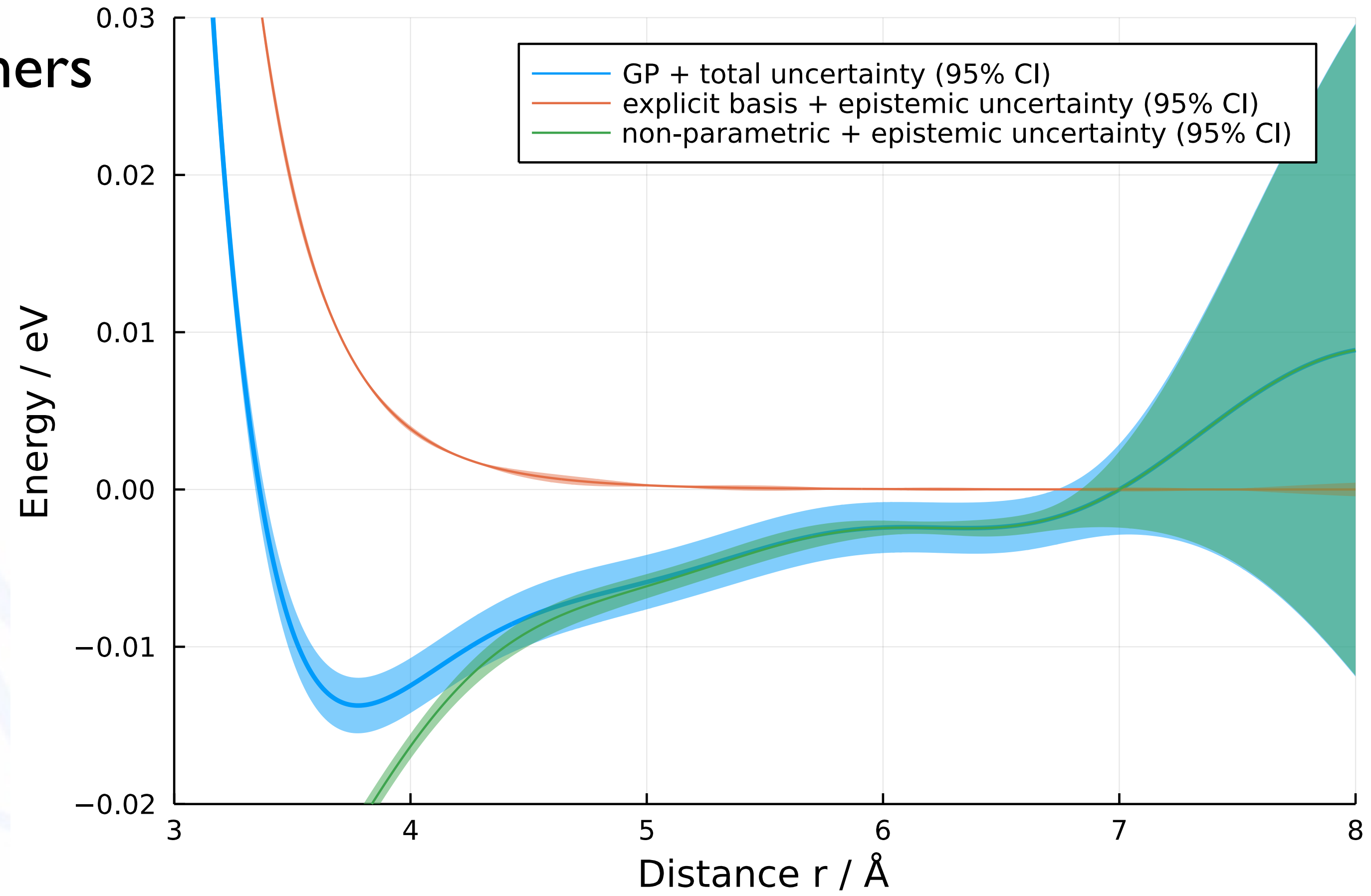
$$V_2(r) \sim \mathcal{GP}\left(\mathbf{h}(r)^T\mathbf{b}, k_2(r, r') + \mathbf{h}(r)^T B\mathbf{x}(r')\right)$$

Take vague prior limit $B^{-1} \to 0$
obtain predictive mean and cov, indpt. of $\mathbf{b}$

$$\bar{\mathbf{V}}_2(X_*) = \bar{\mathbf{f}}_2(X_*) + R^T\bar{\boldsymbol{\beta}}$$

$$\mathrm{cov}[\mathbf{V}_2^*] = \mathrm{cov}[\mathbf{f}_2^*] + R^T(HK_y^{-1}H^T)^{-1}R)$$

Design matrix $H = [\mathbf{h}(r_1), \mathbf{h}(r_2), \ldots]$
$R = H_* - HK_y^{-1}K_*$ and $\bar{\boldsymbol{\beta}} = (HK_y^{-1}H^T)^{-1}K_y^{-1}\mathbf{y}$



LJ-like basis set

$$h_1(r) = \frac{1}{r^{12}}$$

$$h_2(r) = \frac{1}{r^6}$$

2- and 3-body SE kernels

$$k_2(r, r') = \delta_2^2 \exp\left[\frac{|r - r'|^2}{\ell^2}\right]$$

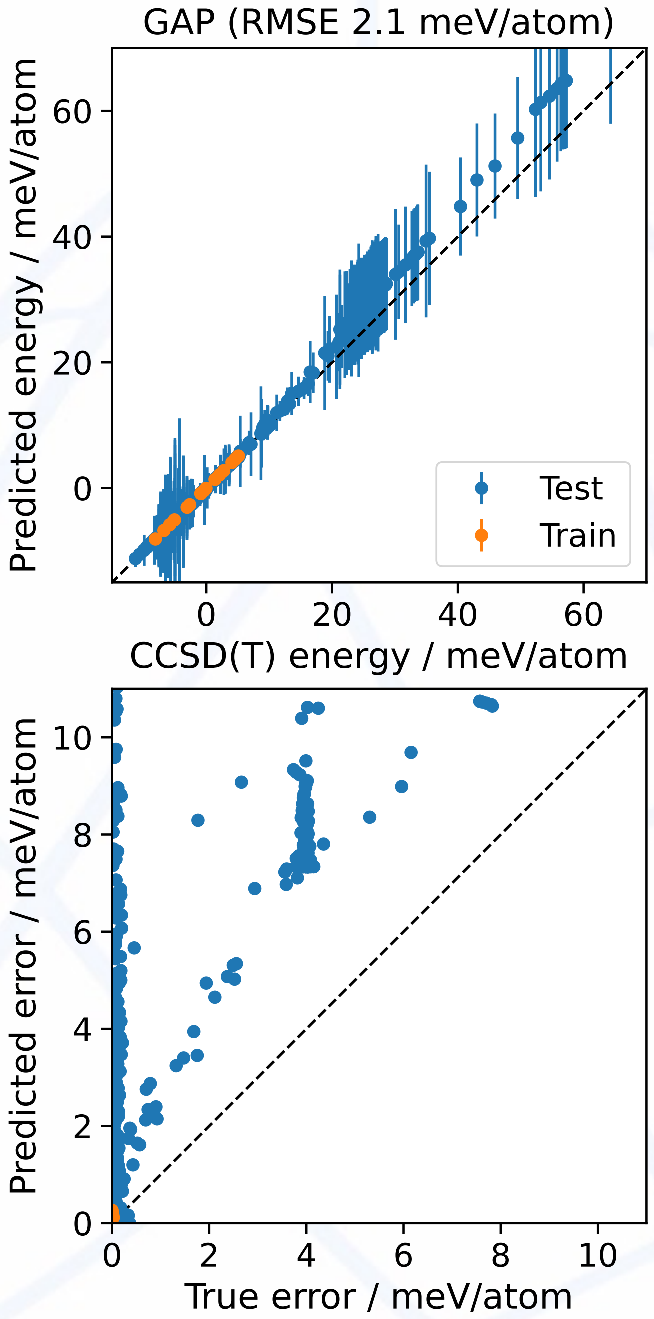
$$k_3(\mathbf{d}, \mathbf{d}') = \delta_3^2 \exp\left[\sum_{i=1}^{3} \frac{|d_i - d_i'|^2}{(\ell_3^i)^2}\right]$$

Gaussian Likelihood $E_i = \mathcal{N}(E(r), \sigma_n^2)$

Even for simple model have 7-dim. hyperspace
$\boldsymbol{\theta} = [\delta_2, \ell_2, \delta_3, \ell_3^1, \ell_3^2, \ell_3^3, \sigma_n$

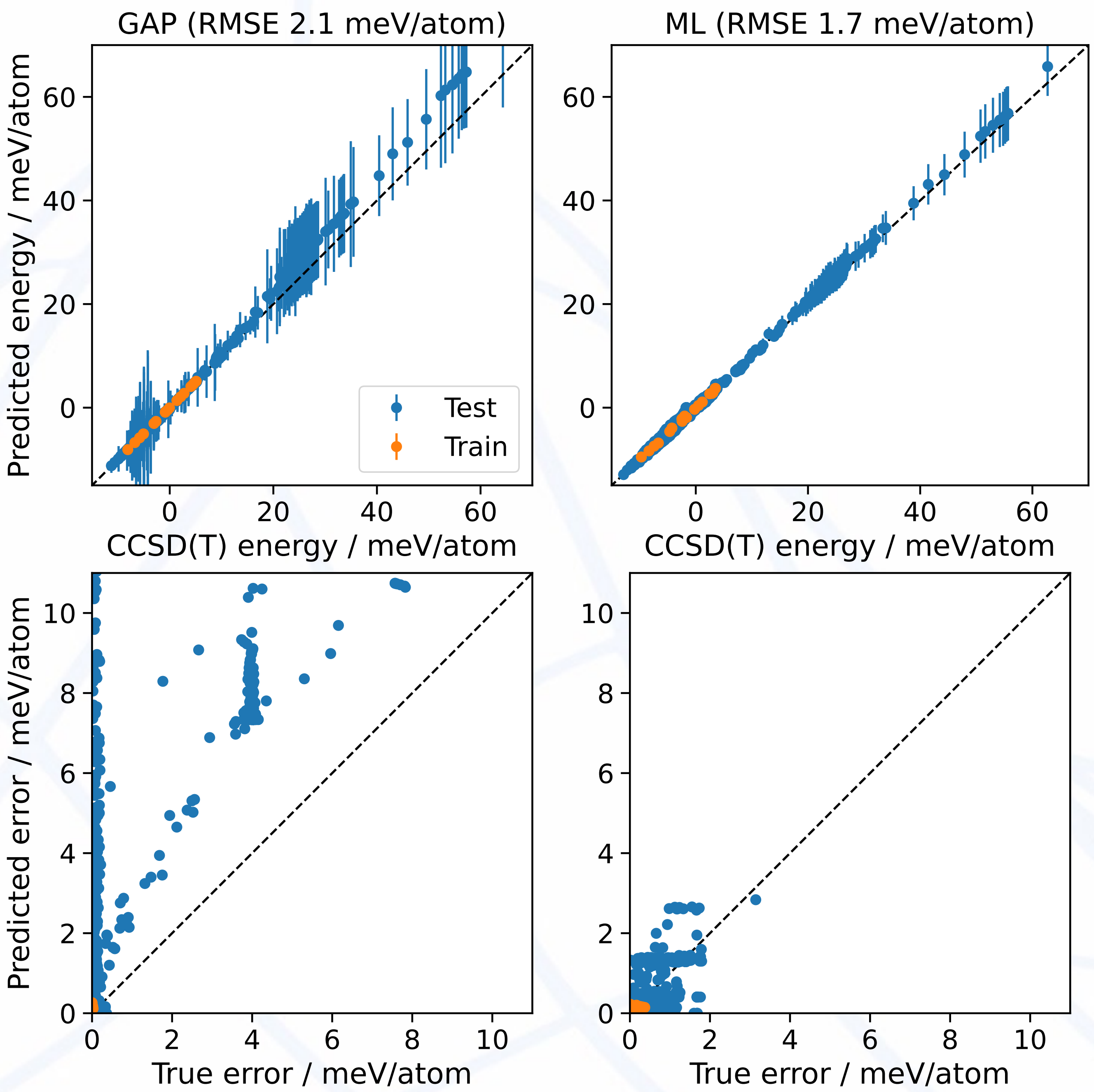
A.P. Bartok and JRK arXiv:2206.08744 (2022)

(i) Standard GAP heuristics for hyperparameters



A.P. Bartok and JRK arXiv:2206.08744 (2022)

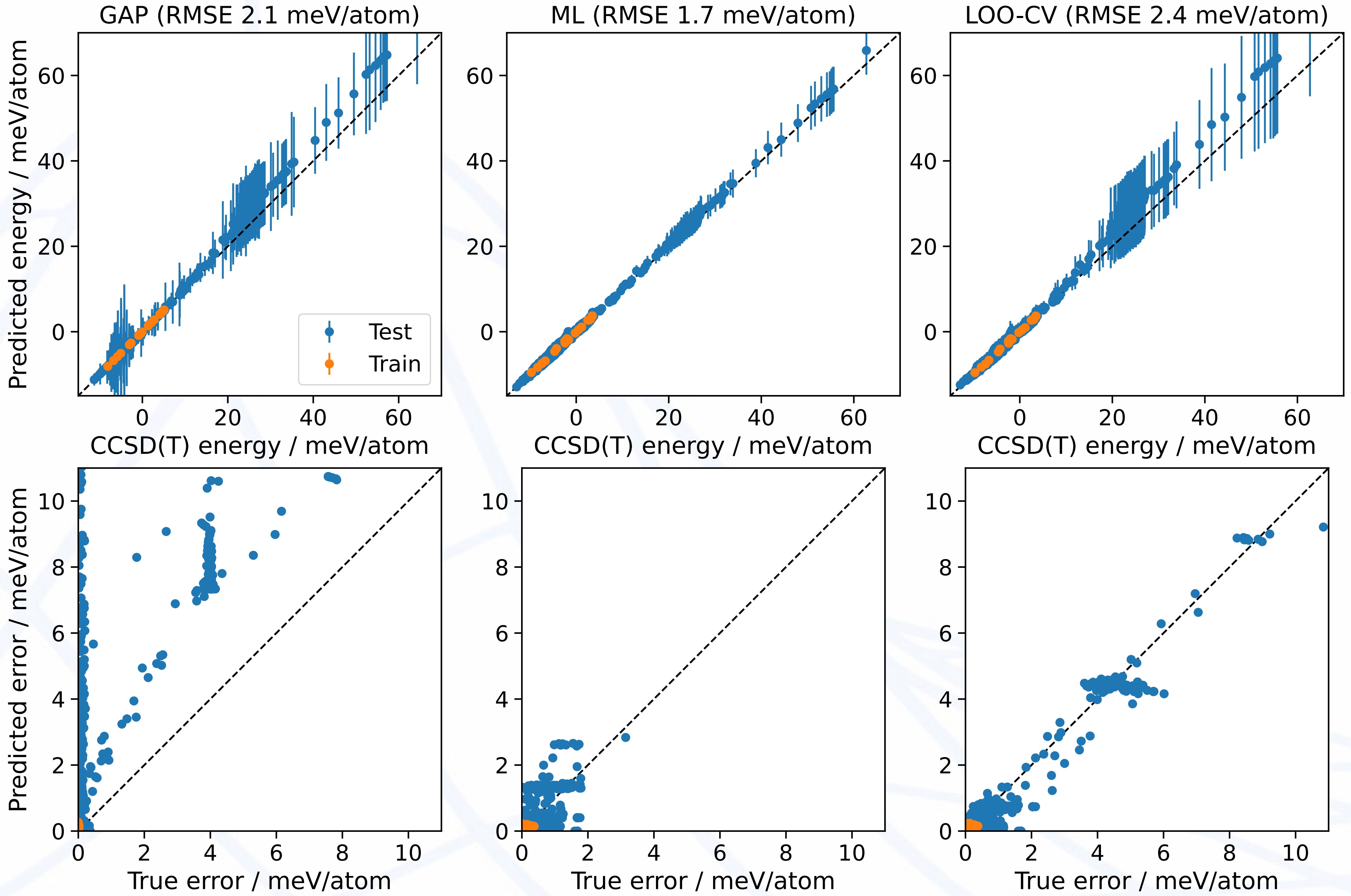
(ii) Optimise hyperparameters to maximise marginal likelihood



$$L(X, \mathbf{y}, \theta) = \log p(\mathbf{y}|X, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T K_y^{-1}\mathbf{y} + \frac{1}{2}\mathbf{y}^T C\mathbf{y} - \frac{1}{2}\log|A| - \frac{1}{2}\log|K_y| - \frac{N-M}{2}\log 2\pi$$

A.P. Bartok and JRK arXiv:2206.08744 (2022)

(iii) Optimise hyperparameters to maximise LOO-CV likelihood



$$L_{LOO}(X, \mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \log p(y_i | X, \mathbf{y}_{-i}, \boldsymbol{\theta})$$

A.P. Bartok and JRK arXiv:2206.08744 (2022)

Summary and Open Questions

- Statistical UQ methods promise to improve error estimates from data-driven models
- *But* we risk conflating epistemic (missing data/physics) and aleatoric (random) errors (cf. discussion group on combining numerical and statistical approaches)
- Gaussian likelihood appealing for practical reasons, but is it realistic for interatomic potential model form errors? Possible remedies:
  - Including explicit basis functions and their contributions to uncertainty
  - Improved description of model discrepancy (à la Kennedy-O'Hagan)
  - Gaussian → Student-t likelihood distribution
- Gaussian process regression predictive variance sensitive to hyperparameter choices:
  - Optimising marginal likelihood doesn't always improve calibration of prediction errors - perhaps because it relies on model assumptions being correct
  - Optimising LOO-CV likelihood is independent of model assumptions
  - *Ideally* MCMC over hypers sampled from suitable priors (or approx inf: VI, LFI)

Acknowledgements

Albert Bartok-Partay, Iain Best, Sascha Klawohn, Reinhard Maurer
Adam McSloy, Berk Onat, Lakshmi Shenoy, Tim Sullivan

Liwei Zhang and Christoph Ortner

Noam Bernstein

James Darby and Gábor Csányi

Geneviève Dusson (CNRS) and Gautam Anand (IIEST)

Financial support and supercomputing resources: