SAND2022-13405C

Sandia National Laboratories

Exceptional service in the national interest

Active Learning of SNAP Potentials using Bayesian Uncertainty Estimation

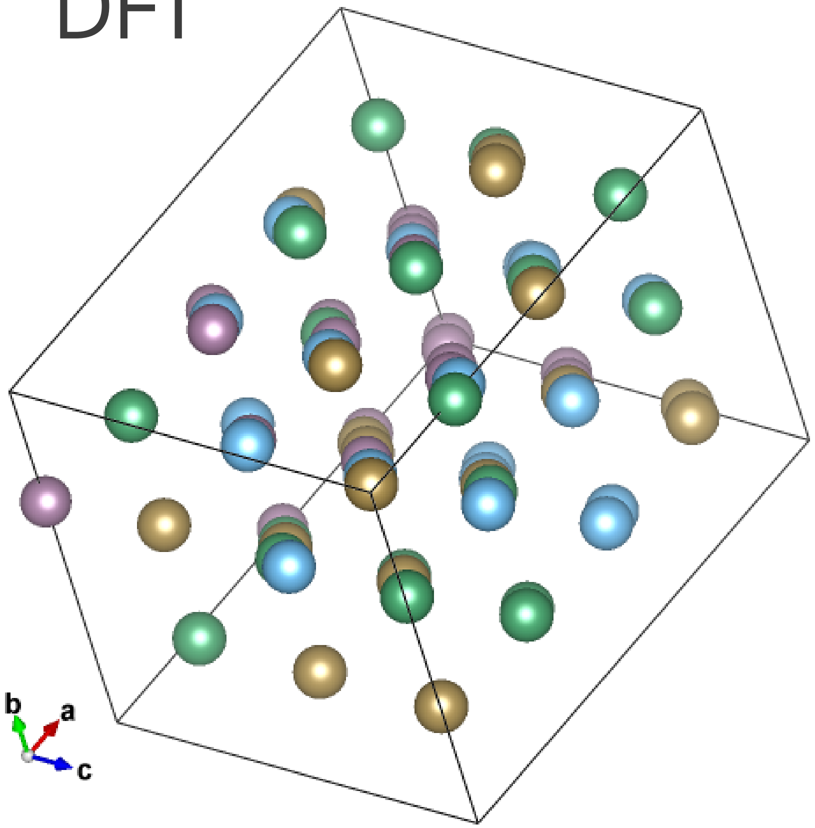
Logan Williams, Khachik Sargsyan, Katherine Johnston, Habib Najm

October 5, 2022

The authors would like to thank the DOE Office of Science, Advanced Scientific Computing Research (ASCR) and Fusion Energy Sciences (FES) for funding.

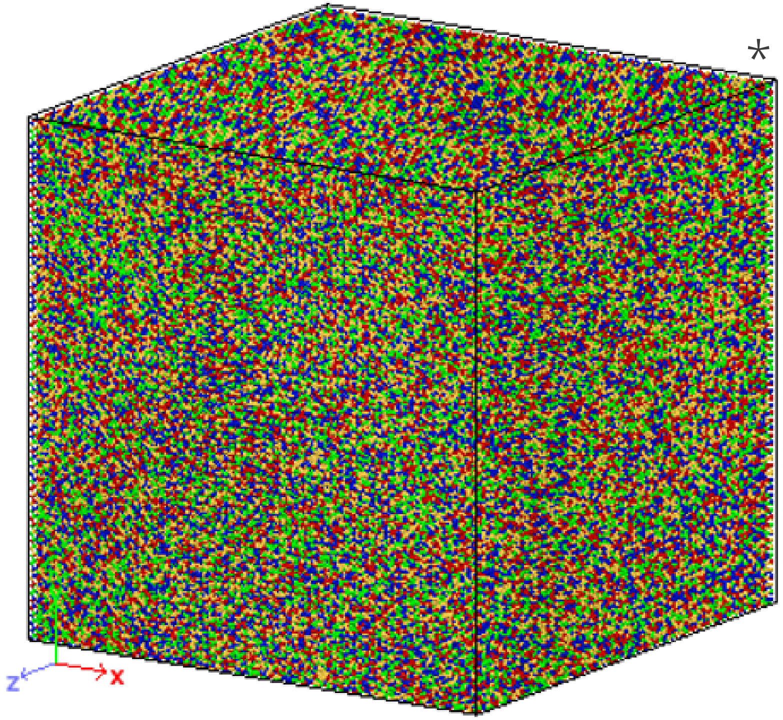
Efficient atomistic computation allows scaling

DFT



System size limit: ~1000 atoms

MD



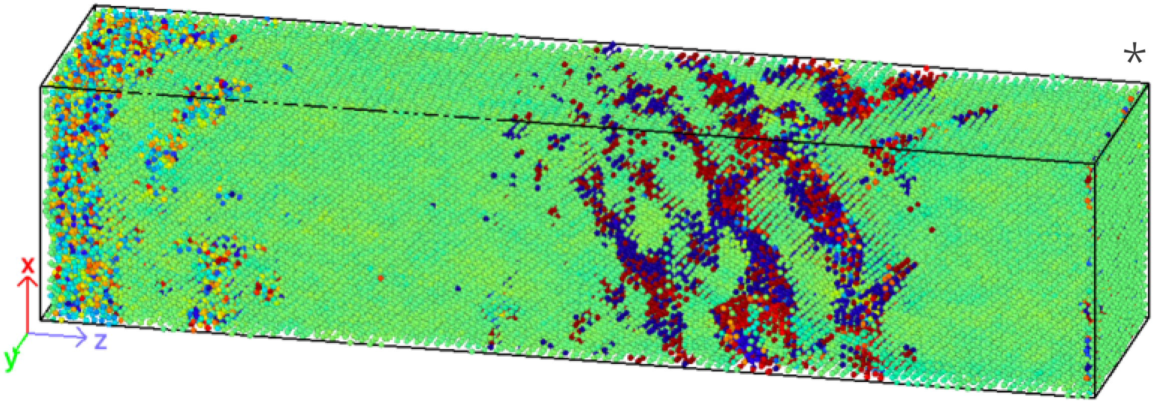
System size limit: ~1,000,000,000 atoms

* Image credit to Megan McCarthy

Uncertainty Quantification (UQ) gives reliability of results

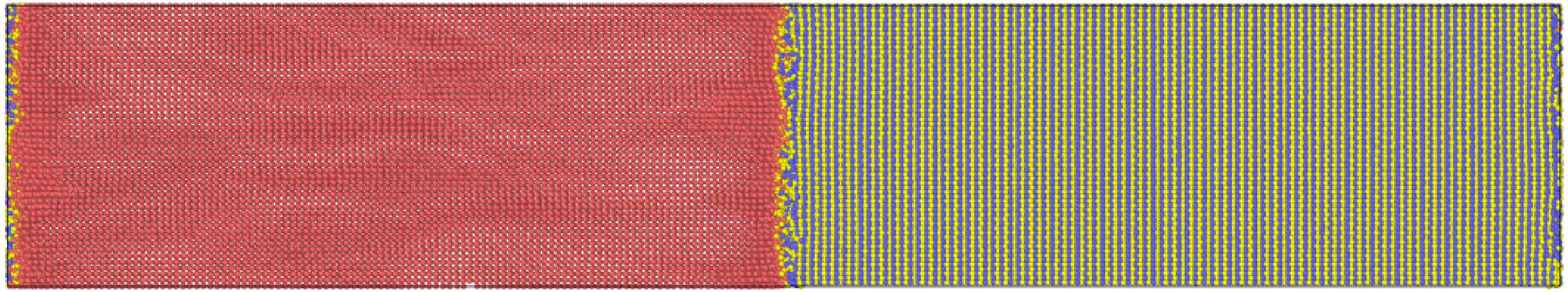
- MD simulations are (almost) always outside the size limit of DFT
- Often studying behavior that can not be replicated at smaller scale
- Trust in the model results is required – UQ allows for building that

MD



*

**



* Image credit to Megan McCarthy
** Image credit to Ember Sikorski

Bayes Rule

- For a model f(x, c) and data $y_i = y(x_i)$, calibrate the model parameters, c.

Likelihood

Prior probability

$$p(c|y) = \frac{p(y|c)\,p(c)}{p(y)} \;\longrightarrow\; p(c|y) \propto p(y|c)\,p(c)$$

Posterior probability

Evidence

Bayesian Parameter Inference

- For a model f(x, c) and data $y_i = y(x_i)$, calibrate the model parameters, c.

Likelihood

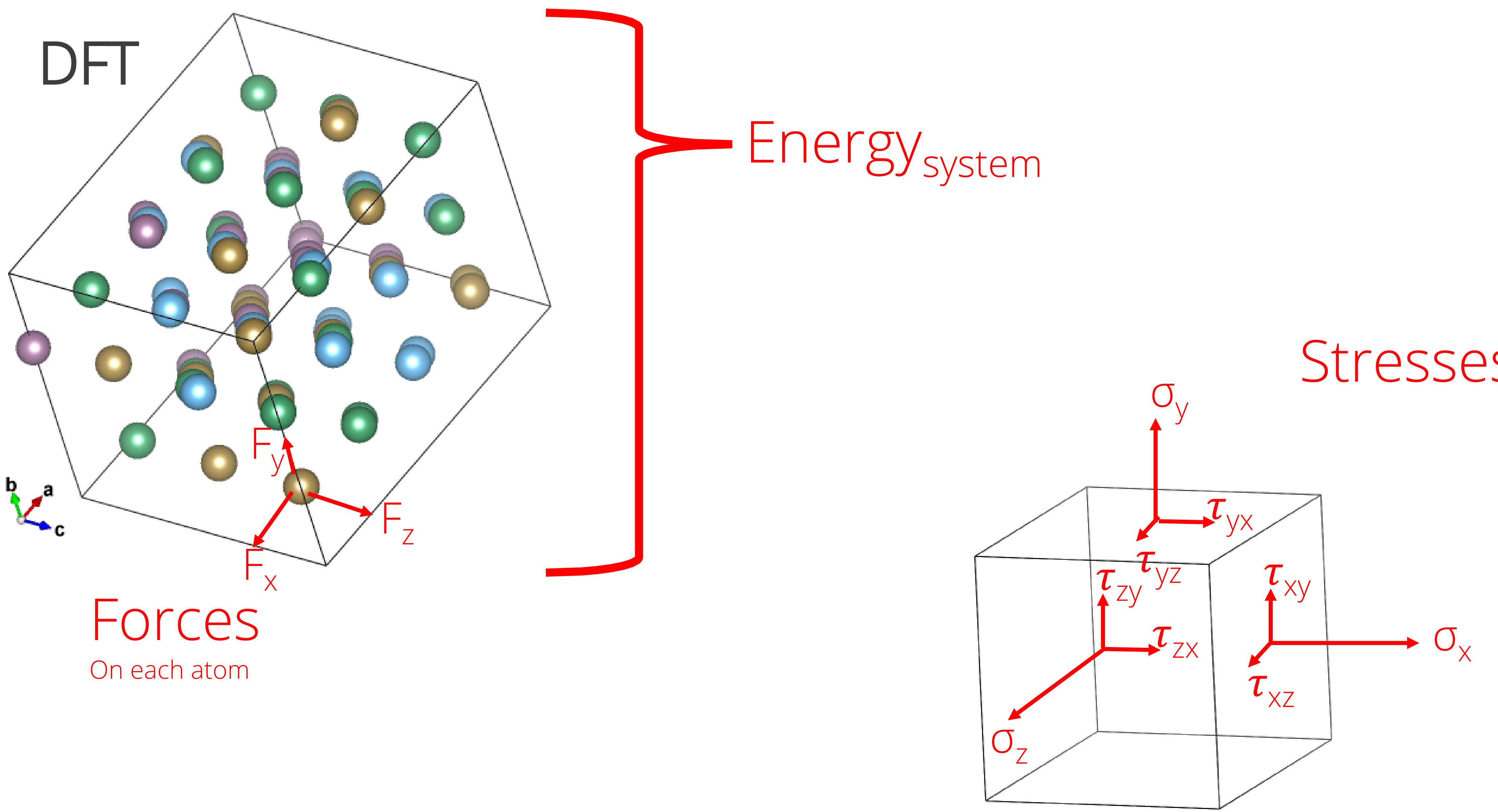
$$p(c|y) = \frac{p(y|c)p(c)}{p(y)}$$

Make some assumptions about noise: $y_i = f(x_i, c) + \sigma_i \epsilon_i, \text{ where } \epsilon_i \sim \mathcal{N}(0, 1)$

Likelihood

$$p(y|c) \propto \prod_{i=1}^{N} exp(-\frac{(f(x_i, c) - y_i)^2}{2\sigma_i^2})$$

DFT provides training data



Bispectrum – formulation of the descriptor

Density around an atom

$$\rho_i(\mathbf{r}) = \delta(\mathbf{r}) + \sum_{r_{ii'} < R_{cut}} f_c(r_{ii'}) w_{i'} \delta(\mathbf{r} - \mathbf{r}_{ii'})$$

Represent different atom types

Switching function, smooth to zero approaching R_{cut}

Modified polar coordinates

$$\theta_0 = \theta_0^{max} \frac{r}{R_{cut}}$$

Points near south pole are excluded

Switch to 4D hyperspherical harmonics

$$\rho(\mathbf{r}) = \sum_{j=0,\frac{1}{2},\ldots}^{\infty} \sum_{m=-j}^{j} \sum_{m'=-j}^{j} u^j_{m,m'} U^j_{m,m'}(\theta_0, \theta, \phi)$$

Hyperspherical harmonics basis functions

where

$$u^j_{m,m'} = U^j_{m,m'}(0,0,0) + \sum_{r_{ii'} < R_{cut}} f_c(r_{ii'}) w_i U^j_{m,m'}(\theta_0, \theta, \phi)$$

Complex valued expansion coefficients

Convert to the real valued, rotation invariant scalar triple products

$$B_{j_1,j_2,j} = \sum_{m_1,m'_1=-j_1}^{j_1} \sum_{m_2,m'_2=-j_2}^{j_2} \sum_{m,m'=-j}^{j} (u^j_{m,m'})^* H^{jmm'}_{\substack{j_1m_1m'_1 \\ j_2m_2m'_2}} u^{j_1}_{m_1,m'_1} u^{j_2}_{m_2,m'_2}$$

Bispectrum components

Coupling coefficients (constants)

($j = 0, 1/2, 1, \ldots$ and $m, m' = -j, -j+1, \ldots, j-1, j$)

Linear SNAP model for interatomic potential – single element model

$$E_{SNAP}(\mathbf{r}^N) = N\beta_0 + \boldsymbol{\beta} \cdot \sum_{i=1}^{N} \mathbf{B}^i$$

Linear coefficients to be fit

Constant energy contribution

$\mathbf{B}^i$ = (flattened) Bispectrum components
= F(atomic positions and geometry)
Defined for each atom in structure

$$\mathbf{F}^j_{SNAP} = -\nabla_j E_{SNAP} = -\boldsymbol{\beta} \cdot \sum_{i=1}^{N} \frac{\partial \mathbf{B}^i}{\partial \mathbf{r}_j}$$

$$\mathbf{W}_{SNAP} = -\sum_{j=1}^{N} \mathbf{r}_j \otimes \nabla_j E_{SNAP} = -\boldsymbol{\beta} \cdot \sum_{j=1}^{N} \mathbf{r}_j \otimes \sum_{i=1}^{N} \frac{\partial \mathbf{B}^i}{\partial \mathbf{r}_j}$$

Cartesian outer product

Example Descriptor Rows

Two elements (W, H). Higher 2J_max for W than for H. 55 non-zero descriptors for W, 14 non-zero descriptors for H

Energy of a W only structure = [34.50, -1.82, -2.94, 2.00, -3.32, -1.92, ... , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...]
Energy of a H only structure = [0, 0, 0, 0, 0, ..., 2.37, 1.77, -0.33, 1.12, -0.50, -1.87, 0, 0.87, 0, 0, 0, 2.87, 0, 0, 0, ...]
Energy of a W+H structure = [16.86, -0.98, -2.27, 4.15, -2.31, -0.46, ..., 0.53, -0.17, 0, 0.09, 0, 0, 0, -0.04, 0, 0, 0, ...]

Force follow the same zero-nonzero patterns based on what atoms surround the atom of interest

	Crystal Structure 1
1 row	$\sum_{i=1}^{N_s} \mathbf{B}^i$
3N rows	$-\sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_i^\alpha}$
6 rows	$-\sum_{j=1}^{N_s} r_j^\alpha \sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_j^\beta}$

Model fitting – linear regression problem

Ab=y

Crystal Structure 1

1 row	$\sum_{i=1}^{N_s} \mathbf{B}^i$
3N rows	$-\sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_j^\alpha}$
6 rows	$-\sum_{j=1}^{N_s} r_j^\alpha \sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_j^\beta}$

Group 1

Group 2

Group 3

Group ...

Group M

$$\cdot \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \vdots \\ E_s^{qm} - E_s^{ref} \\ \\ F_{j,\alpha}^{qm} - F_{j,\alpha}^{ref} \\ \vdots \\ W_{\alpha\beta,s}^{qm} - W_{\alpha\beta,s}^{ref} \\ \vdots \end{bmatrix}$$

Weighted linear regression

$$Ab=y \quad \Rightarrow \quad (w\odot A)b = (w\odot y)$$

$$\tilde{A}b = \tilde{\mathbf{y}}$$

Group 1	weight 1
Group 2	weight 2
Group 3	weight 3
Group ...	weight ...
Group M	weight M

FitSNAP default solver

$$L2\ loss\ function = \sum_{i=1}^{n} (y_{true} - y_{predicted})^2$$

Ab=y

$$\begin{bmatrix} \sum_{i=1}^{N_s} \mathbf{B}^i \\ -\sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_j^\alpha} \\ -\sum_{j=1}^{N_s} r_j^\alpha \sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_j^\beta} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \vdots \\ E_s^{qm} - E_s^{ref} \\ \vdots \\ F_{j,\alpha}^{qm} - F_{j,\alpha}^{ref} \\ \vdots \\ W_{\alpha\beta,s}^{qm} - W_{\alpha\beta,s}^{ref} \\ \vdots \end{bmatrix}$$

b* = best fit solution of least-squares problem: $min(|\tilde{\boldsymbol{y}} - \tilde{A}\boldsymbol{b}|_2)$

(with a vector of group weights)

$(w \odot A)b = (w \odot y)$

$\tilde{A}b = \tilde{y}$

Trained model:

Ab*= predictions

UQ-FitSNAP = Analytical Bayesian solver

$$L2\ loss\ function = \sum_{i=1}^{n} (y_{true} - y_{predicted})^2$$

$$SNAP\ model\ fit = (\tilde{A}^T \cdot \tilde{A})^{-1} \cdot (\tilde{A}^T \cdot \tilde{\boldsymbol{y}}) = \boldsymbol{b}^*$$

(identical to standard least squares regression fit)

$$SNAP\ fit\ covariance = \frac{(\tilde{\boldsymbol{y}} - \tilde{A}\boldsymbol{b}^*)^T \cdot (\tilde{\boldsymbol{y}} - \tilde{A}\boldsymbol{b}^*)}{N - k - 2} \cdot (\tilde{A}^T \cdot \tilde{A})^{-1} = \Sigma$$

$$\boldsymbol{b} = MVN(\boldsymbol{b}^*, \Sigma)$$

Can draw samples / do statistics!

Ab=y

$$\begin{bmatrix} \sum_{i=1}^{N_s} \mathbf{B}^i \\ -\sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_j^\alpha} \\ -\sum_{j=1}^{N_s} r_j^\alpha \sum_{i=1}^{N_s} \frac{\partial \mathbf{B}^i}{\partial r_j^\beta} \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \vdots \\ E_s^{qm} - E_s^{ref} \\ \vdots \\ F_{j,\alpha}^{qm} - F_{j,\alpha}^{ref} \\ \vdots \\ W_{\alpha\beta,s}^{qm} - W_{\alpha\beta,s}^{ref} \\ \vdots \end{bmatrix}$$

(with a vector of group weights)

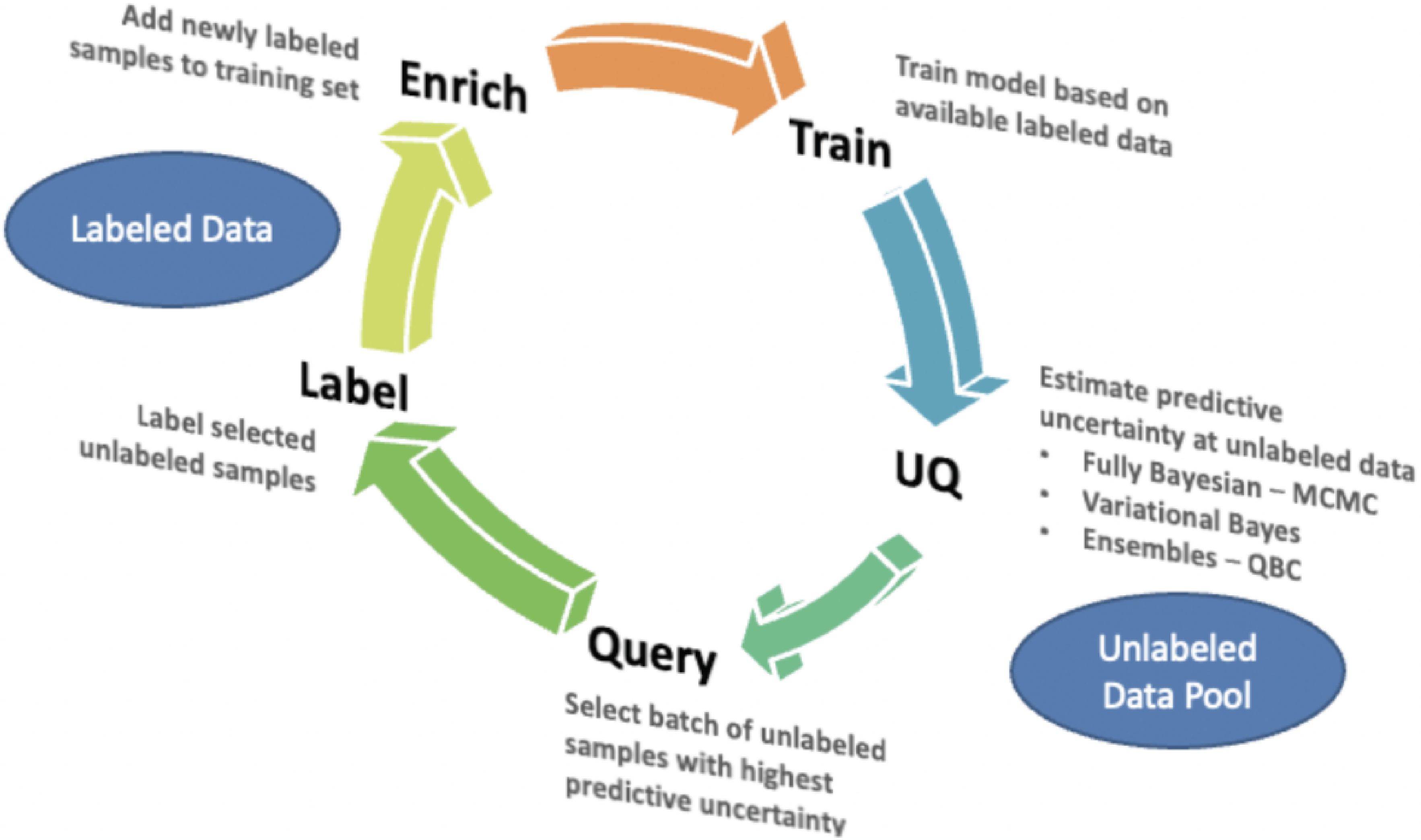
$(w \odot A)b = (w \odot y)$

$\tilde{A}b = \tilde{y}$

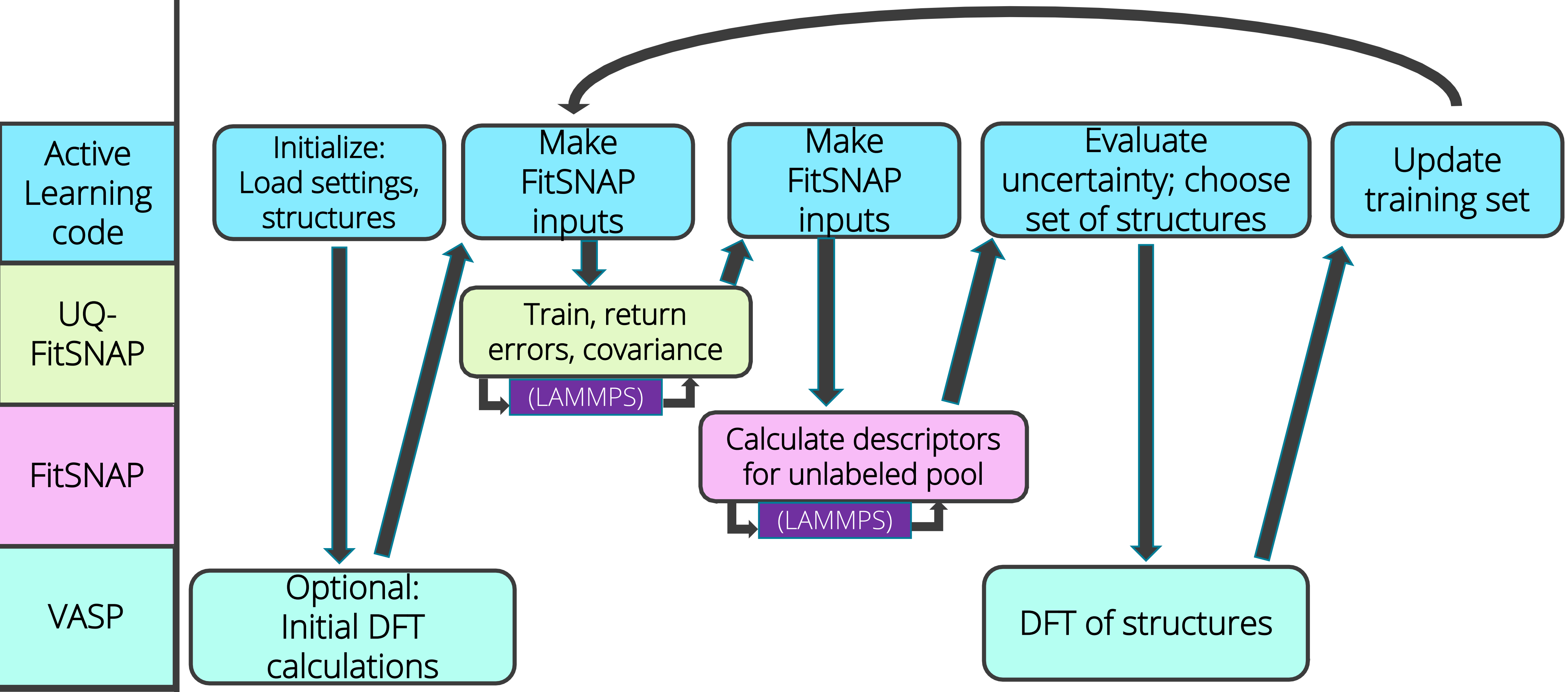
Trained model:

Ab^* = predictions $\mathrm{diag}(\mathbf{A \Sigma A^T})$ = prediction variance

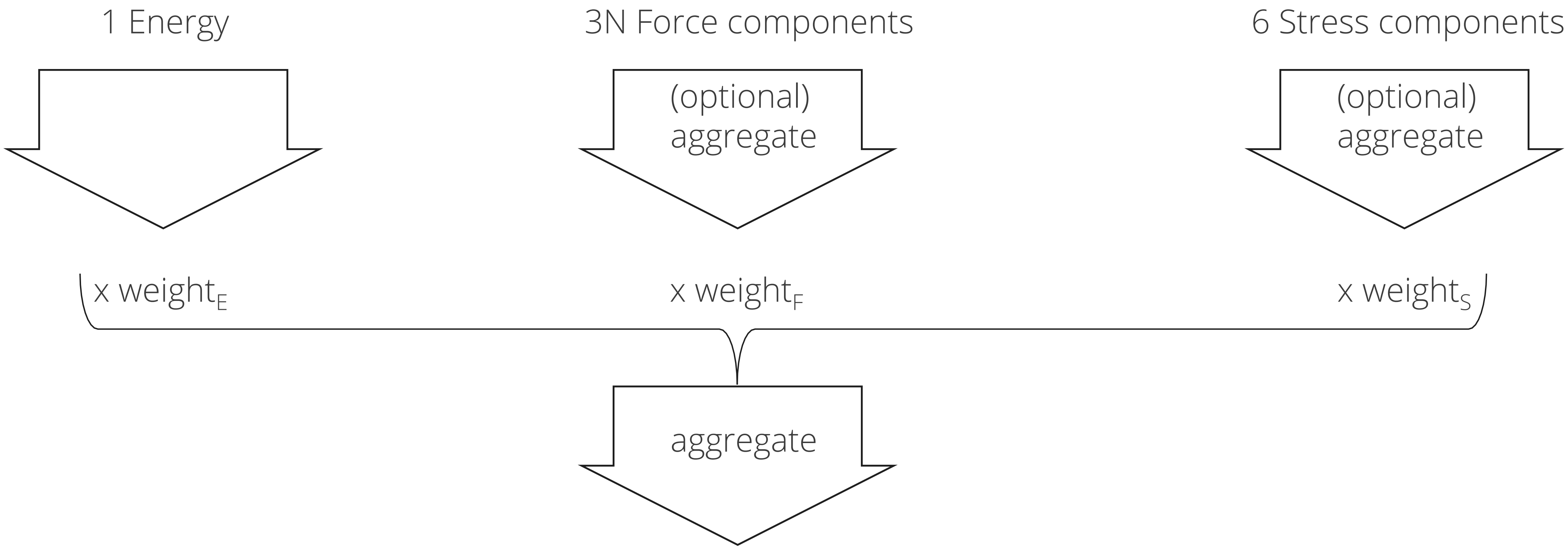
Pool based active learning – helping automate interatomic potential creation



Active Learning Code Structure



Active Learning – structure selection by prediction uncertainties



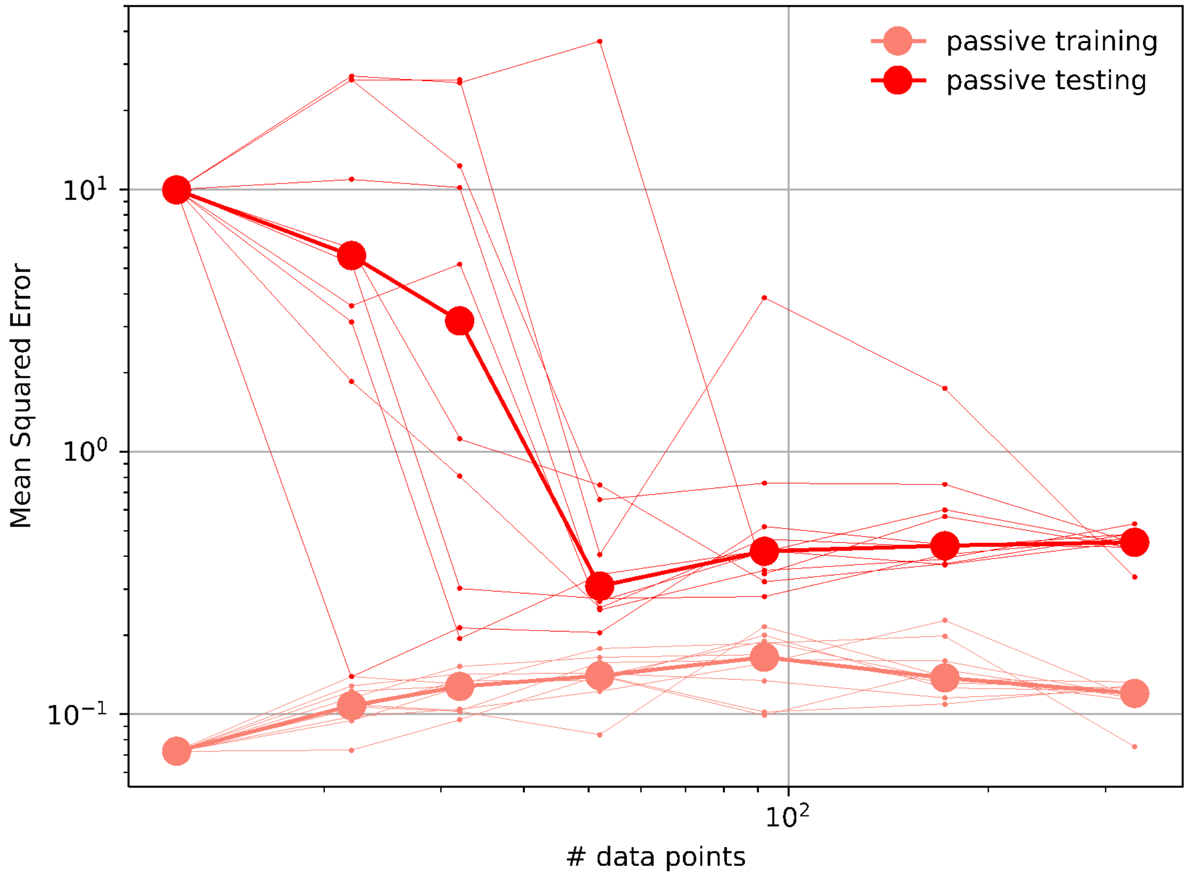
Each aggregate can be: max, mean, median, etc.

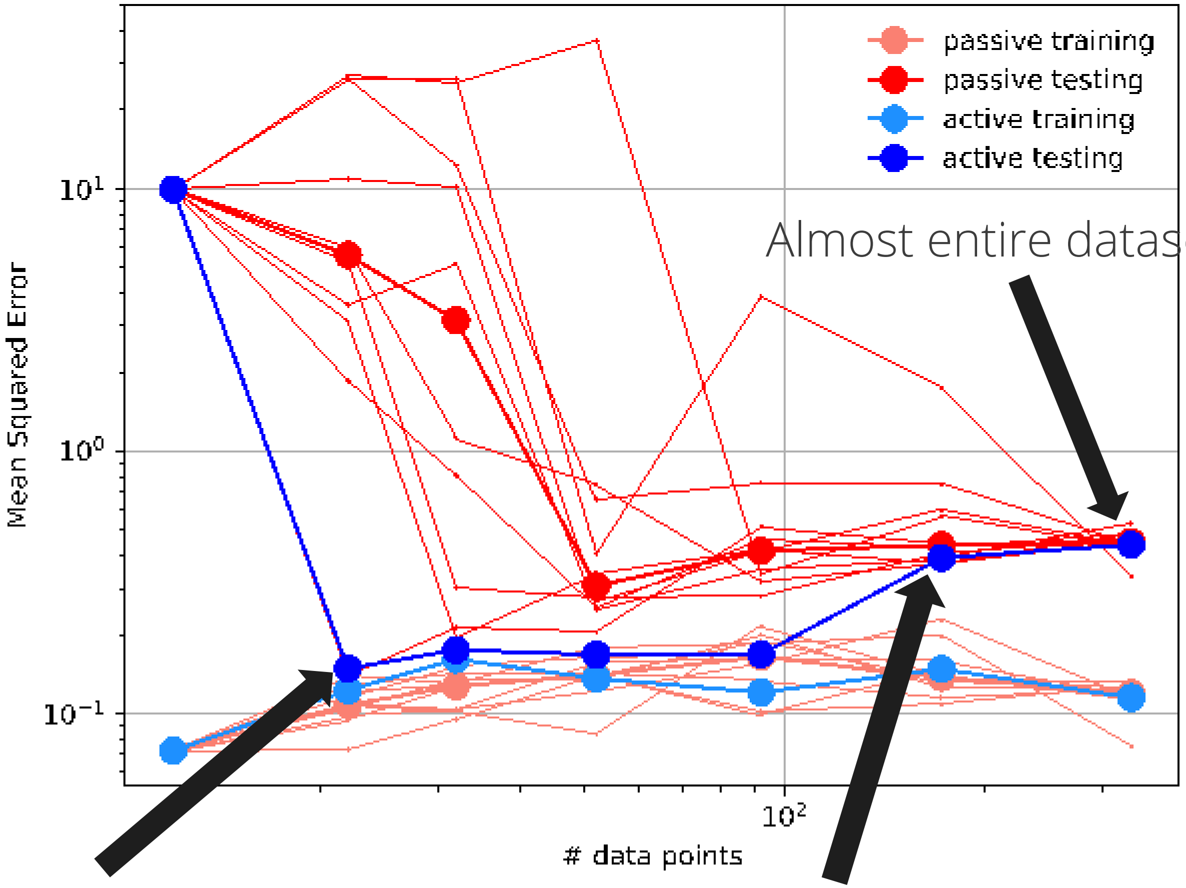
Case Studies

General settings

- Using the example datasets from FitSNAP
  - Begin with a small subset as training set
  - Test set is separated at the start, equally sampled from all defined groups, and held constant
- Use the DAKOTA pre-determined weights from the example files
- Using the ‘anl’ form for UQ approximation

Ta dataset



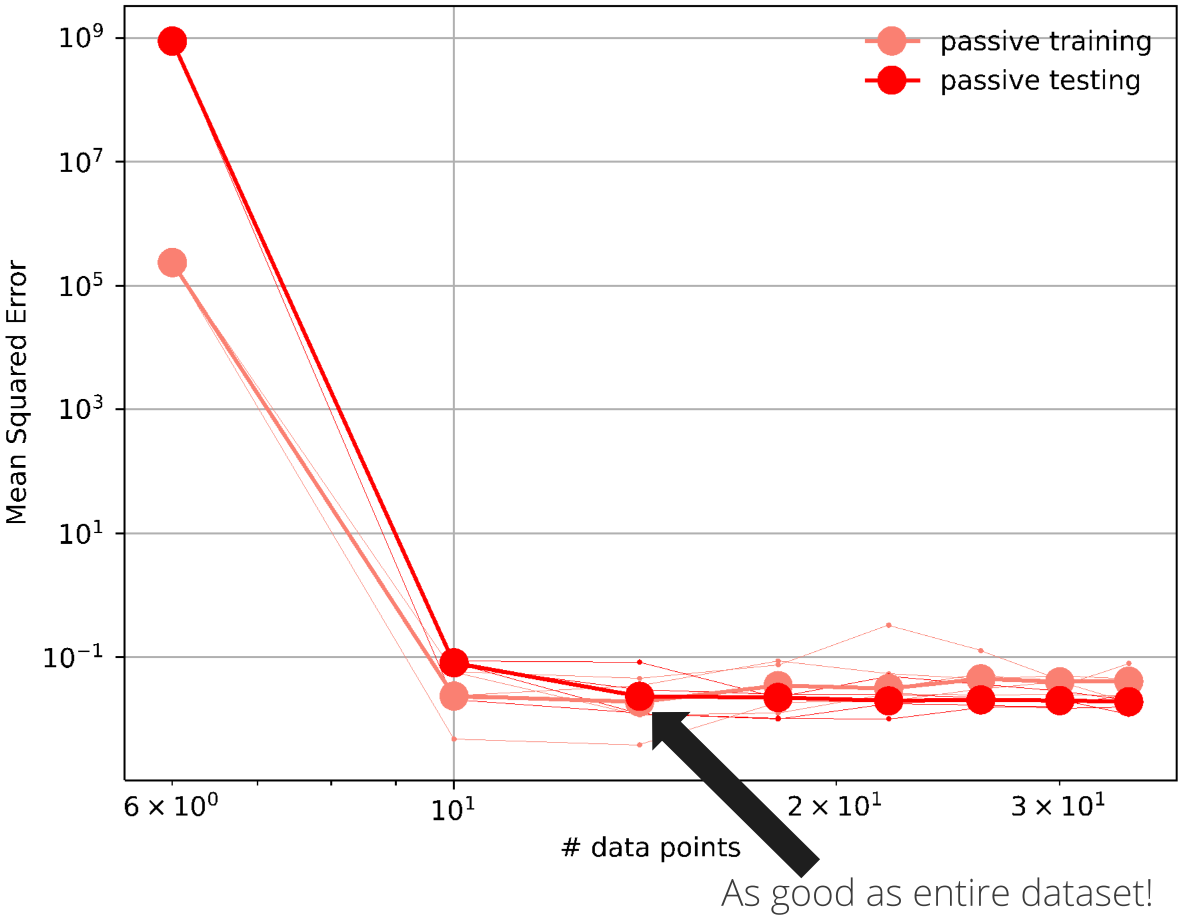


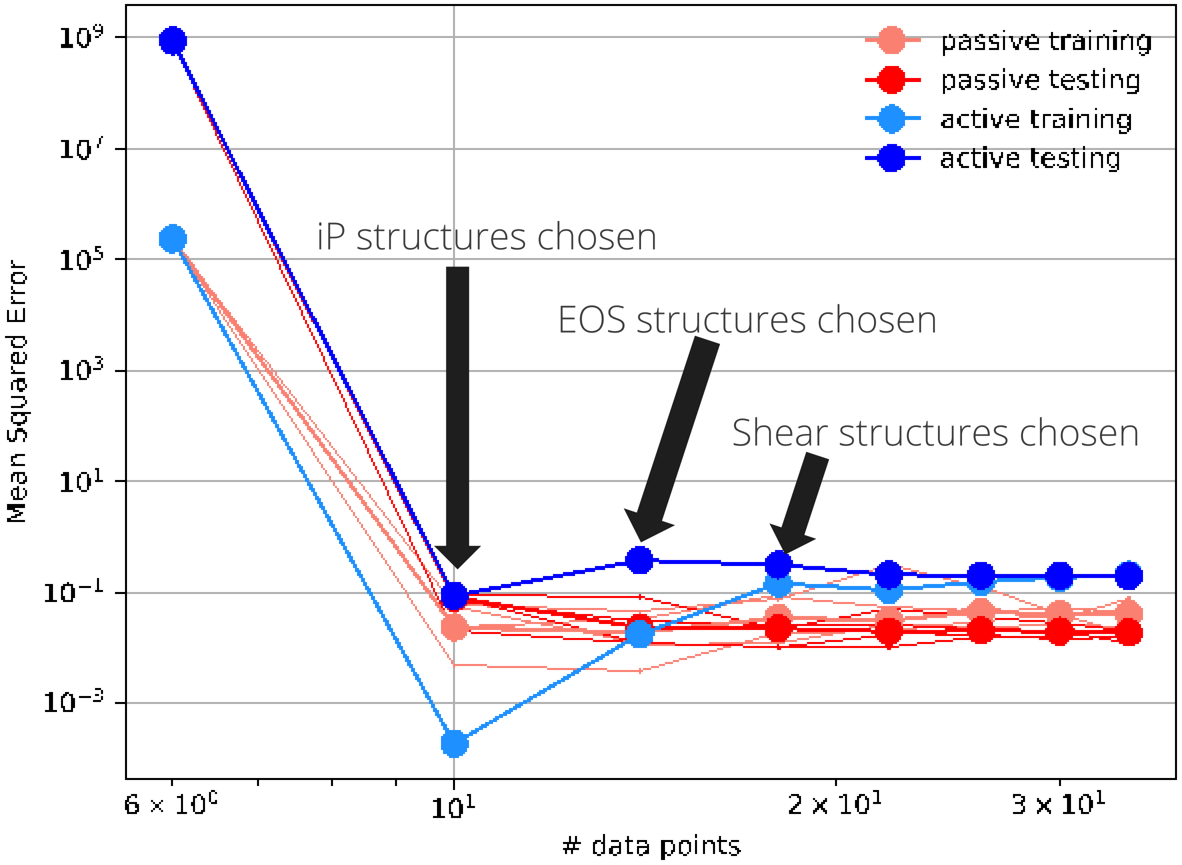
Almost entire dataset

EOS structures chosen

other structures chosen

InP dataset – minimal starting data; models plateau with very few structures

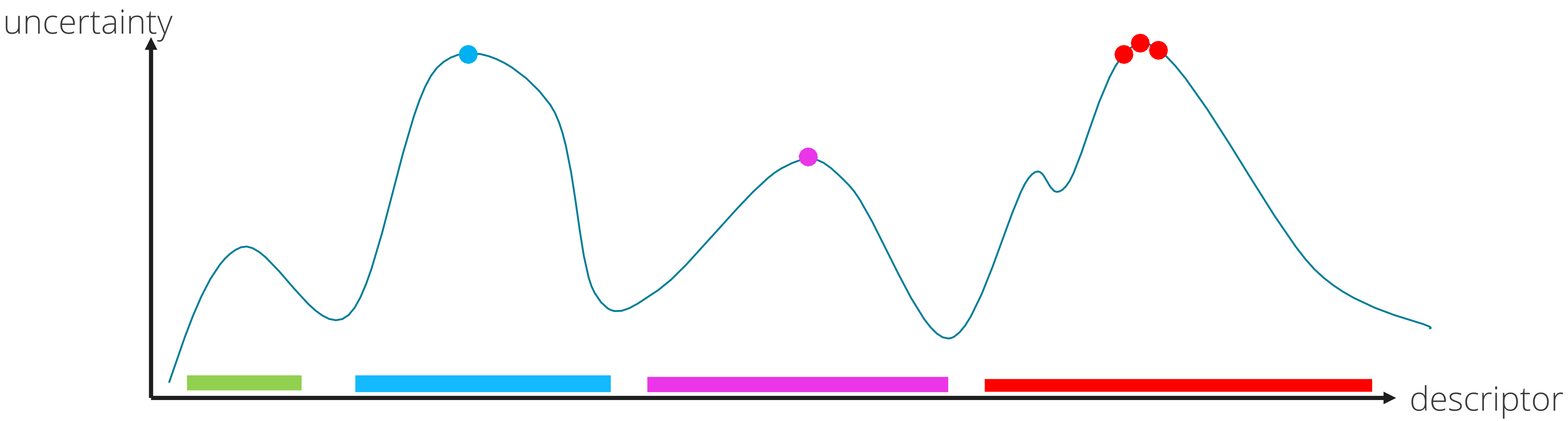




*Passive (random) mostly picks s_iP, Shear, and S_iIn due to dataset group sizes

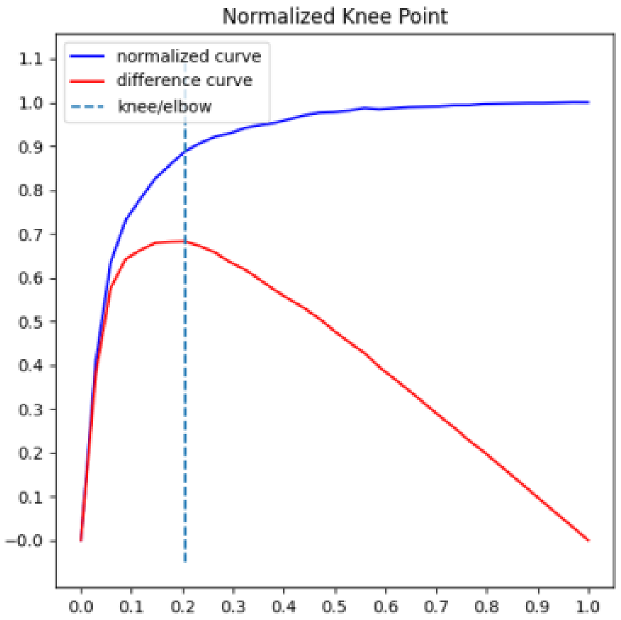
Active learning – motivation for clustering

- Minimal benefit to taking many samples that are very similar
- More efficient to get multiple samples per active learning loop
  - Want to encourage diverse selections

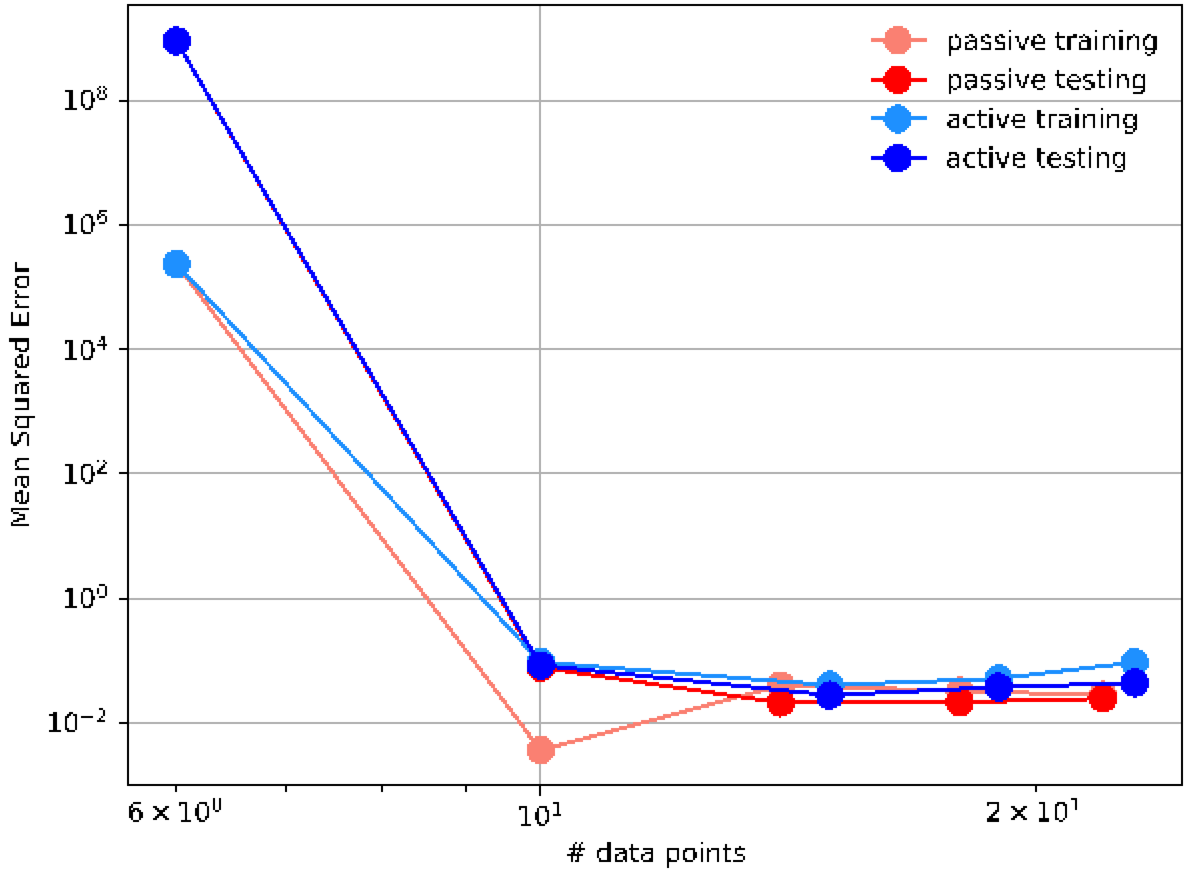


Active learning – clustering

- Kmeans clustering
  - requires assumed # of clusters
  - optimal # of clusters determined by 'knee' method
    - determine sum of squared distances from cluster centers for each # of clusters
    - select point at which the second derivative becomes negative



InP dataset – minimal starting data – clustering



Questions?

The authors would like to thank the DOE Office of Science, Advanced Scientific Computing Research (ASCR) and Fusion Energy Sciences (FES) for funding.