



CHALMERS  
UNIVERSITY OF TECHNOLOGY

# TIF345 / FYM345 Advanced Simulation and Machine Learning

Paul Erhart

Andreas Ekström

Bernhard Mehlig

Arkady Gonoskov



<https://tif345.materialsmodeling.org>

1

## Plan for the next two weeks

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

L1: Cluster expansions (CEs) and using the covariance matrix

L2: Feature selection algorithms

→ P2a: Regression with CEs

L3: Sensitivity analysis

L4: Global optimization

→ P2b: GPs and Bayesian optimization

## tif/fim{345}

Advanced Simulation and Machine Learning

5. Advanced regression ▾

6. Toward applications ▾

6.1. Alloy cluster expansions

6.2. Phonons and force constants

6.3. Interatomic potentials

7. Compressive sensing ▾

8. Sensitivity analysis

9. Global optimization ▾

CHALMERS  
UNIVERSITY OF TECHNOLOGY

Demos



Erhart

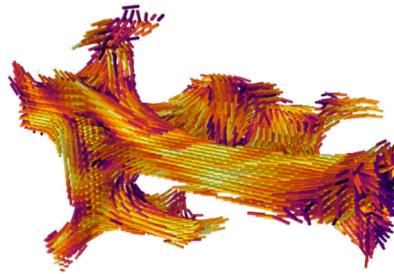
# Practical challenges

(Very) large **data** sets  
Incomplete data sets

(Very) large **parameter** spaces

Errors / **noise**

Potentially functional forms  
that are numerically problematic ("sloppy models")



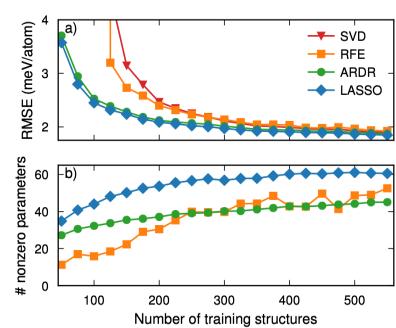
## Consequences

- Direct MCMC often impractical
- No general closed solutions for non-linear problems
- Model design and problem formulation important

## Follow-up questions (1)

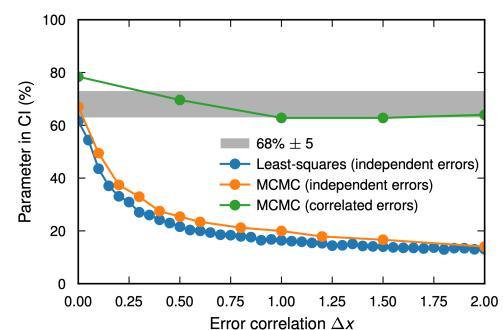
- How do we find out which parameters matter?  
→ sparse models via regularization
- How do we do this efficiently while accounting meaningfully for prior information?  
→ loss function

$$\mathcal{L} = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} + \lambda_1 \|\boldsymbol{\theta}\|_1 + \lambda_2 \|\boldsymbol{\theta}\|_2^2 + \lambda_3 \|\nabla^2 \boldsymbol{\theta}\|_2^2 + \dots$$



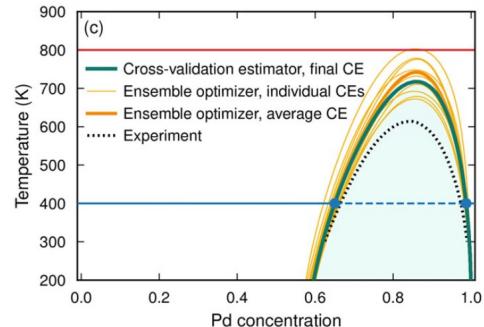
- How we handle error correlation?  
→ "Sigma"

$$\mathcal{L} = \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} + \dots$$

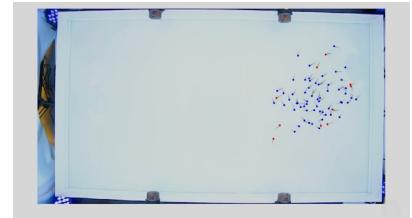


## Follow-up questions (2)

- How can we quantify model uncertainty?  
→ Sensitivity analysis and error propagation  
(e.g., bagging, committee/ensemble models, Mahalanobis norm),



- How do we acquire data?  
→ active learning



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

5

## Cluster expansions

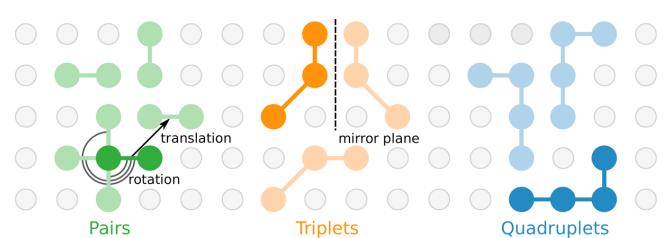
- Cluster expansions: generalized linear models for representing multi-component systems

$$E_{\text{general Ising}} = \sum_{ij}^{\text{1nn pairs}} J_{1p} \sigma_i \sigma_j + \sum_{ij}^{\text{2nn pairs}} J_{2p} \sigma_i \sigma_j \dots$$

↓

$$+ \sum_{ijk}^{\text{1nn triplets}} J_{1t} \sigma_i \sigma_j \sigma_k \dots$$

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



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Erhart

# 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
$N_{\text{ECIs}}$		
$x_1$		$E_1$
$x_2$		$E_2$
$x_3$		$E_3$

$\left. \right\} N_{\text{structures}}$

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

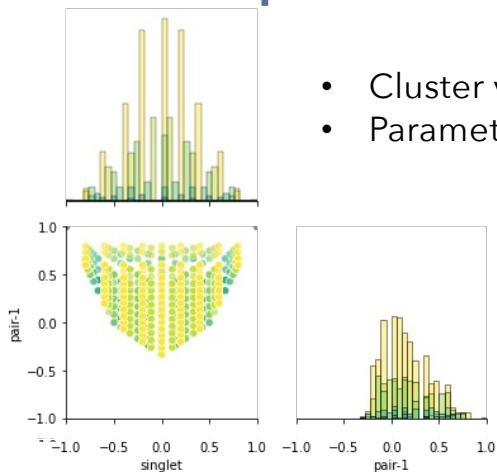
$$\mathbf{J}_{\text{opt}} = \arg \min_{\mathbf{J}} \|\mathbf{X}\mathbf{J} - \mathbf{E}\|_2 + \lambda_2 \|\mathbf{J}\|_2$$

Bayesian

$$p(\mathbf{E} | \mathbf{J}, \mathbf{X}, \sigma^2)$$

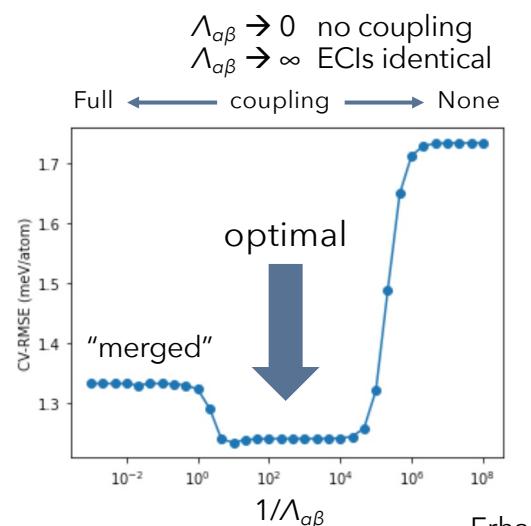
$$\propto \exp\left(-\frac{1}{2\sigma^2}(\mathbf{X}\mathbf{J} - \mathbf{E})^2\right)$$

## Cluster expansions



"Manual" design of covariance matrix

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



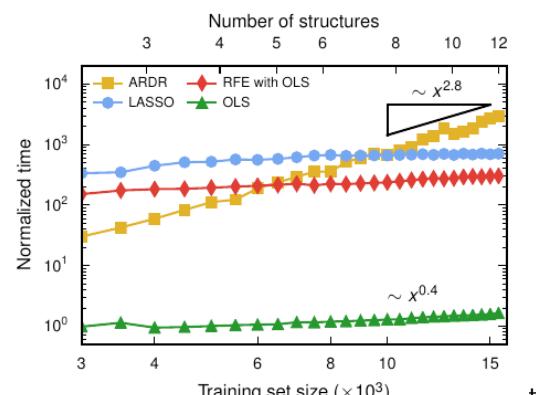
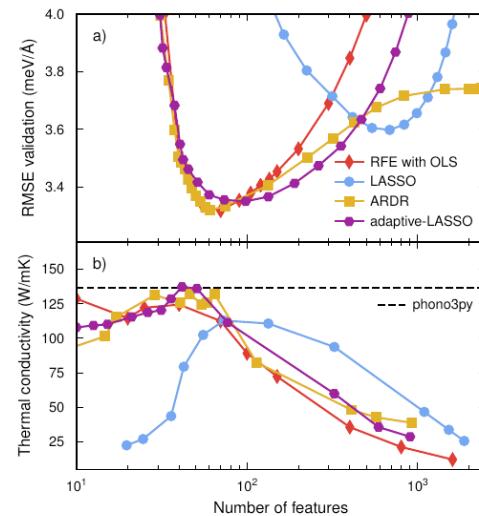
# Force constant expansions

- 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

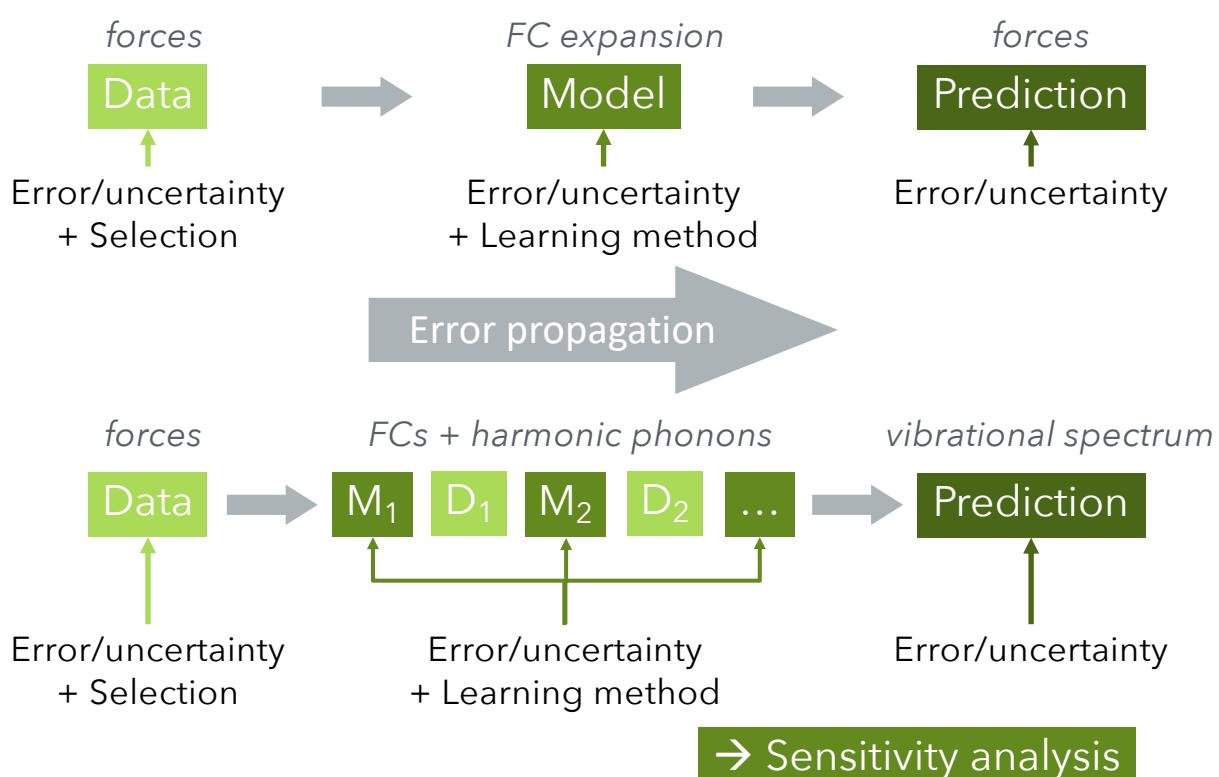
Be mindful of information density  
Judge data cost vs extraction cost



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

9

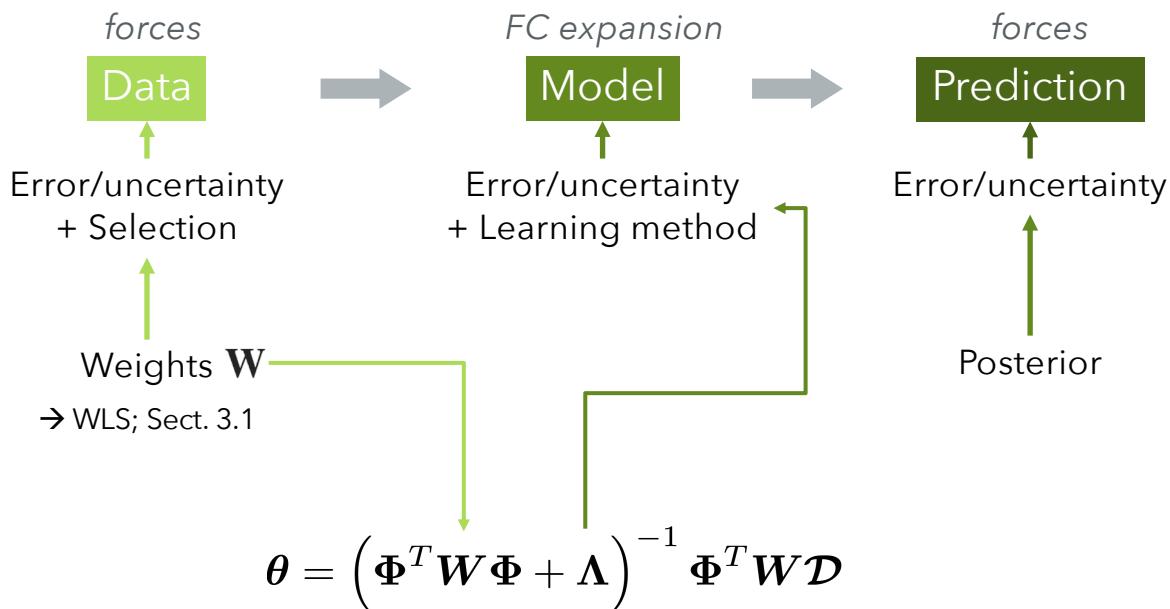
## From model to prediction



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Erhart

# From model to prediction



Combined weighted least squares (WLS) and full covariance matrix

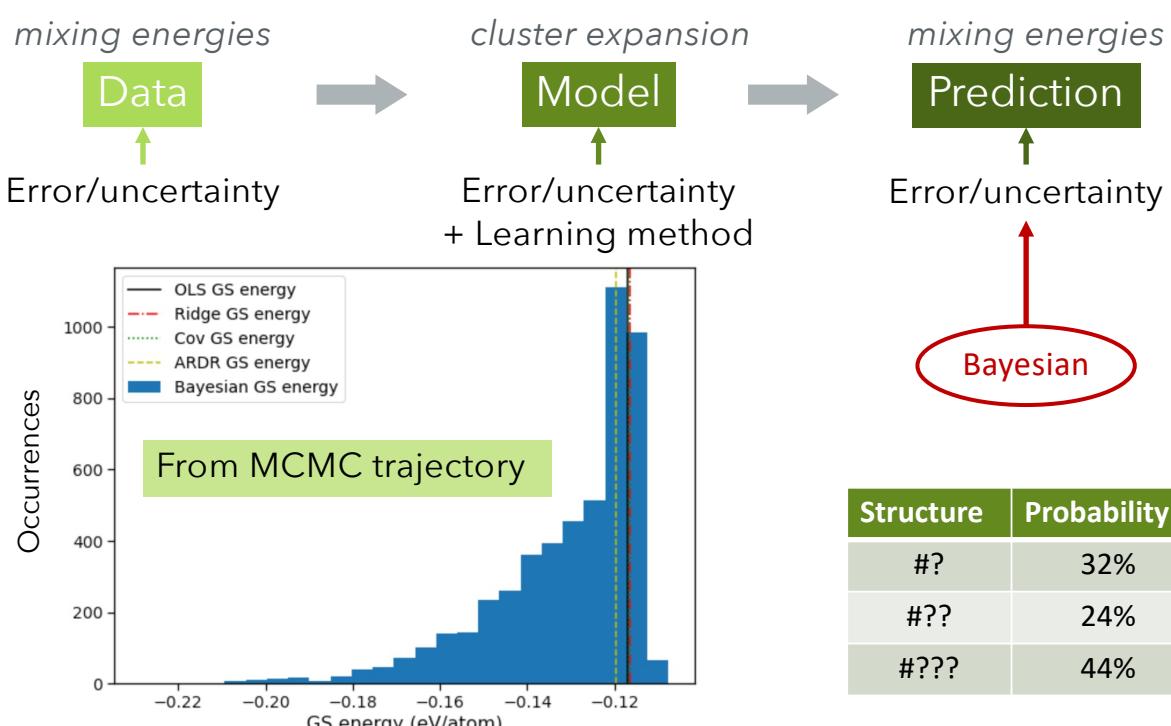
How should the matrix in the bracket be structured?

Erhart

CHALMERS  
UNIVERSITY OF TECHNOLOGY

11

## Example 1: Sampling of posterior



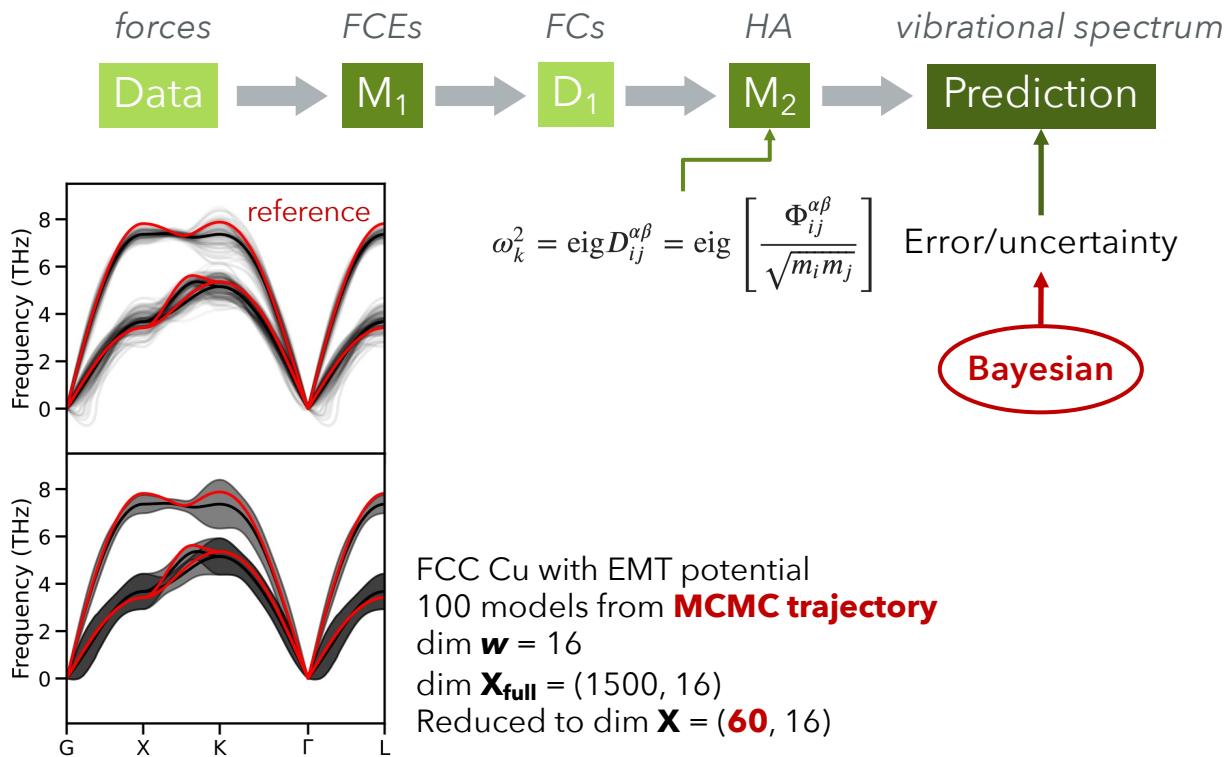
Why can CE succeed in predicting GS?

Erhart

CHALMERS  
UNIVERSITY OF TECHNOLOGY

12

## Example 2: Sampling for predictions

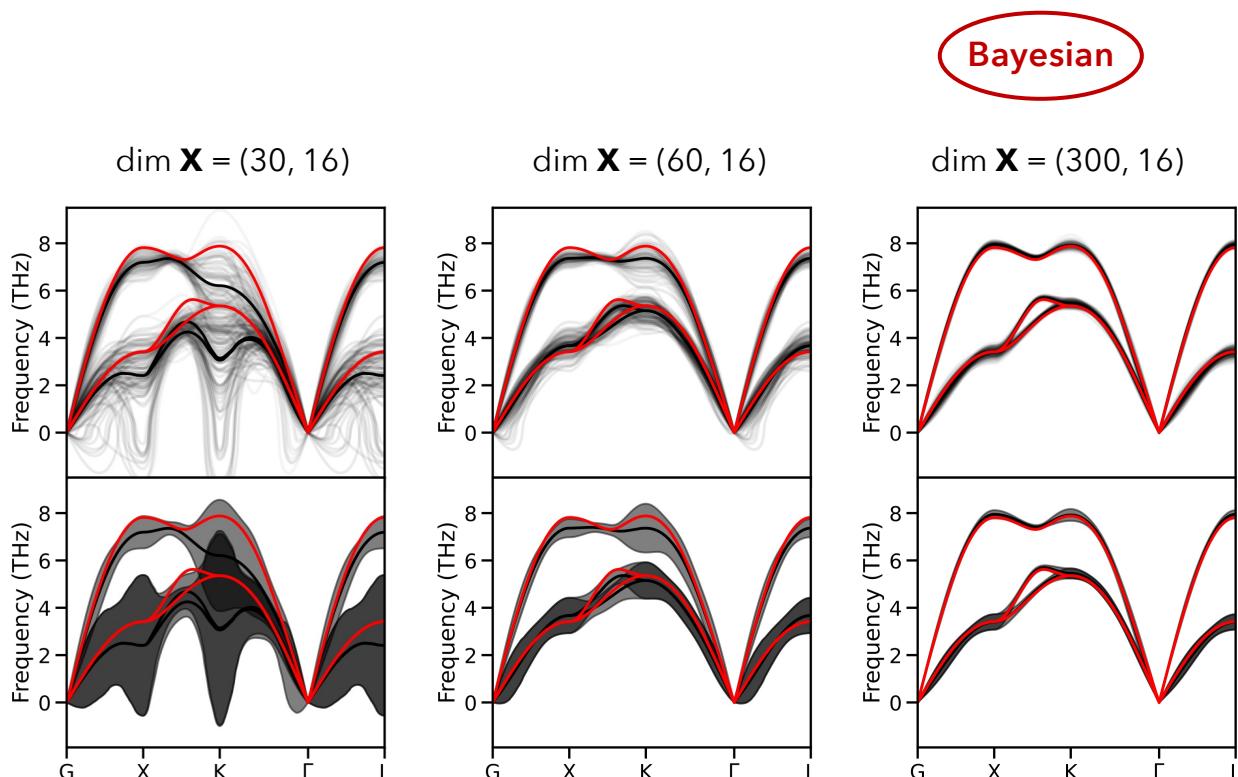


CHALMERS  
UNIVERSITY OF TECHNOLOGY

Erhart

14

## Example 2: Sampling for predictions



CHALMERS  
UNIVERSITY OF TECHNOLOGY

Erhart

# Dependence on input data

$$\theta = \left( \Phi^T W \Phi + \Lambda \right)^{-1} \Phi^T W \mathcal{D}$$

Weight matrix (diagonal)

$$\rightarrow \theta = \left( \Phi_W^T \Phi_W + \Lambda \right)^{-1} \Phi_W^T \mathcal{D}_W$$

Data

$$W_{\alpha\alpha} \in [0, 1]$$

1
2
3
4
5
6

1
2
3
4
5
6

Gradual

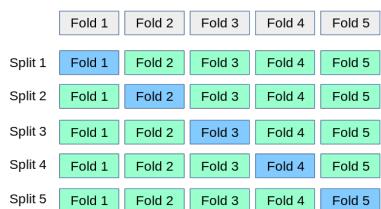
$$W_{\alpha\alpha} \in \{0, 1\}$$

1
2
3
4
5
6

Selection

K-fold

Training data



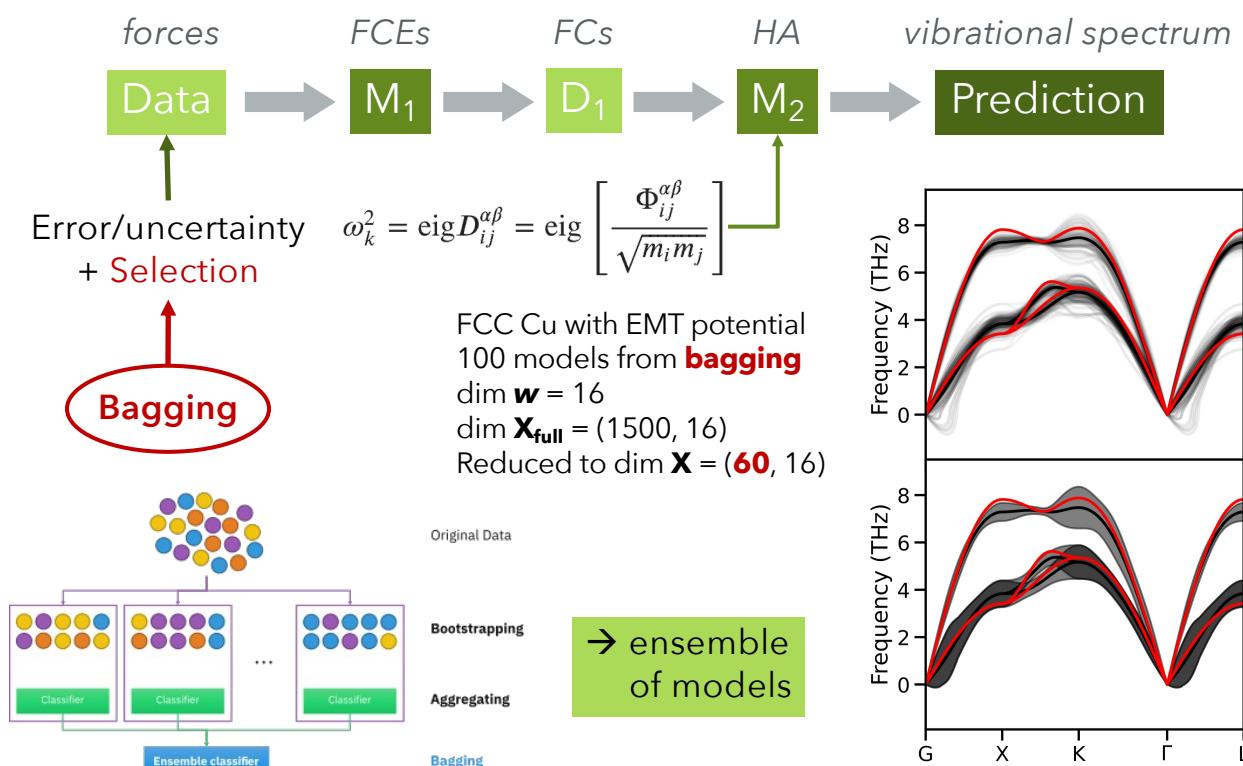
Erhart

CHALMERS  
UNIVERSITY OF TECHNOLOGY

16

## Example 3: Sampling data via bagging

Demo

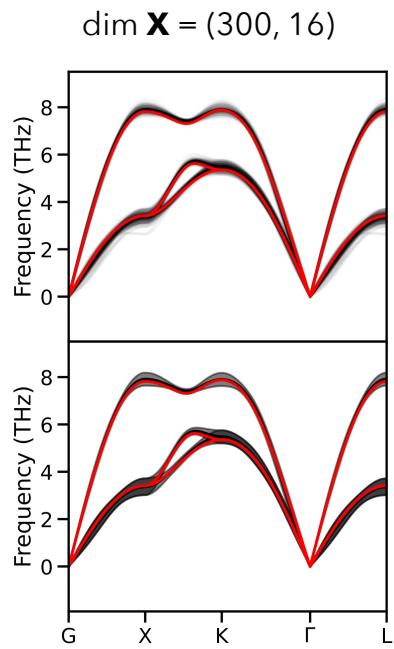
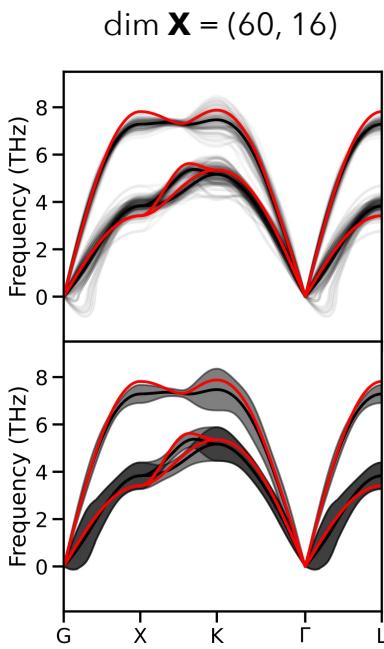
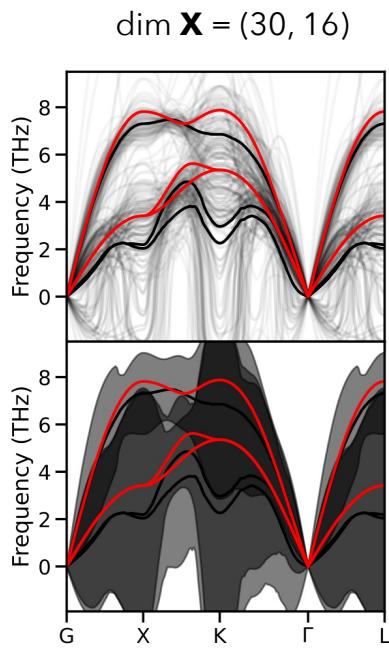


CHALMERS  
UNIVERSITY OF TECHNOLOGY

Erhart

## Example 4: Sampling for predictions

**Bagging**



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Erhart

18

## Example 3 & 4: Comparison

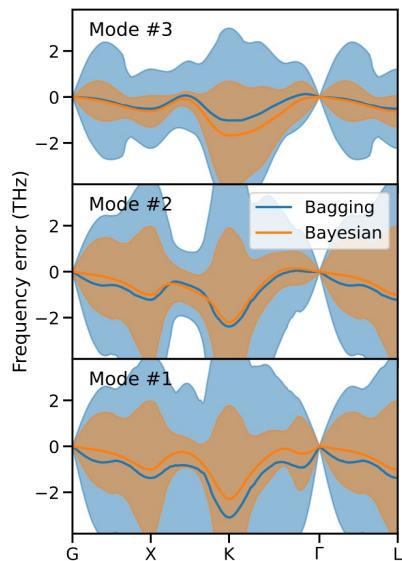
**Bayesian**

Given data, how certain is outcome

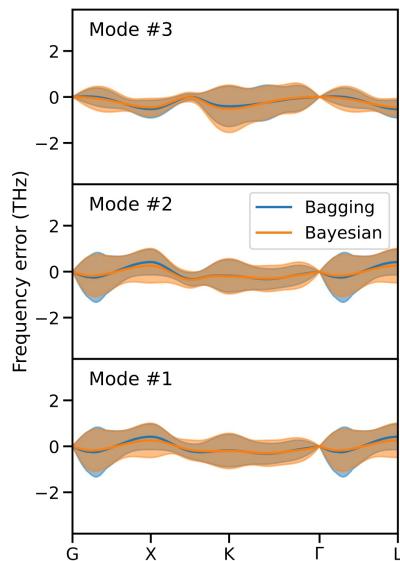
**Both measure sensitivity**

**Bagging**

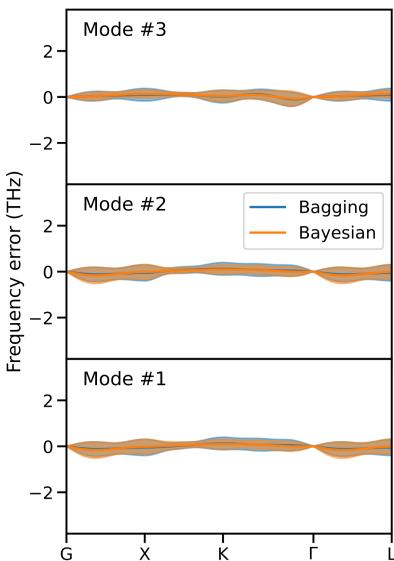
What is effect of **data selection** on outcome



dim  $\mathbf{X} = (30, 16)$   
**CHALMERS**  
UNIVERSITY OF TECHNOLOGY



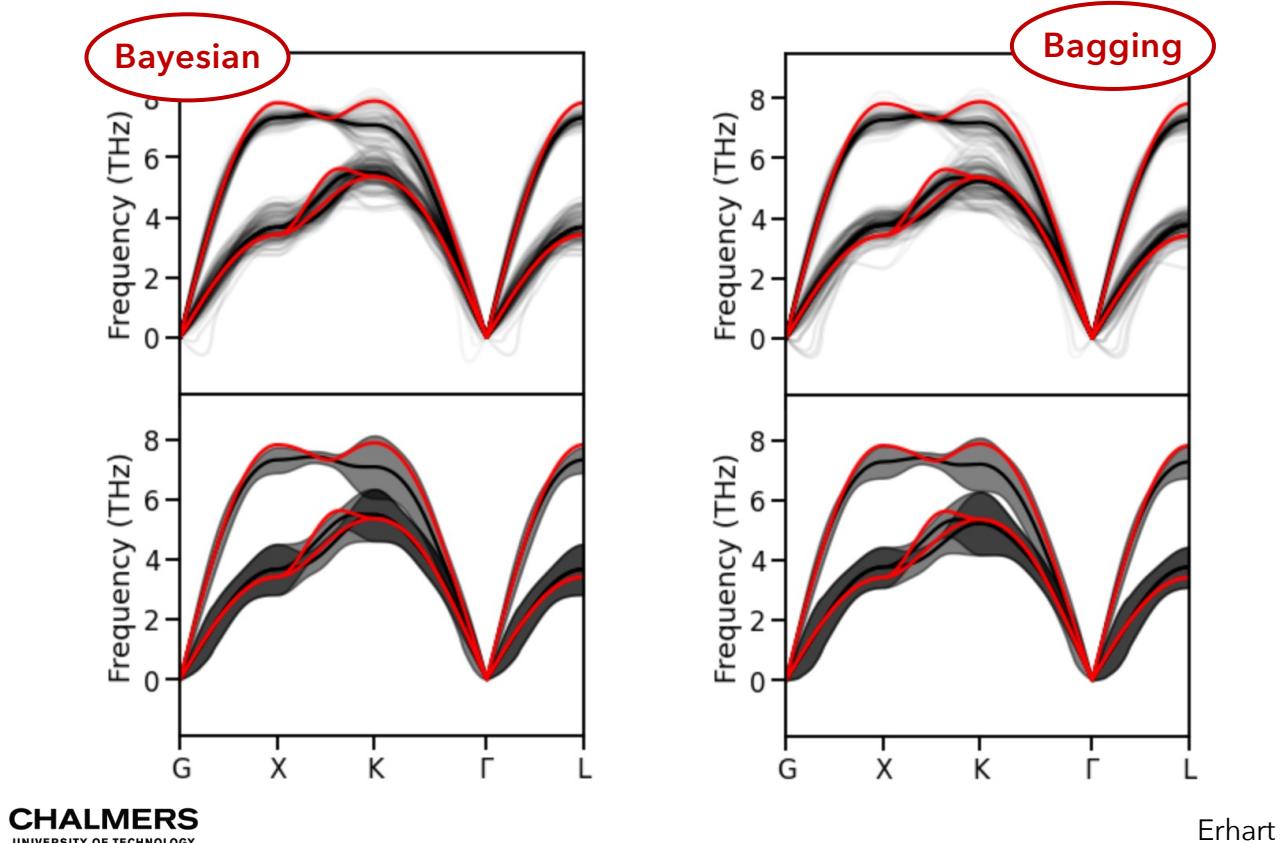
dim  $\mathbf{X} = (60, 16)$



dim  $\mathbf{X} = (300, 16)$   
Erhart

10

## Example 3 & 4: Comparison

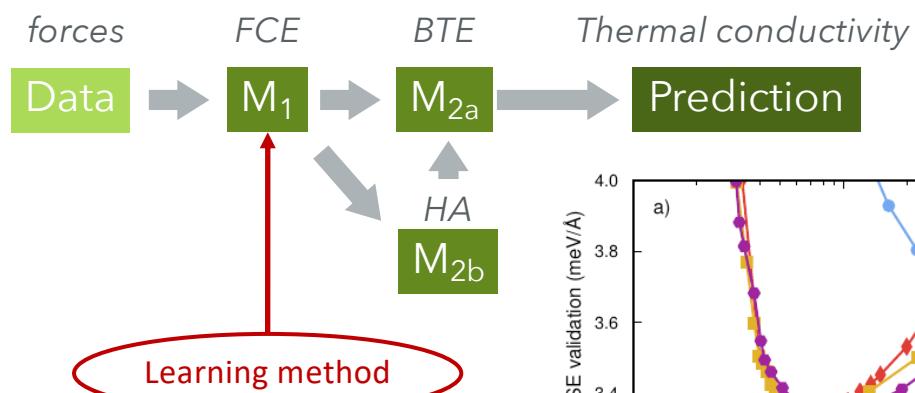


**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

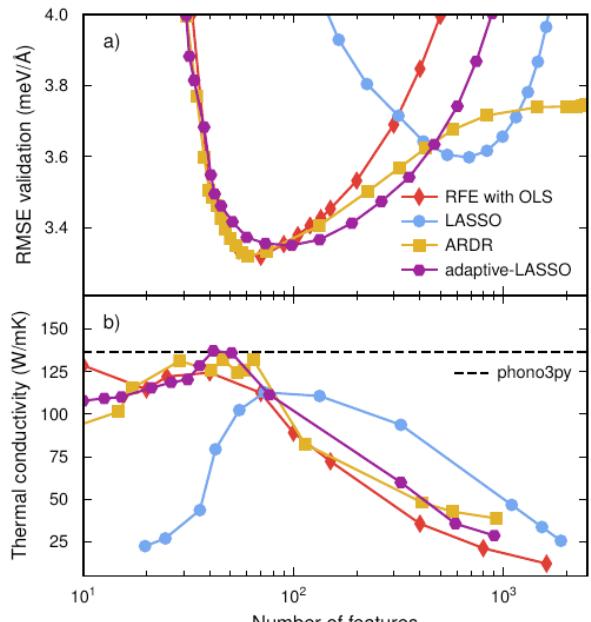
Erhart

20

## Example 4: Learning method



Thermal conductivity  
of bulk silicon

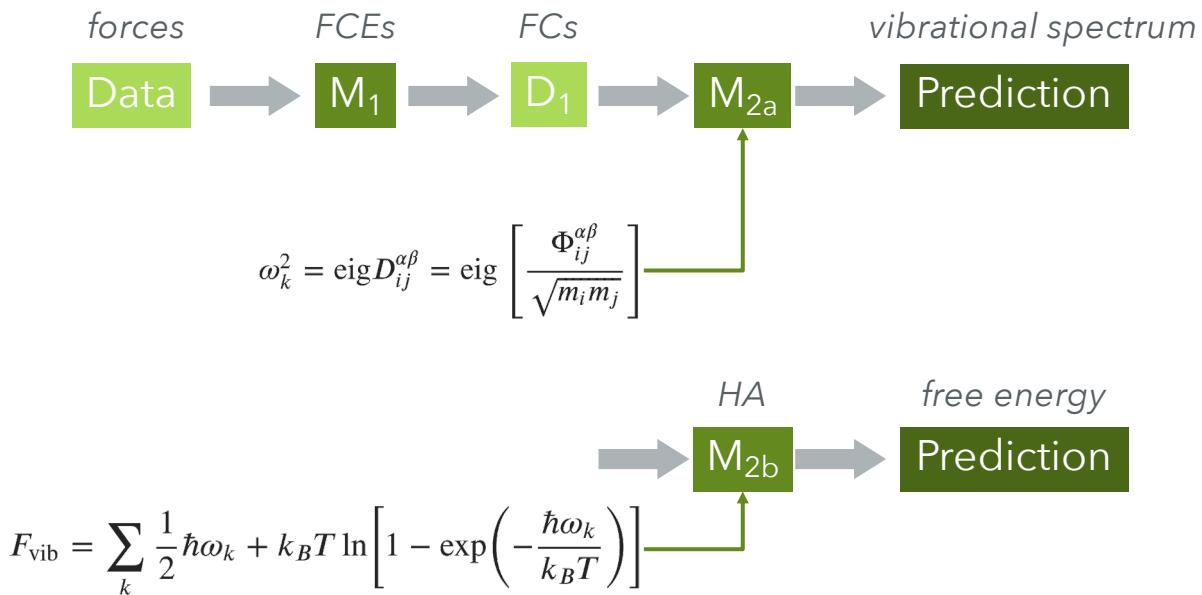


Erhart

**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

21

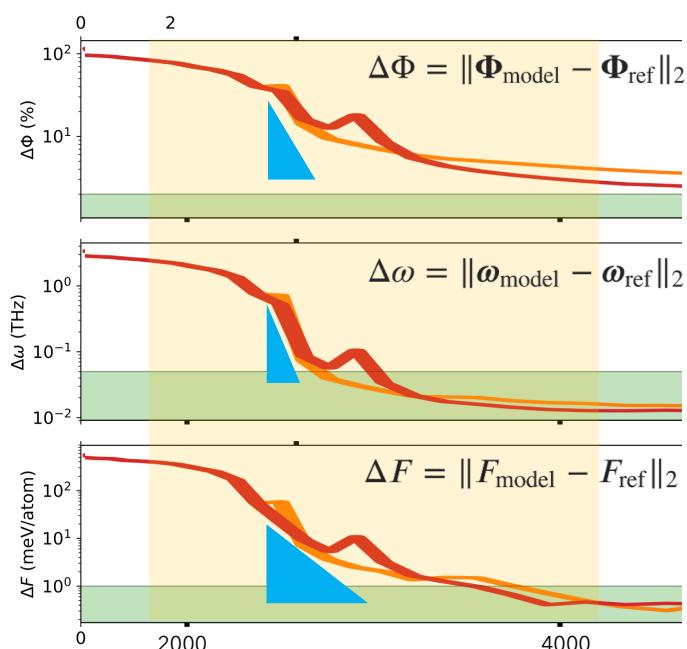
# Different predictors – different sensitivities



# Different predictors – different sensitivities

Convergence with training size

M<sub>1</sub>: FC

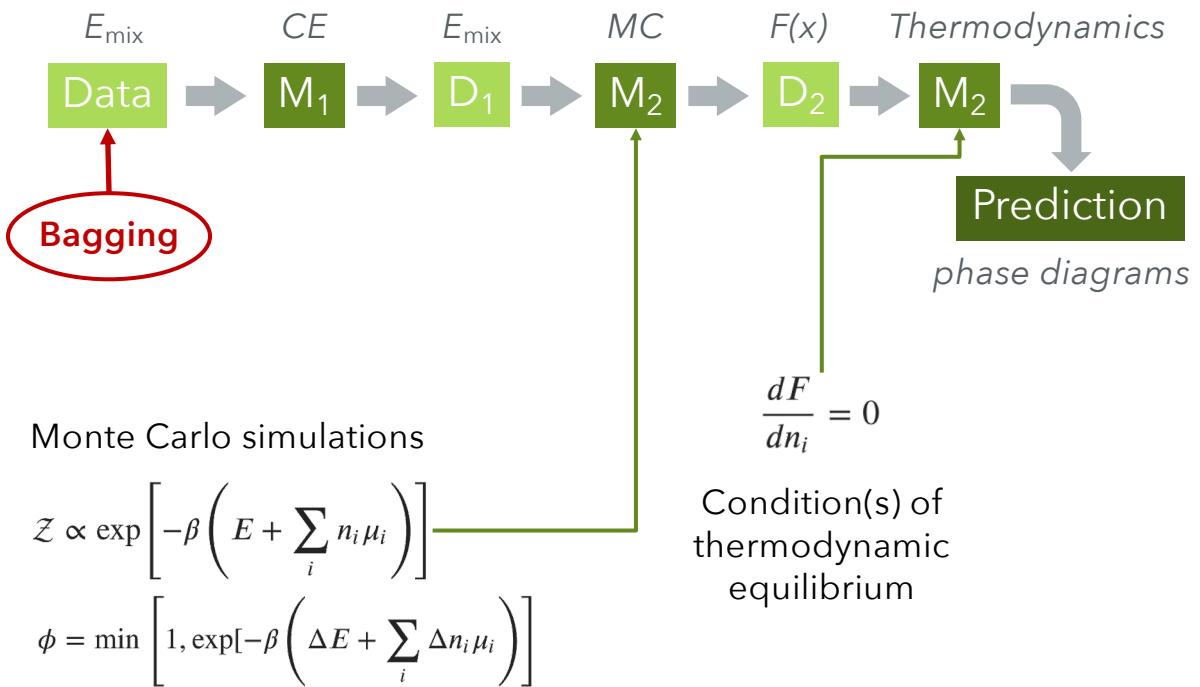


M<sub>2a</sub>: frequencies

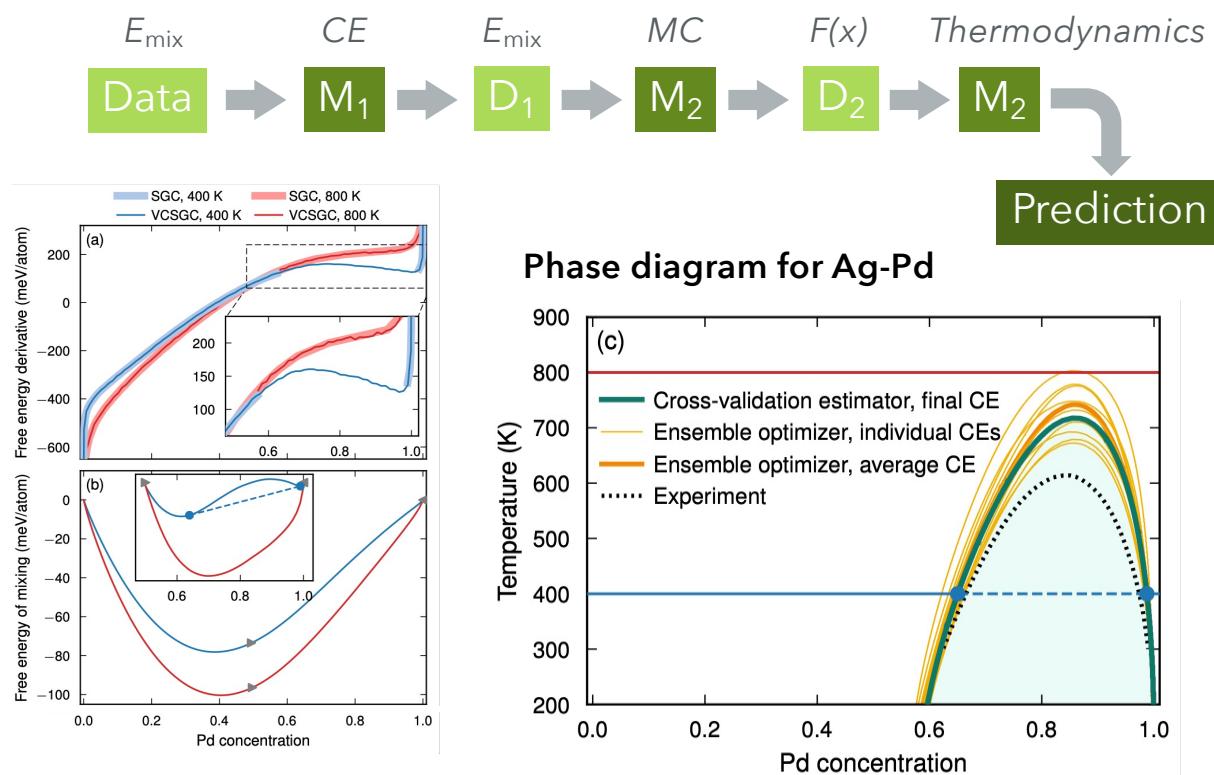
M<sub>2b</sub>: free energies

Can you explain the differences?

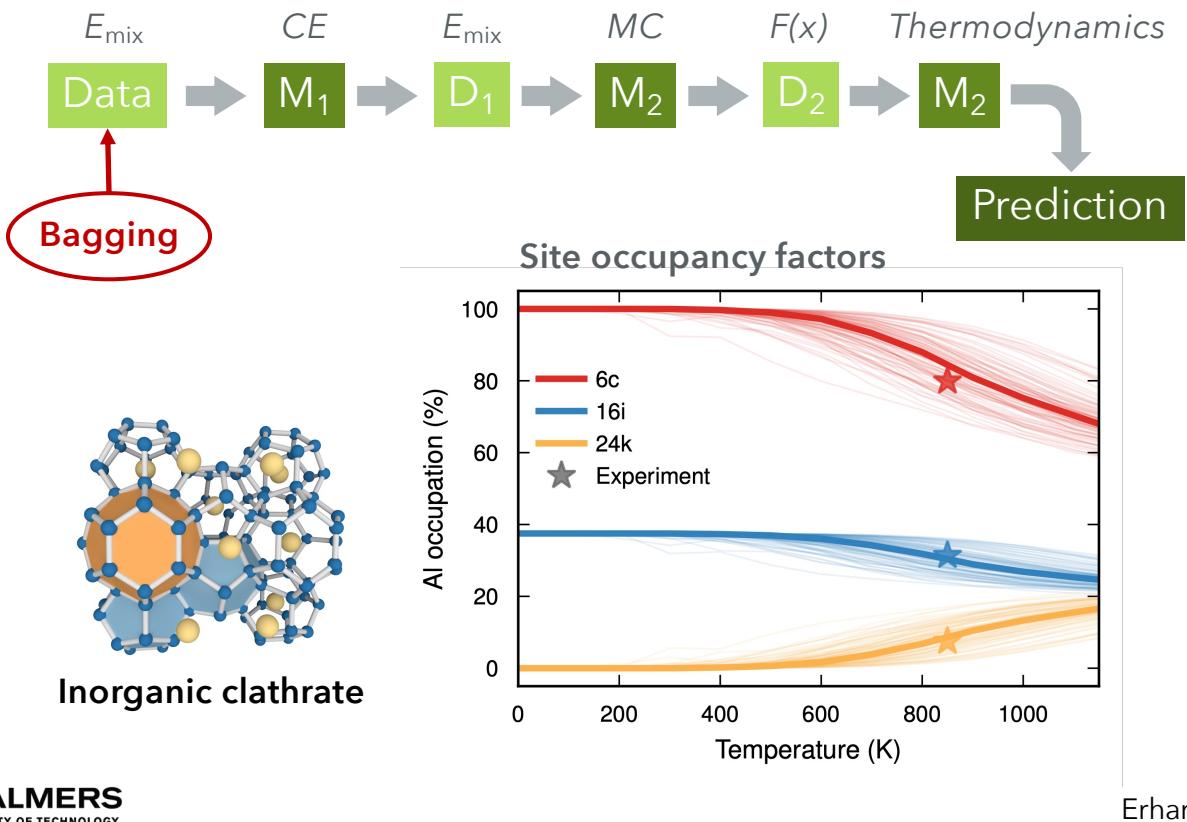
# “Further” error propagation



# “Further” error propagation



# "Further" error propagation

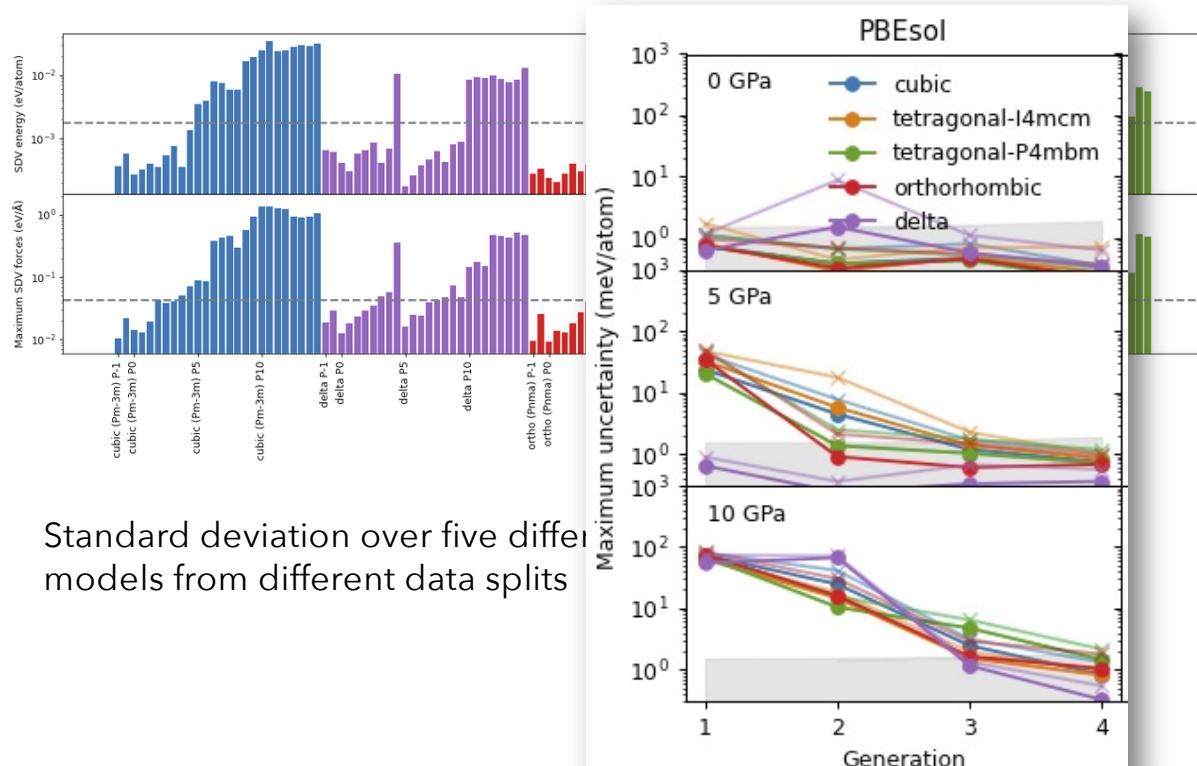


CHALMERS  
UNIVERSITY OF TECHNOLOGY

Erhart

27

## Use uncertainty for active learning

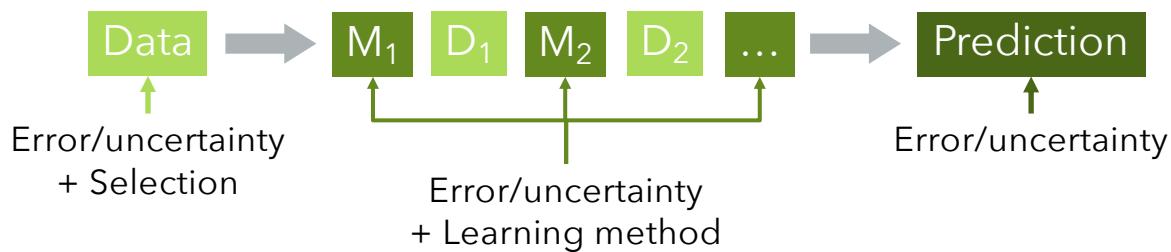


Standard deviation over five different models from different data splits

CHALMERS  
UNIVERSITY OF TECHNOLOGY

Erhart

# Summary



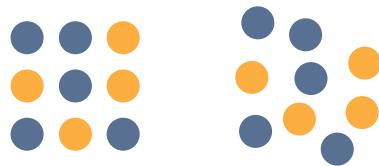
Methods to measure uncertainty

- Errors/uncertainties in data/parameters → Bayesian (MCMC)
  - + Bayesian approach
  - Computationally expensive
  - Does not address data selection/bias
- Data selection/bias → Bagging (Bootstrap aggregating)
  - + easily parallelizable
  - + computationally less demanding than MCMC
  - (usually) only based on maximum likelihood estimate

Interatomic potentials

# Interatomic potentials: 1) Empirical potentials

$$E = E(\{\mathbf{r}_i\}, \{\sigma_i\})$$



Example: A very simple EAM/EMT model with 4 parameters

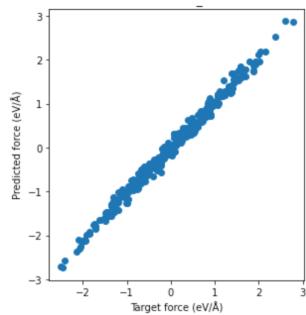
$$E = E(\{\mathbf{r}_i\}, \{\sigma_i\})$$

$$V(r) = A \exp(-\lambda r)$$

$$F(\rho) = -D \exp(-\rho)$$

$$\rho_i = \sum_j \rho(r_{ij})$$

$$\rho(r) = \exp(-2\mu r)$$



Fit model against forces from MD simulations at temperatures below melting

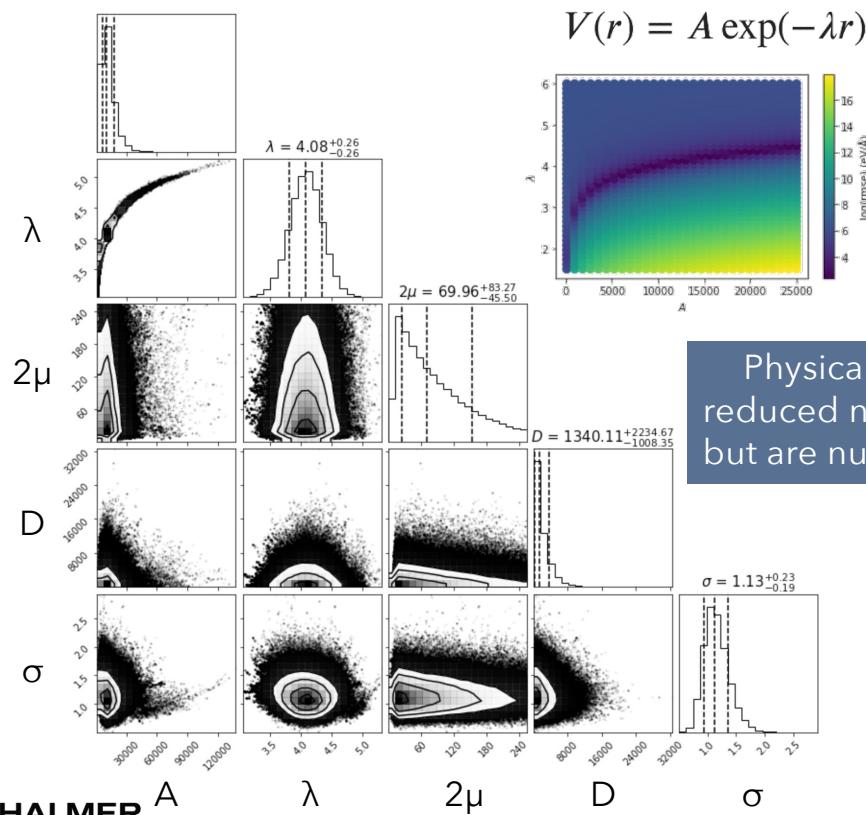


Erhart

**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

31

## Empirical potentials are “sloppy models”

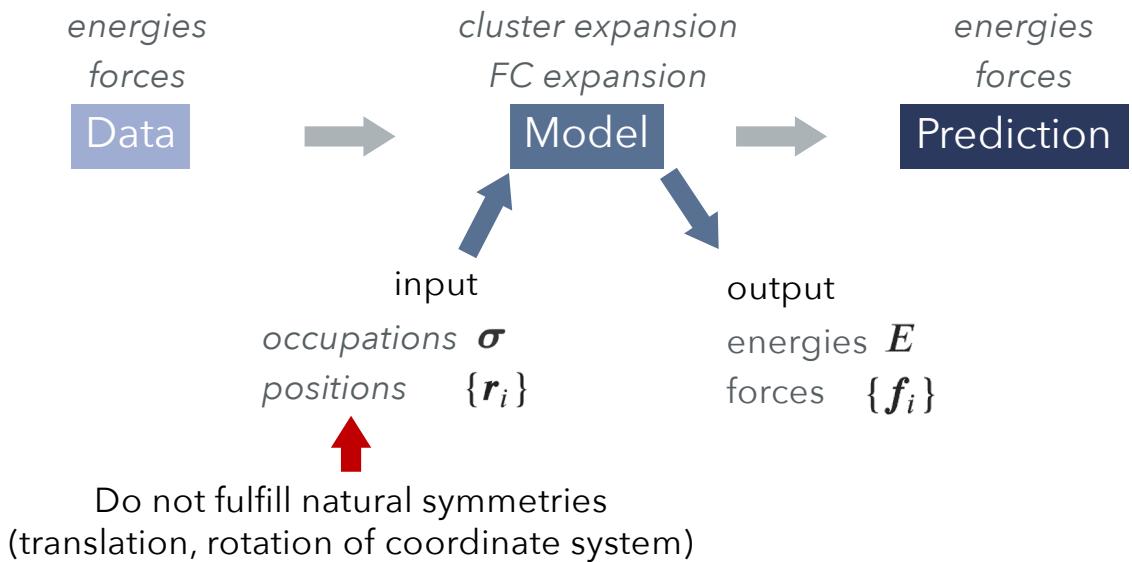


Physically motivated forms  
reduced number of parameters  
but are numerically challenging

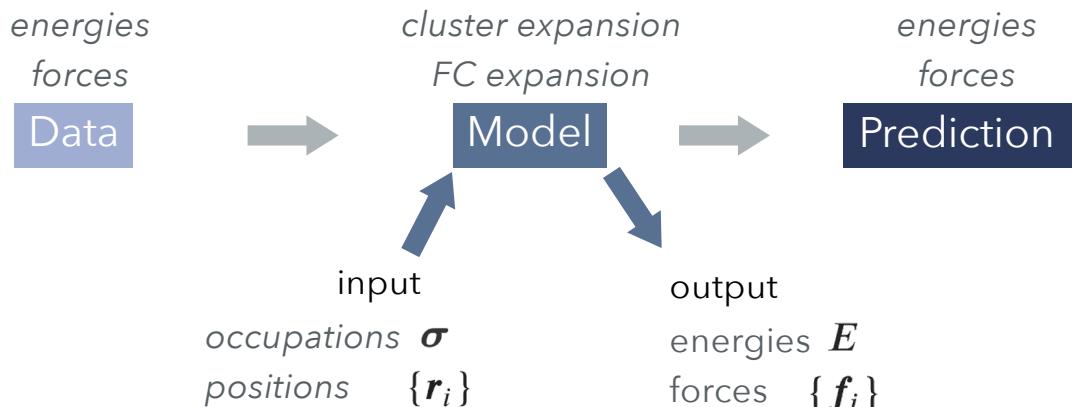
**CHALMER**  
UNIVERSITY OF TECHNOLOGY

Erhart

# How do we represent the state of a system?



# How do we represent the state of a system?



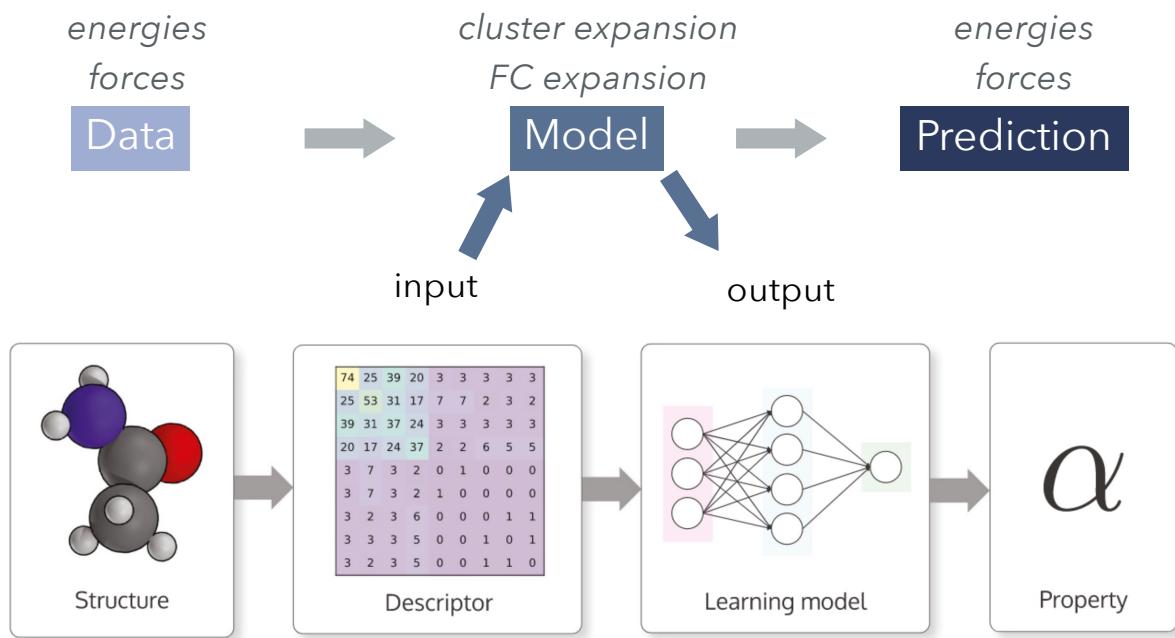
Consider a cluster expansion

$$E(\sigma) = E_0 + \sum_i J_i \sigma_i + \sum_{ij} J_{ij} \sigma_i \sigma_j + \sum_{ijk} J_{ijk} \sigma_i \sigma_j \sigma_k + \dots$$

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

Coordinate transformation

# How do we represent the state of a system?



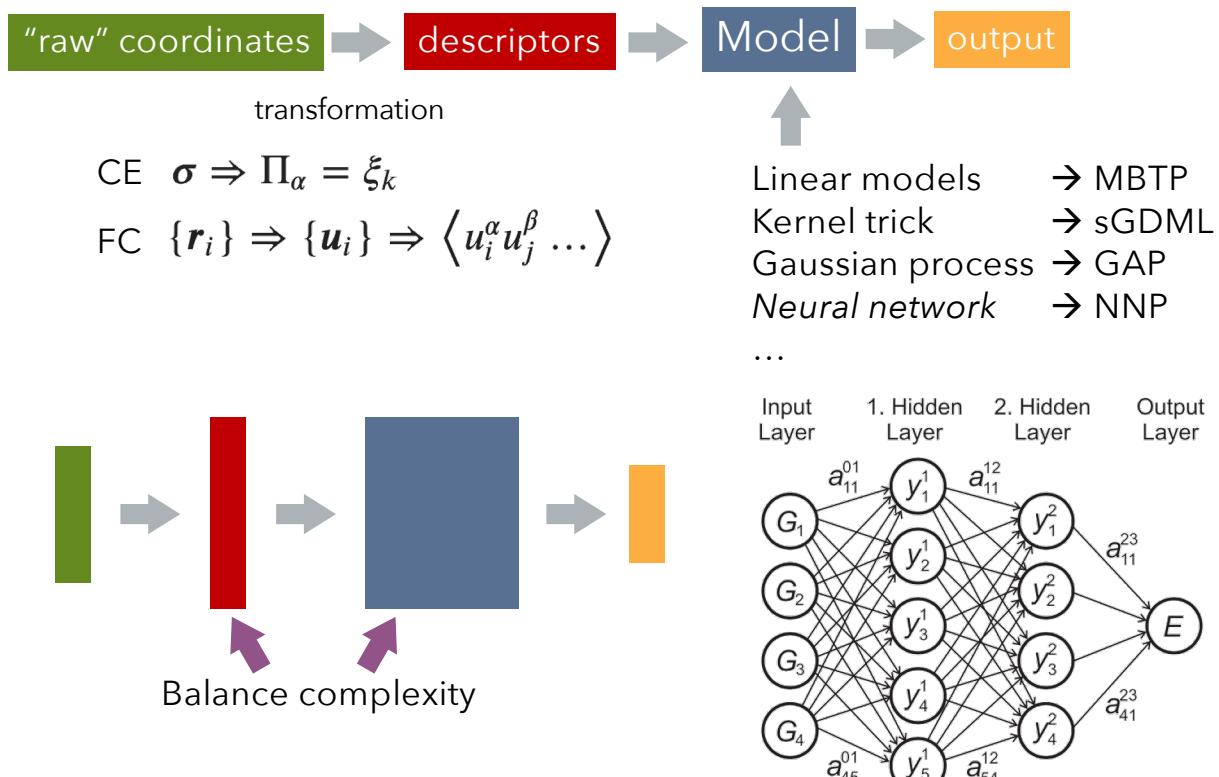
**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Figure from Comp. Phys. Comm. 247, 106949 (2020)

Erhart

35

## Model vs descriptor complexity



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

Figure from J. Chem. Phys. 134, 074106 (2011)

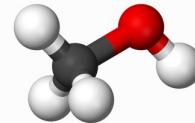
Erhart

# Descriptors: Coulomb matrix



$$M_{ij}^{\text{Coulomb}} = \begin{cases} 0.5Z_i^{2.4} & \text{for } i = j \\ \frac{Z_i Z_j}{R_{ij}} & \text{for } i \neq j \end{cases}$$

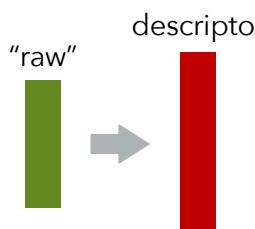
ethanol



$$\begin{bmatrix} 36.9 & 33.7 & 5.5 & 3.1 & 5.5 & 5.5 \\ 33.7 & 73.5 & 4.0 & 8.2 & 3.8 & 3.8 \\ 5.5 & 4.0 & 0.5 & 0.35 & 0.56 & 0.56 \\ 3.1 & 8.2 & 0.35 & 0.5 & 0.43 & 0.43 \\ 5.5 & 3.8 & 0.56 & 0.43 & 0.5 & 0.56 \\ 5.5 & 3.8 & 0.56 & 0.43 & 0.56 & 0.5 \end{bmatrix}$$

- Fast to compute
- Low dimensional
- Applicable to any chemistry
- Too unspecific for accurate predictions
- Useful for high-throughput fingerprinting of molecules

# Descriptors: Smooth overlap of atomic positions (SOAP)



For each atom

Density field

$$\rho^Z(\mathbf{r}) = \sum_i^{|Z|} e^{-\frac{1}{2\sigma^2} |\mathbf{r} - \mathbf{R}_i|^2}$$

Projection on spherical harmonics

$$\rho^Z(\mathbf{r}) = \sum_{nlm} c_{nlm}^Z g_n(r) Y_{lm}(\theta, \phi)$$

$$c_{nlm}^Z = \iiint_{\mathcal{R}^3} dV g_n(r) Y_{lm}(\theta, \phi) \rho^Z(\mathbf{r})$$

$$\text{Partial power spectrum } p_{nn'l}^{Z_1, Z_2} = \pi \sqrt{\frac{8}{2l+1}} \sum_m \left( c_{nlm}^{Z_1} \right)^* c_{n'lm}^{Z_2}$$

Full power spectrum by concatenating partial PS

- Much more expensive compute
- Applicable to any chemistry
- Suitable for accurate predictions
- Basis for, e.g., GAP and more recent NNPs

# Summary

- Interatomic potentials
  - Empirical potentials  
→ physically motivated → functionally challenged
  - Machine learning potentials  
→ physically challenged → functionally flexible
- Balance model vs descriptor complexity
- Descriptors (examples)
  - General
    - Coulomb matrix, Sine matrix
    - Smooth overlap of atomic orbitals (SOAP)
    - Atom centered symmetry functions
    - Many-body tensor representations
  - Restricted
    - Displacement or spin products