

Applications: Cluster and Force Constant Expansions

Paul Erhart (examiner)

Andreas Ekström

Bernhard Mehlig

Arkady Gonoskov



1

Plan for the next two weeks

- 1-2 Foundations: Linear models, Gaussian processes
- 3-4 Applications: Linear models, Regression, (Gaussian processes)
- 5-7 Neural networks and other ML techniques

L1: Advanced regression

L2: Applications

→ P2a: Regression with cluster expansions

L3: Sensitivity analysis

L4: Global optimization

→ P2b: GPs and Bayesian optimization

tif/fim{345}

Advanced Simulation and Machine Learning

5. Advanced regression

- 5.1. Ridge regression and beyond
- 5.2. Robust regression
- 5.3. Error correlation
- 5.4. Additional considerations
- 5.5. A brief detour into frequentist statistics
- 5.6. Feature selection and sparse models
- 5.7. Key take-aways

6. Toward applications

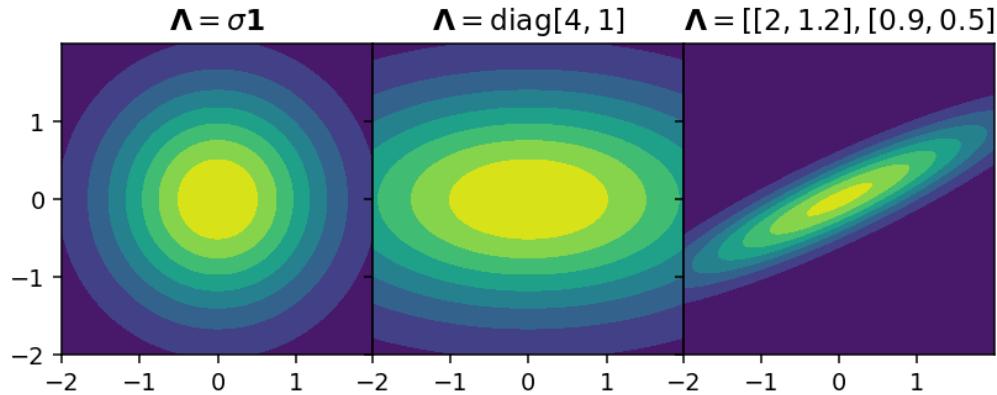
- 6.1. Alloy cluster expansions
- 6.2. Phonons and force constants
- 6.3. Interatomic potentials



Error and parameter correlations

$$\mathcal{L} = \hat{\varepsilon}^T \Sigma^{-1} \hat{\varepsilon} + \boldsymbol{\theta}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{\theta}$$

↑
Error information
↓ Model information



CHALMERS
UNIVERSITY OF TECHNOLOGY

3

Ridge regression and beyond

Ridge regression $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|0, \tau^2 \mathbf{I})$

Bayesian ridge regression $p(\mathcal{D}; \Phi, \boldsymbol{\theta}, \alpha) = \mathcal{N}(\mathcal{D}|\Phi\boldsymbol{\theta}, \alpha)$
 $p(\boldsymbol{\theta}|\lambda) = \mathcal{N}(\boldsymbol{\theta}|0, \lambda^{-1} \mathbf{I})$

α and λ from Gamma distributions
estimated jointly during fitting by maximizing LML

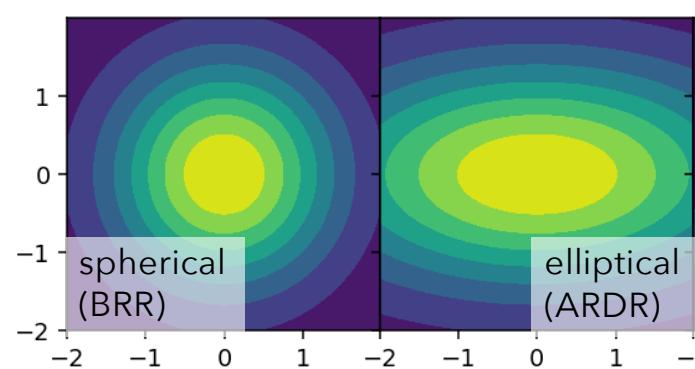
Automatic relevance detection (ARD) regression

$$p(\boldsymbol{\theta}|\boldsymbol{\Lambda}) = \mathcal{N}(\boldsymbol{\theta}|0, \boldsymbol{\Lambda}^{-1})$$

$$\text{diag}\boldsymbol{\Lambda} = [\lambda_1, \lambda_2 \dots \lambda_{N_p}]$$

Pruning step

$$\hat{\theta}_i = \begin{cases} \theta_i & \text{if } \lambda_i < \lambda_{\text{threshold}} \\ 0 & \text{else} \end{cases}$$



CHALMERS
UNIVERSITY OF TECHNOLOGY

(Automated) feature selection

"Embedded"-type

Algorithms that "take advantage of their own variable selection"

Automatic relevance detection (ARD) regression

$$p(\boldsymbol{\theta}|\boldsymbol{\Lambda}) = \mathcal{N}(\boldsymbol{\theta}|0, \boldsymbol{\Lambda}^{-1})$$

$$\text{diag}\boldsymbol{\Lambda} = [\lambda_1, \lambda_2 \dots \lambda_{N_p}]$$

Followed by pruning step

$$\hat{\theta}_i = \begin{cases} \theta_i & \text{if } \lambda_i < \lambda_{\text{threshold}} \\ 0 & \text{else} \end{cases}$$

Principle can be similarly applied to LASSO

Schematic of relation between methods

Method	Likelihood	Parameter prior	Hyperparameter prior
OLS	Gaussian	Uniform	n/a
Ridge	Gaussian	Gaussian	n/a
Bayesian ridge	Gaussian	Gaussian	spherical Gaussian
ARD	Gaussian	Gaussian	elliptical Gaussian
LASSO	Gaussian	Laplace	n/a
Robust regression	Laplace	Uniform	n/a

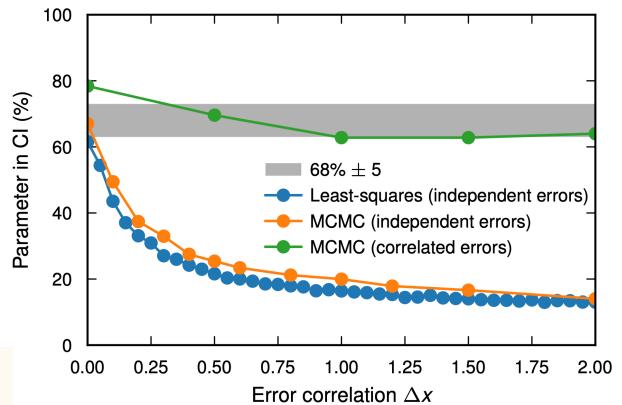
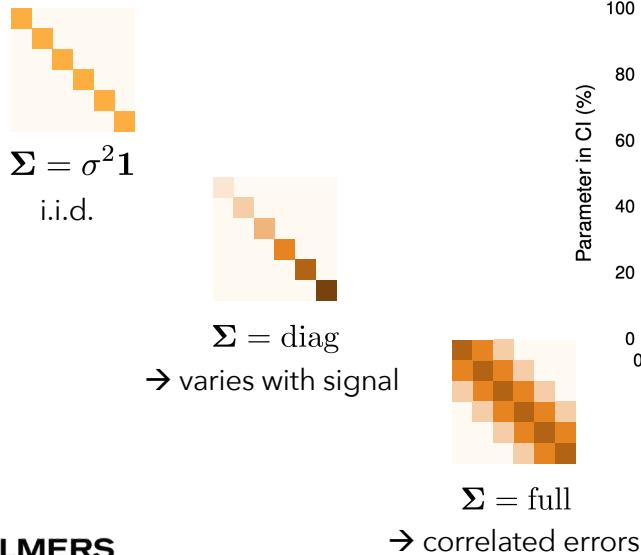
Error correlation

Generalize the variance = go beyond i.i.d. errors

$$\frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{\sigma^2} \rightarrow \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$$

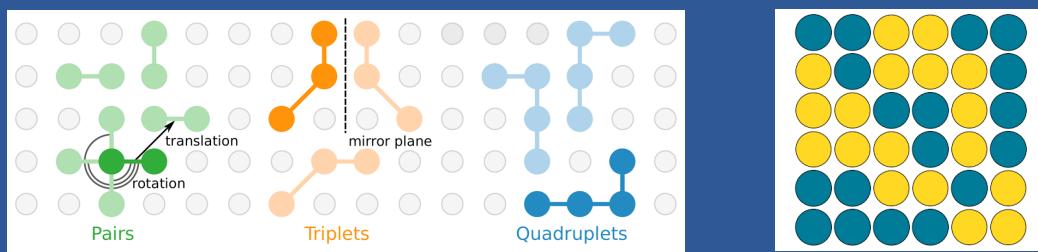
Covariance matrix

$$p(\mathcal{D}|\boldsymbol{\theta}, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{N_d/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right\}$$



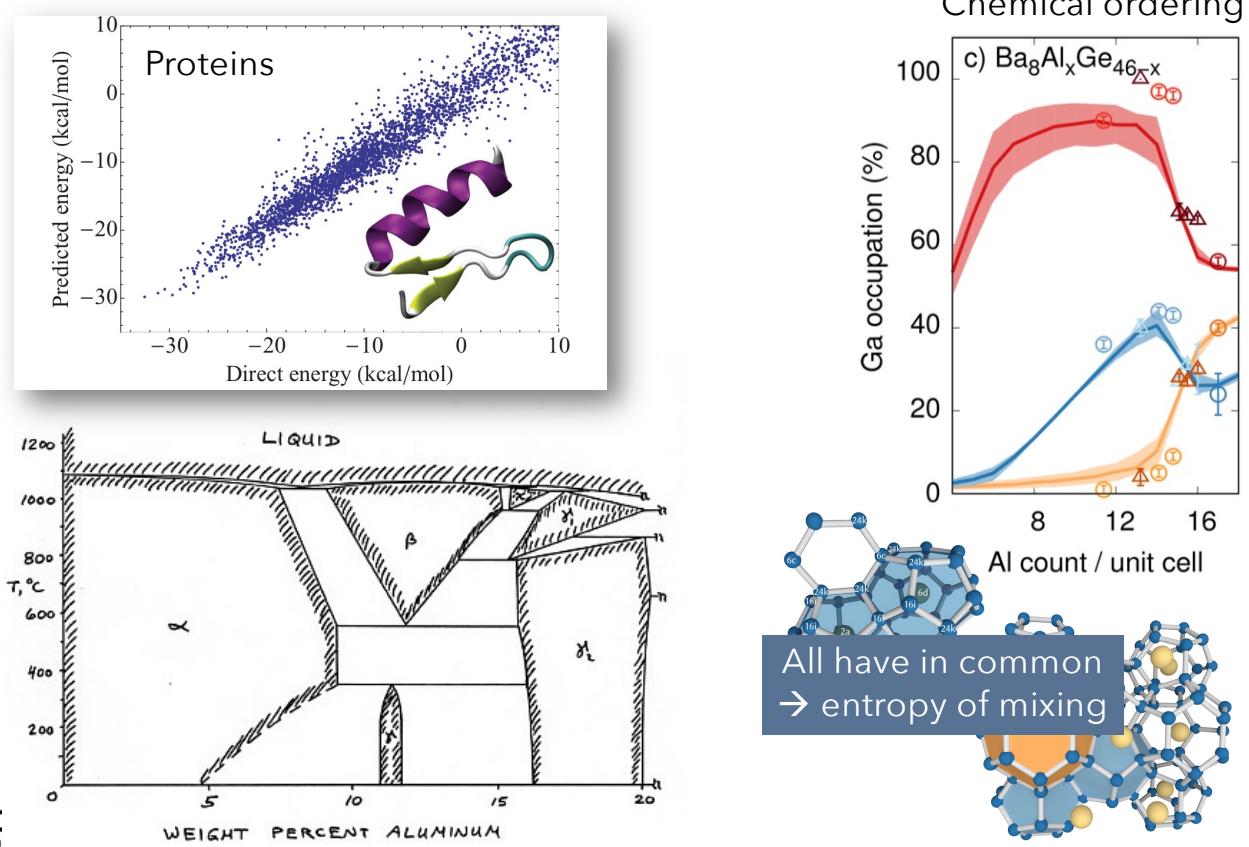
CHALMERS
UNIVERSITY OF TECHNOLOGY

7



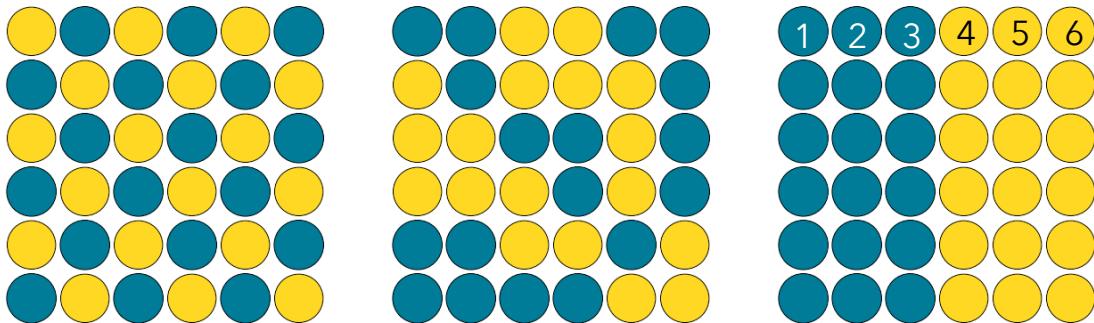
Cluster expansions:
Generalized linear models
for multi-component systems

What is the challenge?



9

Modelling multi-component systems



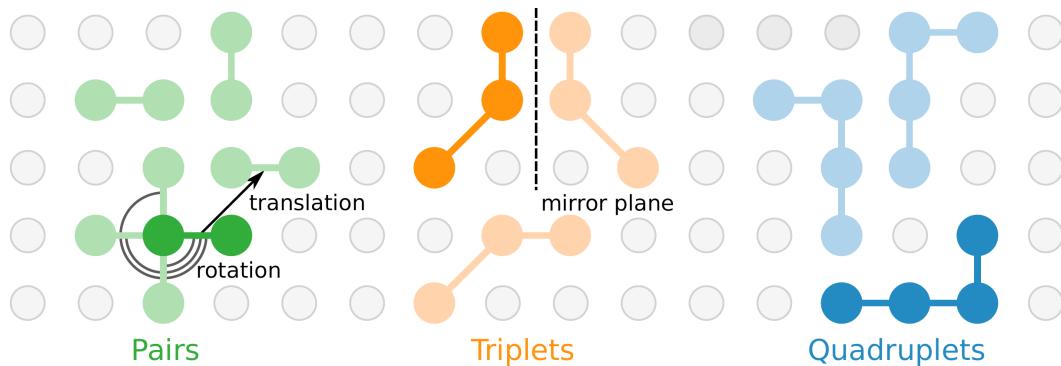
$$\sigma = [-1, -1, -1, +1, +1, +1 \dots]$$

Translational and symmetry variant representation

$$\rightarrow E_{\text{Ising}} = \frac{1}{N} \sum_{ij} J \sigma_i \sigma_j \quad \text{Ising model}$$

Generalized Ising model

Clusters → Orbits (group theory)



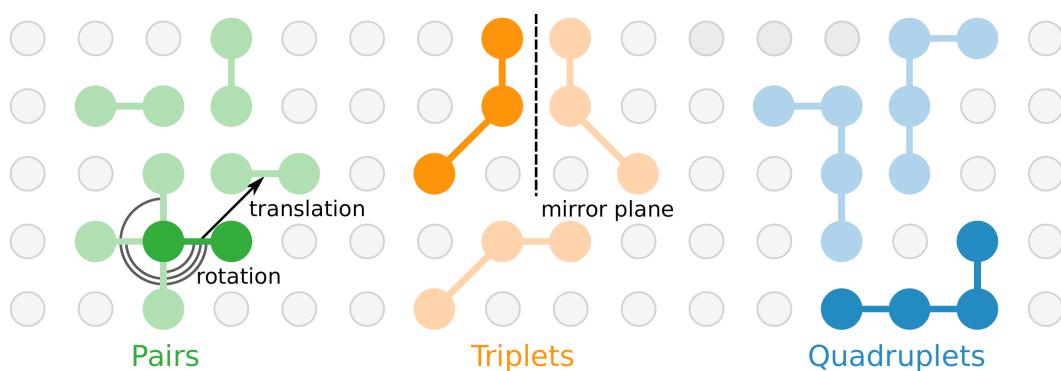
$$E_{\text{generalized Ising}} = \underbrace{\frac{1}{N} \sum_{ij}^{\text{1nn pairs}} J_{\text{pair-1}} \sigma_i \sigma_j}_{E_{\text{pair-1}}} + \underbrace{\frac{1}{N} \sum_{ij}^{\text{2nn pairs}} J_{\text{pair-2}} \sigma_i \sigma_j}_{E_{\text{pair-2}}} \\ + \underbrace{\frac{1}{N} \sum_{ijk}^{\text{1nn triplet}} J_{\text{triplet-1}} \sigma_i \sigma_j \sigma_k}_{E_{\text{triplet-1}}} + \dots$$

CHALMERS
UNIVERSITY OF TECHNOLOGY

11

Formalized alloy cluster expansions

Clusters → Orbits (group theory)



Sum over orbits

$$E_{\text{CE}} = J_0 + \sum_{\alpha} m_{\alpha} J_{\alpha} \Pi_{\alpha}(\sigma)$$

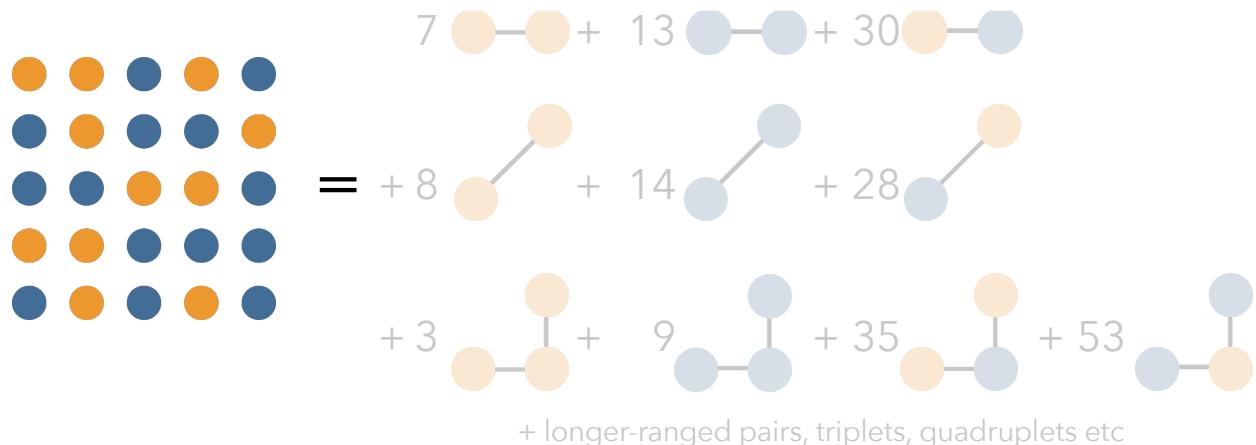
Multiplicity Averaged spin products

Effective cluster interaction (ECI)

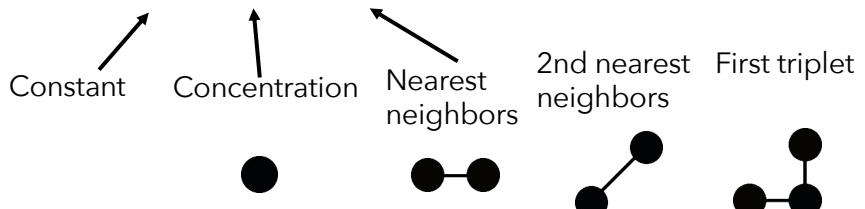
CHALMERS
UNIVERSITY OF TECHNOLOGY

12

A simple example



Cluster vector $\rightarrow = [1.0, 0.12, -0.2, -0.12, -0.12, \dots]$



CHALMERS
UNIVERSITY OF TECHNOLOGY

13

A simple example

Constant 0.12 -0.20 -0.12 -0.12 $J = \begin{bmatrix} J_1 \\ J_2 \\ J_3 \\ J_4 \\ J_5 \end{bmatrix}$

Solution: $J^T = [-37.8, 34.1, -9.9, 25.4, 2.67]$



CHALMERS
UNIVERSITY OF TECHNOLOGY

Alloy cluster expansions

Separate chemical and positional degrees of freedom

→ Map chemical degrees onto lattice
(relaxations are included implicitly)

$$E_{\text{CE}} = J_0 + \sum_{\alpha} m_{\alpha} J_{\alpha} \Pi_{\alpha}(\sigma)$$

Cluster expansion can be cast as a linear problem

$$\mathbf{E} = \mathbf{X} \mathbf{J}$$

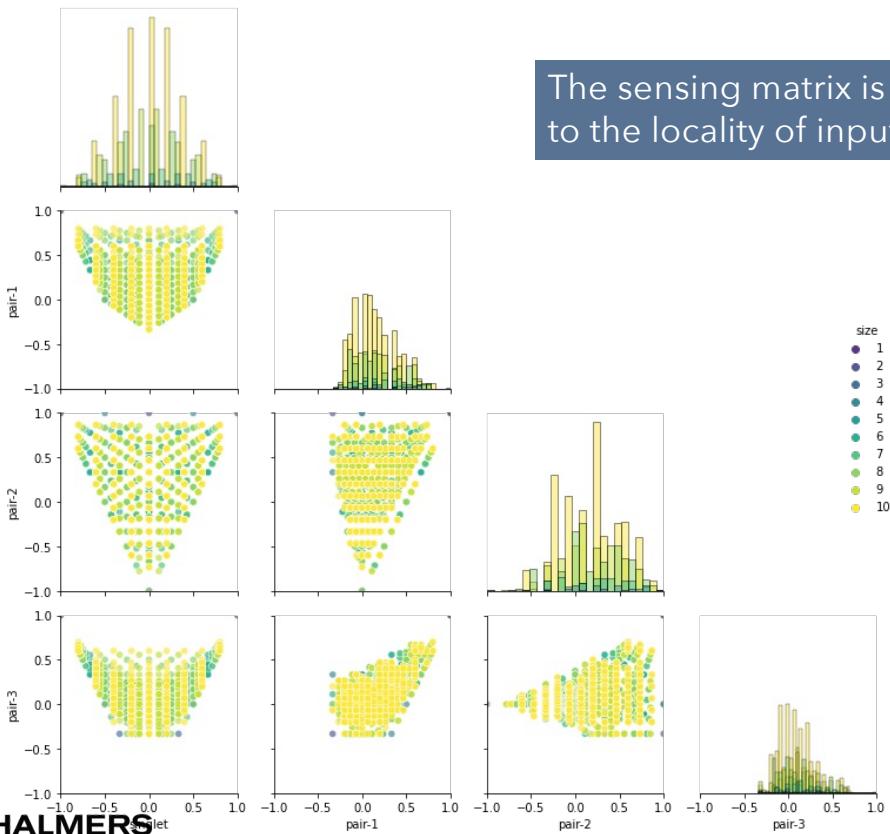


\mathbf{X} Matrix of cluster vectors (sensing matrix)

\mathbf{J} Effective cluster interactions

\mathbf{E} Target energies (usually from DFT)

Correlations in the sensing matrix



Incorporating physical insight

Insight 1: interactions decay with distance and the number of bodies

$$p(\mathbf{J}|\mathbf{X}) \sim \prod_{\alpha} \exp\left(-\frac{J_{\alpha}^2}{2\sigma_{\alpha}^2}\right)$$

Insight 2: similar orbits should have similar ECIs

$$p(\mathbf{J}|\mathbf{X}) \sim \prod_{\alpha, \beta \neq \alpha} \exp\left(-\frac{(J_{\alpha} - J_{\beta})^2}{2\sigma_{\alpha\beta}^2}\right)$$



$$\mathbf{J}_{\text{opt}} = (\mathbf{X}^T \mathbf{X} + \boldsymbol{\Lambda}^{-1}) \mathbf{X}^T \mathbf{E}$$

$$\begin{aligned}\boldsymbol{\Lambda}_{\alpha,\alpha} &= \frac{\sigma^2}{\sigma_{\alpha}^2} + \sum_{\beta | \beta \neq \alpha} \frac{\sigma^2}{\sigma_{\alpha\beta}} \\ \boldsymbol{\Lambda}_{\alpha\beta} &= \boldsymbol{\Lambda}_{\beta\alpha} = -\frac{\sigma^2}{\sigma_{\alpha\beta}^2}.\end{aligned}$$

Incorporating physical insight

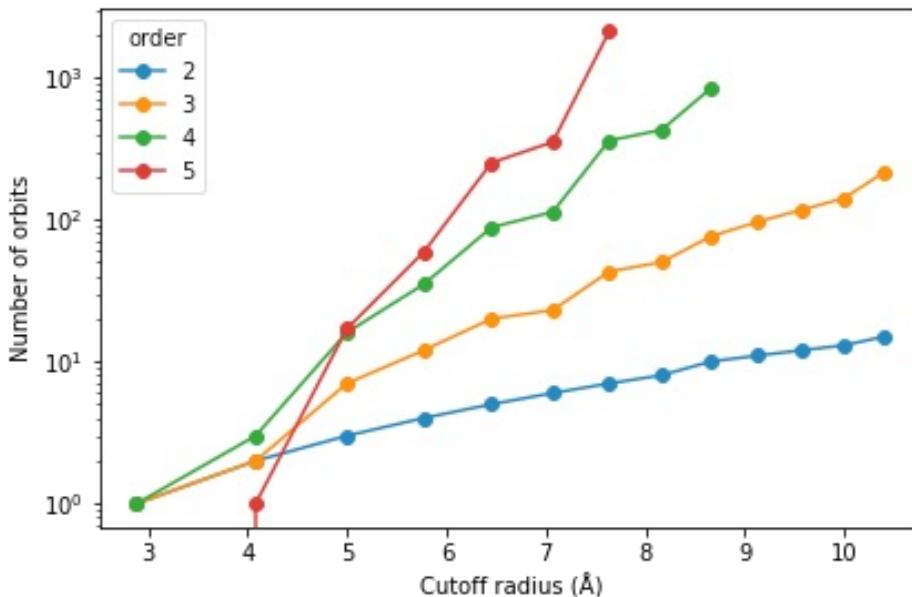
$$\begin{aligned}\boldsymbol{\Lambda}_{\alpha,\alpha} &= \frac{\sigma^2}{\sigma_{\alpha}^2} + \sum_{\beta | \beta \neq \alpha} \frac{\sigma^2}{\sigma_{\alpha\beta}} & \lambda_{\alpha} &= \sigma^2 / \sigma_{\alpha}^2 \\ \boldsymbol{\Lambda}_{\alpha\beta} &= \boldsymbol{\Lambda}_{\beta\alpha} = -\frac{\sigma^2}{\sigma_{\alpha\beta}^2}. & \lambda_{\alpha\beta} &= \sigma^2 / \sigma_{\alpha\beta}^2\end{aligned}$$

1. $\lambda_{\alpha} \rightarrow \infty$: force **ECI** for orbit α to be zero (i.e. remove the orbit from the **CE**)
2. $\lambda_{\alpha} \rightarrow 0$ and $\lambda_{\alpha\beta} \rightarrow 0$: all **ECI** values are equally likely; this recovers **OLS**
3. $\lambda_{\alpha\beta} \rightarrow 0$: no correlation (coupling) between **ECIs**; this recovers ridge regression if λ_{α} is the same for all orbits
4. $\lambda_{\alpha\beta} \rightarrow \infty$: force two orbits to have the same **ECI**

Cluster space dimensionality

Orbit count increases exponentially with cutoff and order

$$n_{\text{orbits}} \sim r_{\text{cut}}^{\gamma n}$$



Demo
(surface)

CHALMERS
UNIVERSITY OF TECHNOLOGY

19

Finding the ECIs is a linear problem

General procedure

1. Compile structures into one fit matrix \mathbf{X}
2. Solve the linear problem $\mathbf{X}\mathbf{J} = \mathbf{E}$

$$\mathbf{X} \quad \mathbf{J} \quad \mathbf{E}$$

Structures Parameters Energies

$$N_{\text{ECIs}}$$

$$\left[\begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{array} \right] \quad \bullet \quad \left[\begin{array}{c} \mathbf{j}_1 \\ \mathbf{j}_2 \\ \mathbf{j}_3 \end{array} \right] = \left[\begin{array}{c} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \end{array} \right]$$

$N_{\text{structures}}$

Least-squares (OLS)

$$\mathbf{J}_{\text{opt}} = \arg \min_{\mathbf{J}} \|\mathbf{X}\mathbf{J} - \mathbf{E}\|_2$$

LASSO

$$\mathbf{J}_{\text{opt}} = \arg \min_{\mathbf{J}} \|\mathbf{X}\mathbf{J} - \mathbf{E}\|_2 + \lambda_1 \|\mathbf{J}\|_1$$

Ridge regression

$$\mathbf{J}_{\text{opt}} = \arg \min_{\mathbf{J}} \|\mathbf{X}\mathbf{J} - \mathbf{E}\|_2 + \lambda_2 \|\mathbf{J}\|_2$$

Bayesian

$$p(\mathbf{E} | \mathbf{J}, \mathbf{X}, \sigma^2) \propto \exp \left(-\frac{1}{2\sigma^2} (\mathbf{X}\mathbf{J} - \mathbf{E})^2 \right)$$

CHALMERS
UNIVERSITY OF TECHNOLOGY

20

Low symmetry: Bayesian regression

Solving the linear CE system

Least-squares (OLS)

$$\mathbf{J}_{\text{opt}} = \arg \min_{\mathbf{J}} \|\mathbf{X}\mathbf{J} - \mathbf{E}\|_2 \longrightarrow \mathbf{J}_{\text{opt}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E}$$

Ridge regression

$$\mathbf{J}_{\text{opt}} = \arg \min_{\mathbf{J}} \|\mathbf{X}\mathbf{J} - \mathbf{E}\|_2 + \lambda_2 \|\mathbf{J}\|_2 \longrightarrow \mathbf{J}_{\text{opt}} = (\mathbf{X}^T \mathbf{X} + \lambda_2 \mathbf{1})^{-1} \mathbf{X} \mathbf{E}$$

General case [1]

$$\mathbf{J}_{\text{opt}} = (\mathbf{X}^T \mathbf{X} + \Lambda)^{-1} \mathbf{X}^T \mathbf{E}$$

Inverse covariance matrix

```
sc = StructureContainer(cs)
for atoms in list_of_training_structures:
    sc.add_structure(atoms)
X, E = sc.get_fit_data()
```

Impose physical insight through priors

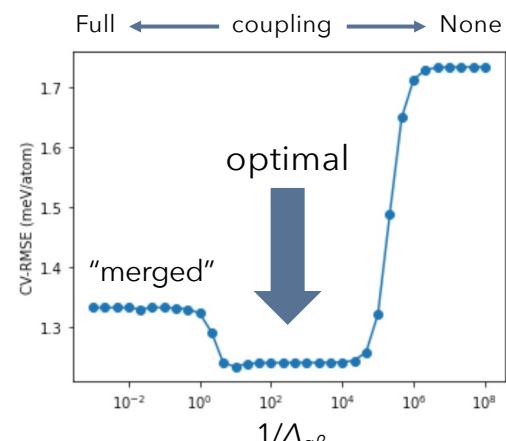
Insight 2: ECIs of similar orbits should be similar → off-diagonal terms

- $\Lambda_{\alpha\beta} \rightarrow 0$ no coupling between α and β
- $\Lambda_{\alpha\beta} \rightarrow \infty$ force ECIs for α and β to be identical

Simple example:
All off-diagonal terms identical

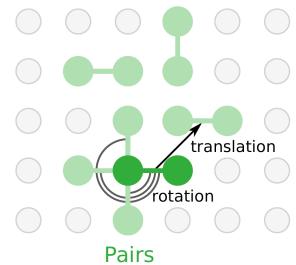
Some possible strategies

- Manually force ECIs to be identical → merging approach
- Use similarity measure to set $\Lambda_{\alpha\beta}$
- Couple ECIs to bulk values, e.g., in a surface case



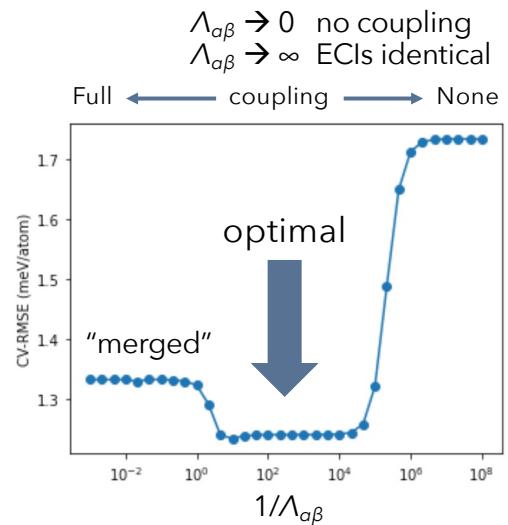
Take aways

- Cluster expansions: generalized linear models for representing multi-component systems
→ scalar quantities!
- Components of cluster vector are correlated
- Parameters are correlated



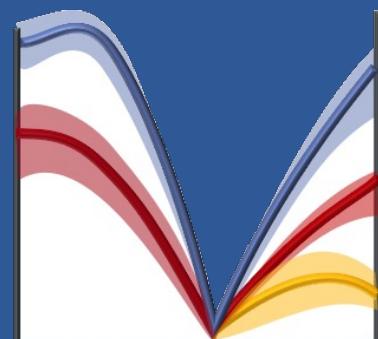
Some possible strategies

- OLS
- ARDR
- RFE
- "manual" design of covariance matrix



CHALMERS
UNIVERSITY OF TECHNOLOGY

23

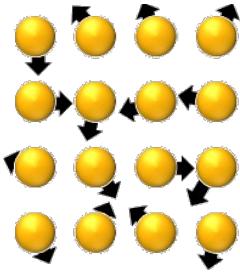


Force constant expansions:
Generalized linear models
connected to phonons

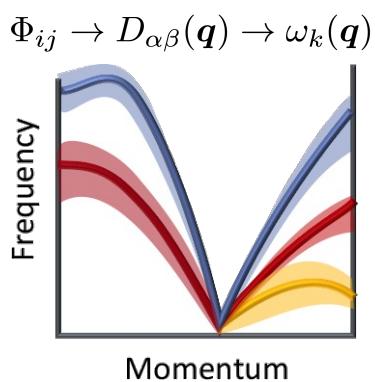
Force constant expansions

Taylor expansion of the total energy

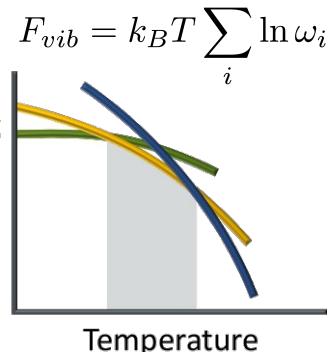
$$E = E_0 + \sum_i \Phi_i u_i + \frac{1}{2!} \sum_{ij} \Phi_{ij} u_i u_j + \frac{1}{3!} \sum_{ijk} \Phi_{ijk} u_i u_j u_k \dots$$



Phonon dispersions



Free energies



Need efficient means to extract Φ matrices

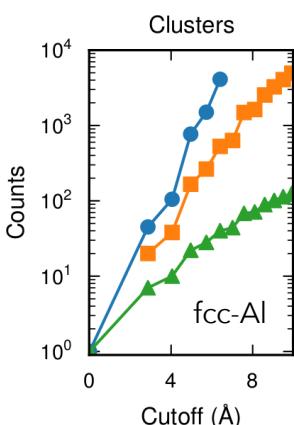
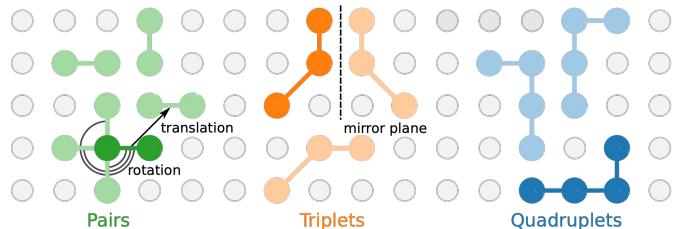
CHALMERS
UNIVERSITY OF TECHNOLOGY

27

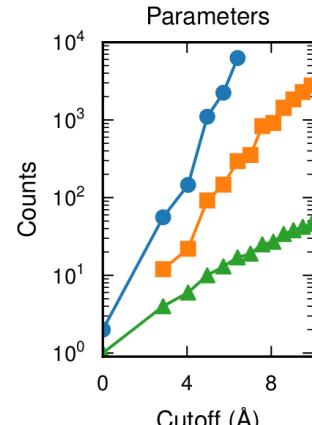
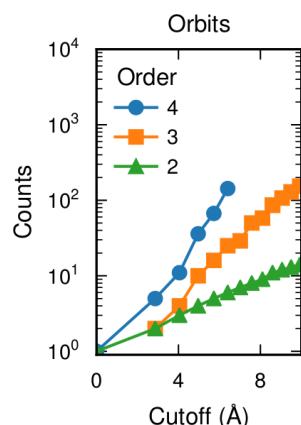
Step 1: Cluster decomposition

General procedure

1. Generate list of clusters
2. Obtain symmetry operations
3. Categorize clusters into orbits
4. Identify independent parameters
5. Apply sum rules



Number of clusters increases rapidly with order and size

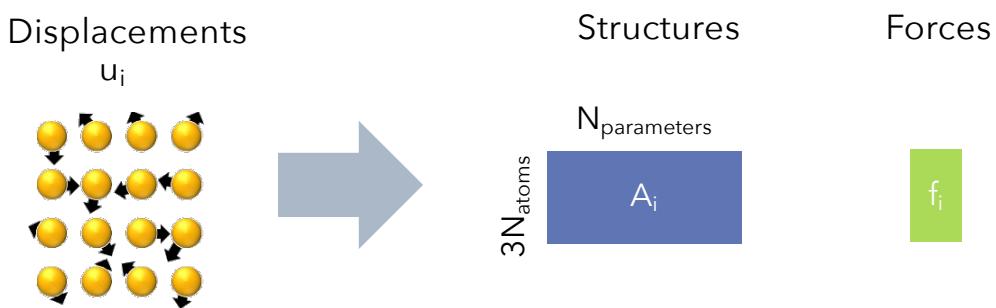


CHALMERS
UNIVERSITY OF TECHNOLOGY

Step 2: Structure decomposition

General procedure

1. Generate structure e.g., by applying random displacements or superposing normal modes
2. Convert displacements u_i into cluster vectors
→ each structure yields a matrix A_i with $N_{\text{parameters}}$ columns and $3N_{\text{atoms}}$ rows
3. Each structure comes with a target force vector



CHALMERS
UNIVERSITY OF TECHNOLOGY

29

FC expansions also a linear problem

Demo

General procedure

1. Compile multiple structures into one (large) fit matrix \mathbf{X}
2. Solve the linear problem $\mathbf{X} \mathbf{p} = \mathbf{f}$

$$\mathbf{X} \quad \mathbf{p} \quad \mathbf{f}$$

Structures	Parameters	Forces
$N_{\text{parameters}}$		
x_1		f_1
x_2		f_2
x_3		f_3

\mathbf{X} is a $3N_{\text{atoms}} \times N_{\text{structures}}$ matrix. \mathbf{p} is a $N_{\text{structures}} \times 1$ vector. \mathbf{f} is a $3N_{\text{atoms}} \times 1$ vector.

How?

Least-squares (OLS)

$$p_{\text{opt}} = \arg \min_p \|\mathbf{X}p - \mathbf{f}\|_2$$

LASSO

$$p_{\text{opt}} = \arg \min_p \|\mathbf{X}p - \mathbf{f}\|_2 + \lambda_1 \|\mathbf{f}\|_1$$

Ridge regression

$$p_{\text{opt}} = \arg \min_p \|\mathbf{X}p - \mathbf{f}\|_2 + \lambda_2 \|\mathbf{f}\|_2$$

Automatic relevance detection (ARD)

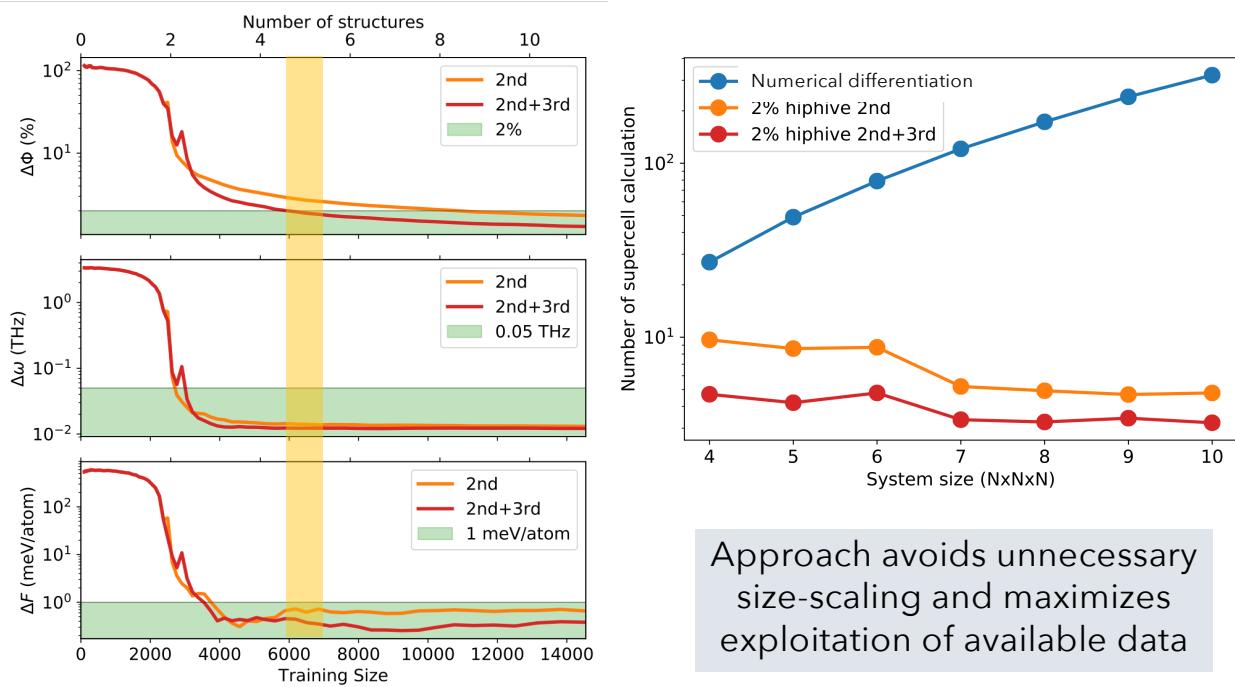
Recursive feature selection (RFE)

CHALMERS
UNIVERSITY OF TECHNOLOGY

20

OLS: Vacancy in BCC-Ta (2nd order)

Convergence with training size

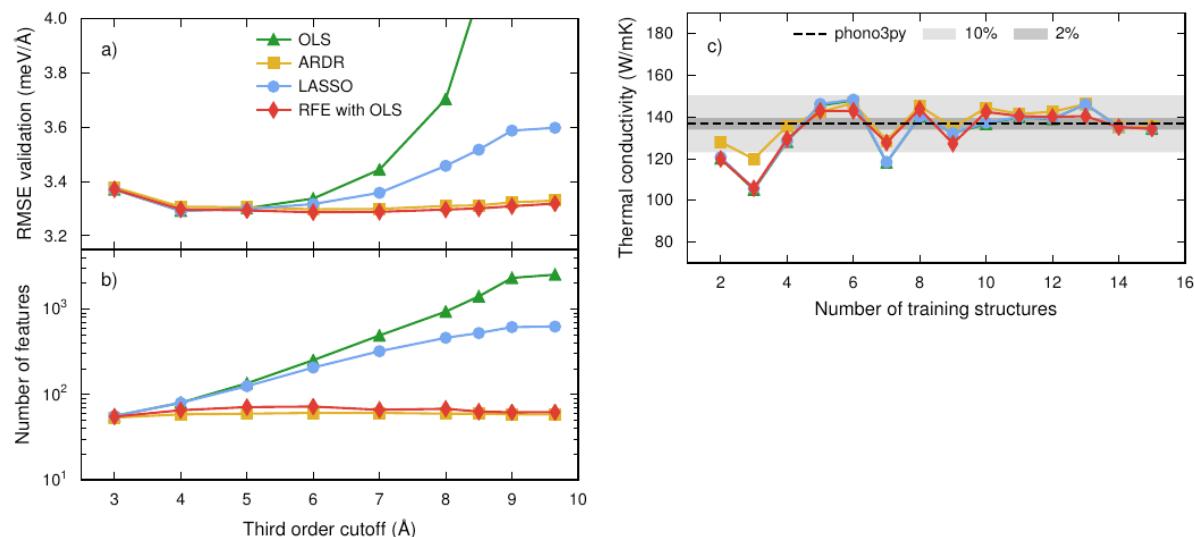


CHALMERS
UNIVERSITY OF TECHNOLOGY

31

Manual vs automated feature selection

Manual selection (via cutoff)



3rd order model for Si

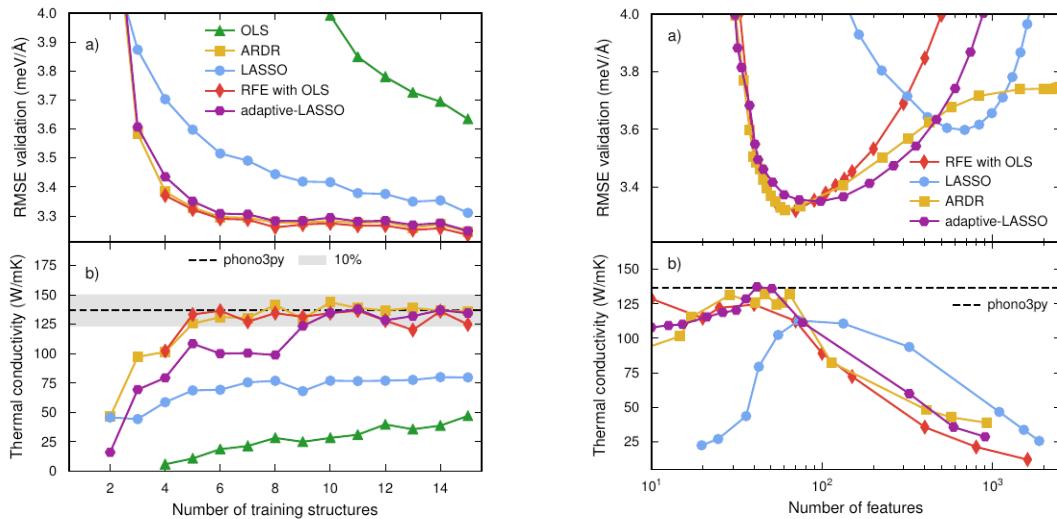
CHALMERS
UNIVERSITY OF TECHNOLOGY

Fransson *et al.*, npg Comp. Mater. **6**, 135 (2020)

22

Manual vs automated feature selection

Automated selection



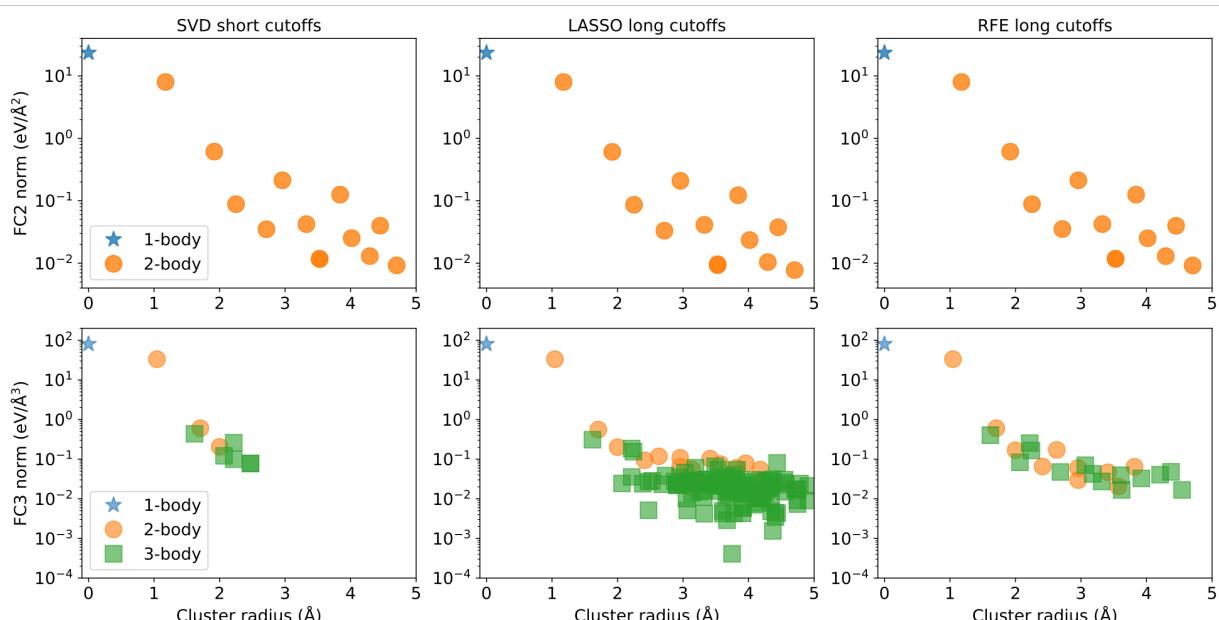
- LASSO overselects (known) and fails at recovering derived property
- ARDR and RFE perform by far the best (reliable derived properties)
- CV-RMSE poor predictor for quality of derived property

CHALMERS
UNIVERSITY OF TECHNOLOGY

Fransson *et al.*, npg Comp. Mater. **6**, 135 (2020)

33

Efficiency: Thermal conductivity of Si

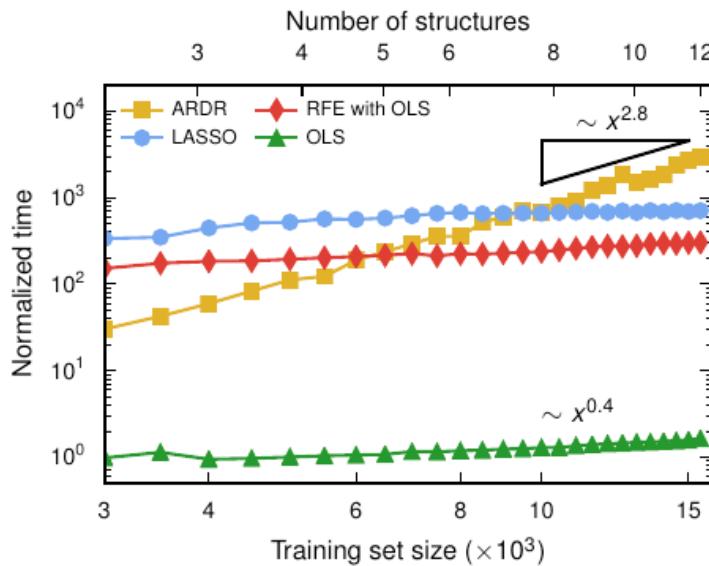


2nd-order FCs independent of method
3rd-order FCs much more sensitive
→ explore e.g., sequential training

CHALMERS
UNIVERSITY OF TECHNOLOGY

24

Cost analysis



- Cost becomes significant for large problems (FC not CE)
- ARDR (Bayesian) becomes too costly due to scaling
- RFE or OLS with manual selection often the most performant (*bummer*)

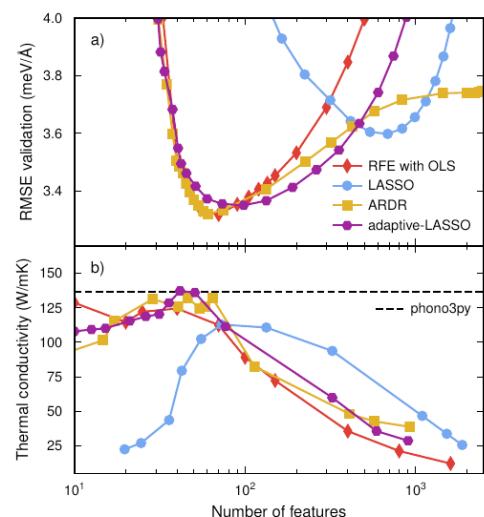
CHALMERS
UNIVERSITY OF TECHNOLOGY

Fransson *et al.*, npg Comp. Mater. **6**, 135 (2020)

35

Take aways

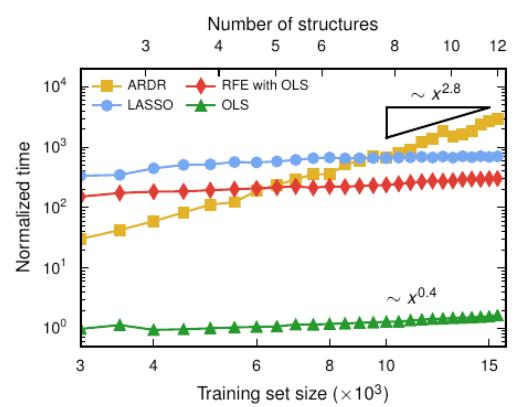
- Force constant expansions: generalized linear models for representing forces in crystalline systems
- Components of cluster vector are correlated
- Parameters are correlated



Some possible strategies

- OLS
- RFE

Be mindful of information density
Judge data cost vs extraction cost



CHALMERS
UNIVERSITY OF TECHNOLOGY

26



icet Workflow

