Data driven estimation of atomistic support for continuum stress using the Gaussian mixture model

Sean J. Moran

School of Informatics, University of Edinburgh, Scotland, UK

Manfred Ulz
Graz University of Technology, Austria

Darmstadt, March 2012

Overview

Motivation

Overview

Overview

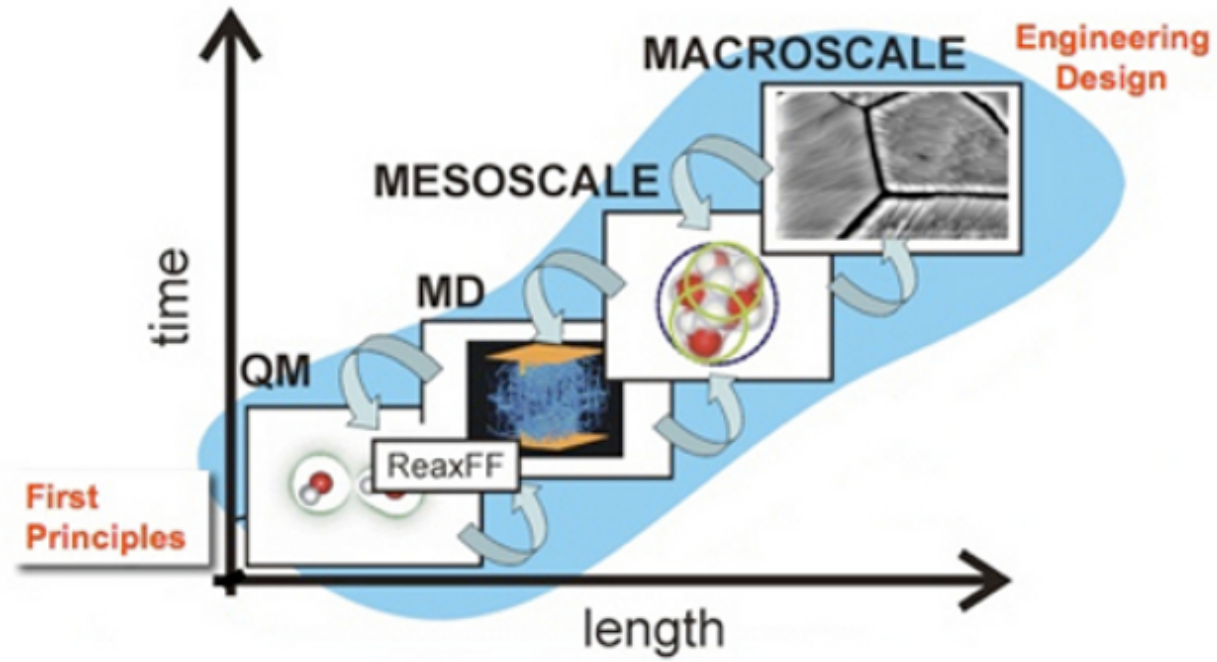
Motivation

Atomistic Support by Gaussian Mixture Modelling

Numerical example: edge dislocation

Motivation

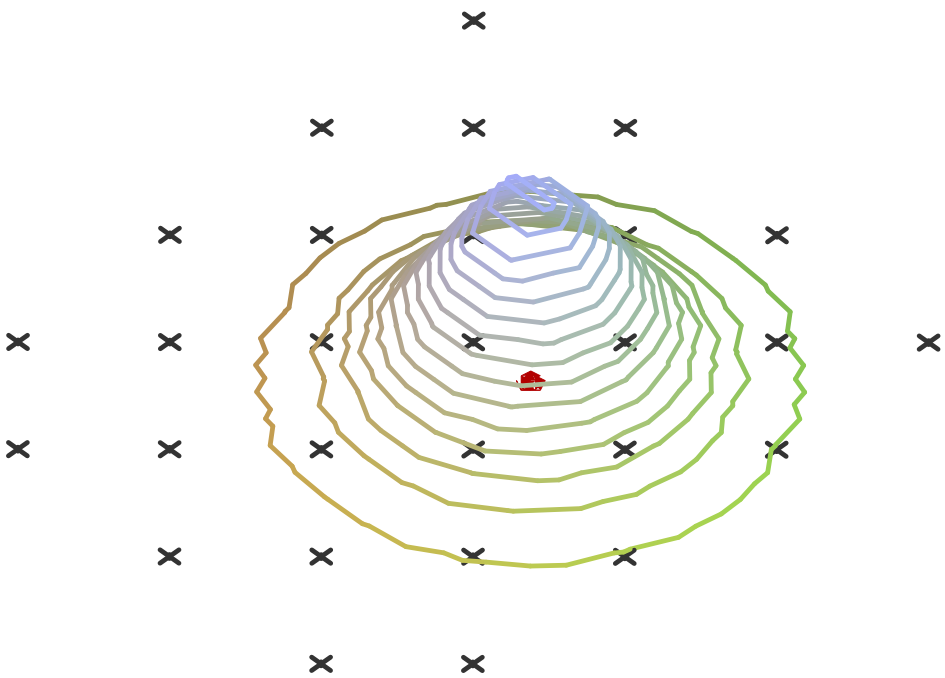
Atomistic to continuum coupling: time and length scale



Atomistic to continuum coupling

Atomistic definition of stress (Hardy stress)

$$\boldsymbol{\sigma}(\mathbf{x}) := -\sum_{i=1}^{N} m^i \mathbf{u}^i \otimes \mathbf{u}^i \psi(\mathbf{x} - \mathbf{x}^i) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \mathbf{f}^{ij} \otimes \mathbf{x}^{ij} B^{ij}(\mathbf{x})$$



Kernel function

$\psi(\mathbf{x} - \mathbf{x}^i)$

Bond function

$$B^{ij}(\mathbf{x}) = \int_0^1 \psi(\mathbf{x} - \mathbf{x}^i + \lambda \mathbf{x}^{ij})\, d\lambda$$

Atomistic Support by Gaussian Mixture Modelling

Hypothesis

Atoms with similar values for position and stress are contained within the support of the kernel function.

We can discover the size of this support by *clustering* related atoms together in position-stress space.

Atomistic Support by Gaussian Mixture Modelling

Gaussian Mixture Model

- Model each component as a Gaussian distribution parameterized by μ_k, Σ_k. Denote the data by $\mathbf{X}$, $\mathbf{X} \in \mathbb{R}^d$. The density of component k is:

$$P(\mathbf{x}|k) = \phi(\mathbf{x}|\mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} exp(\frac{-(\mathbf{x}-\mu_k)^t \mathbf{\Sigma}_k^{-1}(\mathbf{x}-\mu_k)}{2})$$

- The prior probability (weight) of component k is π_k. The mixture density is:

$$P(\mathbf{x}) = \sum_{k=1}^{K} \pi_k P(\mathbf{x}|k) = \sum_{k=1}^{K} \pi_k \phi(\mathbf{x}|\mu_k, \mathbf{\Sigma}_k)$$

Atomistic Support by Gaussian Mixture Modelling

Letting the data speak for itself

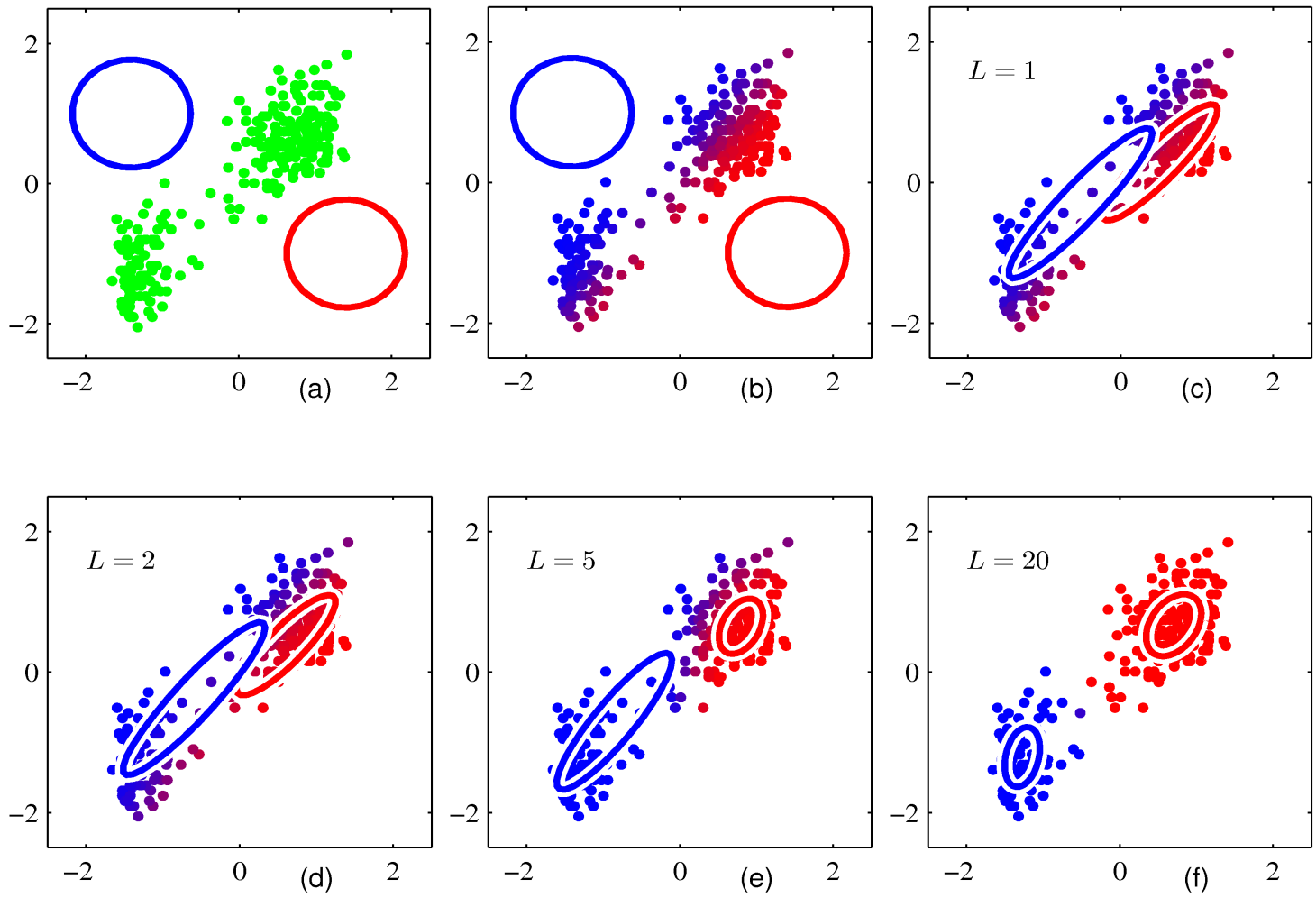
- We make no a-priori assumptions on how many clusters there are or which atoms belong to which cluster: **let the data "speak" for itself**.
- Two challenges:
  - Which mixture component(s) should be responsible for which data points?
  - How to automatically select the optimal number of mixture components?
- Note on terminology: Component⇔Cluster⇔Continuum Point

Atomistic Support by Gaussian Mixture Modelling

EM Algorithm for Mixture of Gaussians

- Expectation Maximization (EM)(*A. Dempster et al. Journal of the Royal Statistical Society, 1977*): optimization algorithm to learn GMM parameters from data.
- In a nutshell:
  - Start with K randomly placed Gaussians $(\mu_1, \mathbf{\Sigma}_1)$,...,$(\mu_k, \mathbf{\Sigma}_k)$.
  - **E-Step:** For each point $\mathbf{x}_i$: $P(k|\mathbf{x}_i)$ = does it look like it came from Gaussian k?
  - **M-Step:** Adjust $(\mu_1, \mathbf{\Sigma}_1), \ldots, (\mu_k, \mathbf{\Sigma}_k)$ to fit assigned points.
  - Stop when we maximize the likelihood of our data: $\text{argmax}_{((\mu_1,\mathbf{\Sigma}_1),\ldots,(\mu_k,\mathbf{\Sigma}_k))} P(\mathbf{X}|(\mu_1, \mathbf{\Sigma}_1), \ldots, (\mu_k, \mathbf{\Sigma}_k))$.
- Guaranteed to converge to local maximum.

Atomistic Support by Gaussian Mixture Modelling



Atomistic Support by Gaussian Mixture Modelling

BIC for Mixture of Gaussians

- Bayesian Information Criterion (BIC) (*G. Schwarz. Annals of Statistics, 1978*). Method for model order selection:

$$BIC \Rightarrow -\log P(\mathbf{X}|(\mu_1, \mathbf{\Sigma}_1), \ldots, (\mu_k, \mathbf{\Sigma}_k)) + \frac{1}{2} M log N$$

- $N = \#$data points, $M = \#$parameters.
- Model minimizing BIC is a robust estimate of underlying data generating process.

Numerical example: edge dislocation in fcc copper

Data space

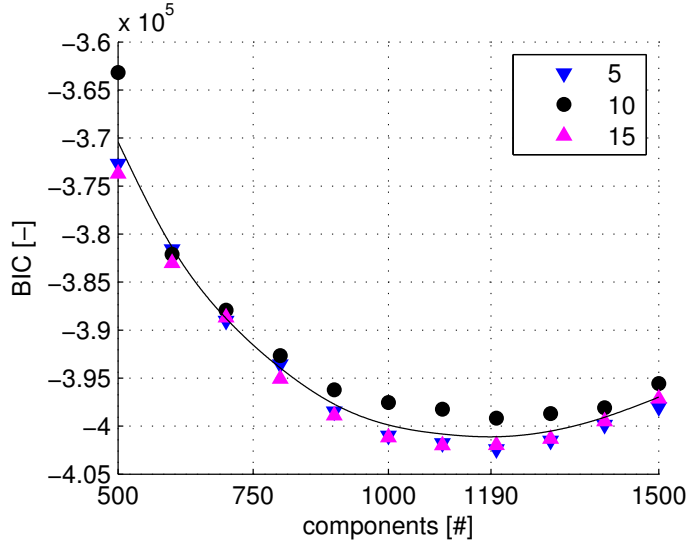
- Crystal at zero Kelvin, energy minimization using LAMMPS.
- Each atom about edge dislocation core gives a data point in 5d position-stress space.
- Correlated data points are discovered by the GMM model.

Numerical example: edge dislocation in fcc copper

Data space

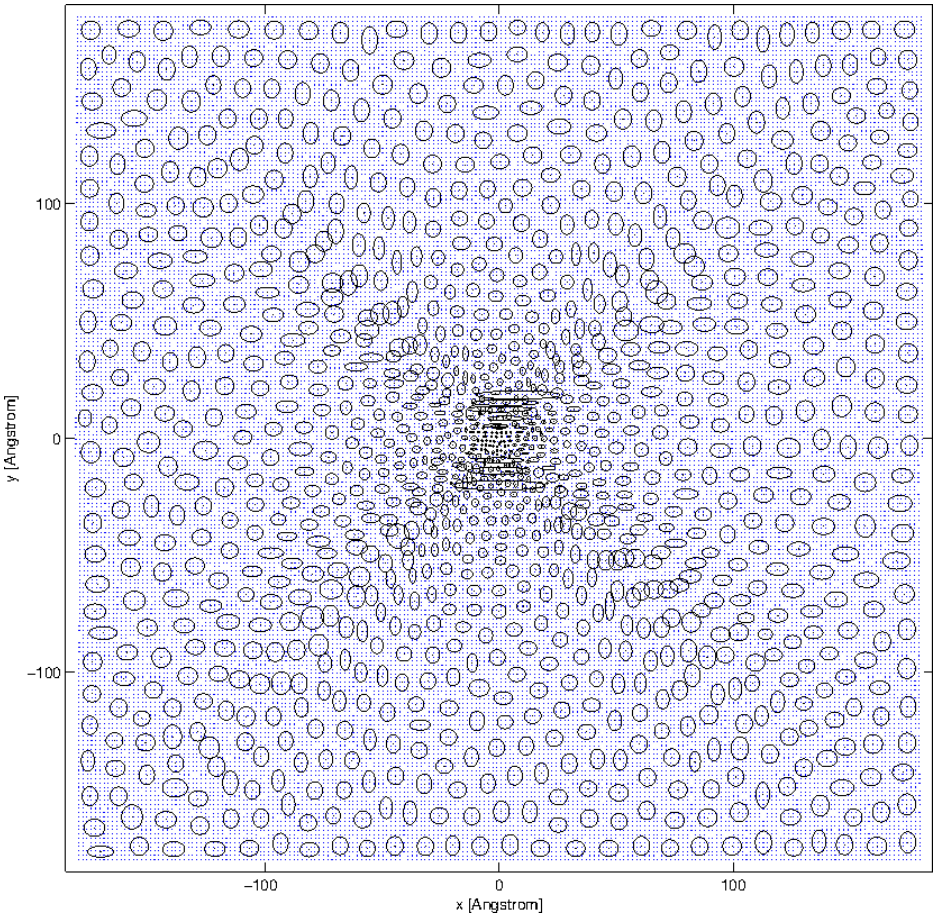
- Crystal at zero Kelvin, energy minimization using LAMMPS.
- Each atom about edge dislocation core gives a data point in 5d position-stress space.
- Correlated data points are discovered by the GMM model.

Minimum BIC search



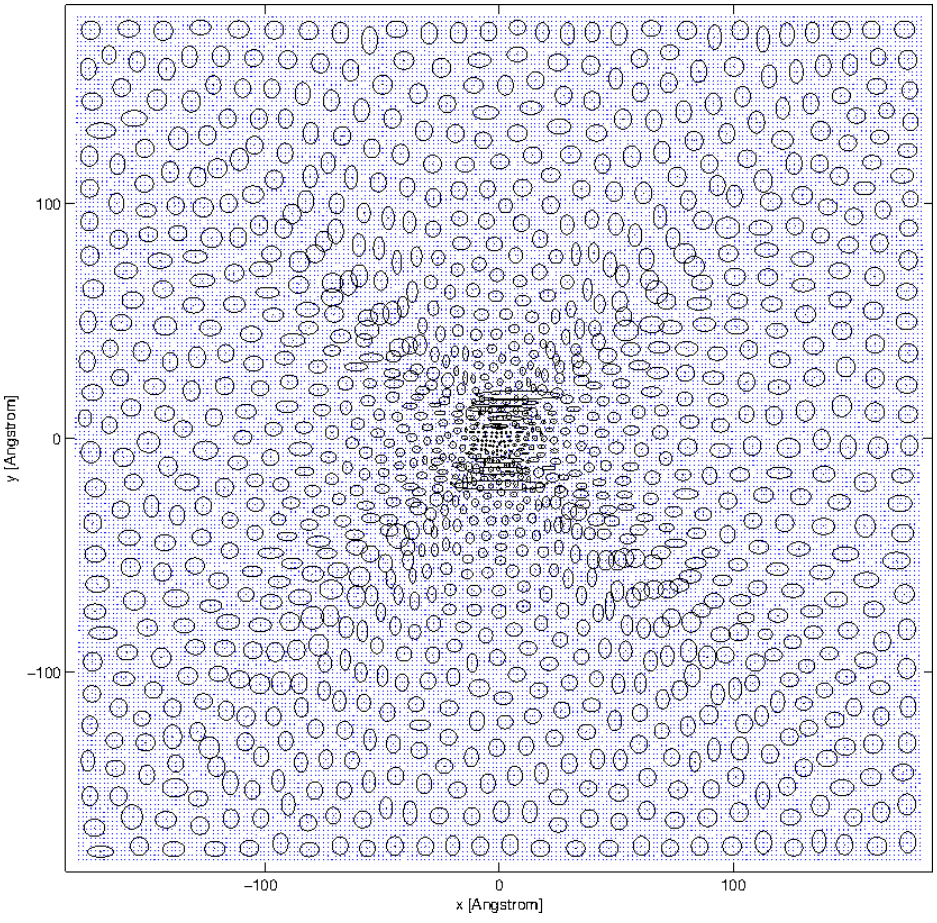
Numerical example: edge dislocation in fcc copper

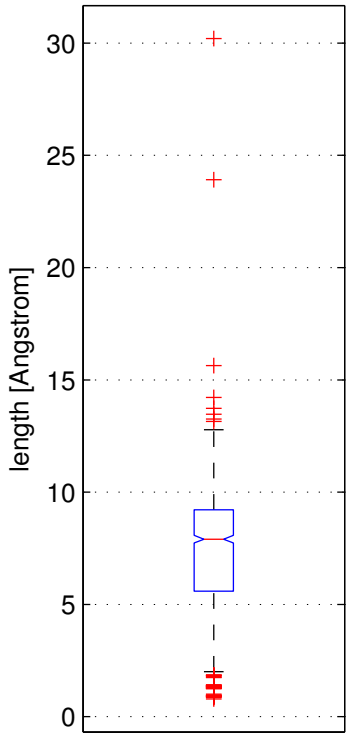
Location of components



Numerical example: edge dislocation in fcc copper

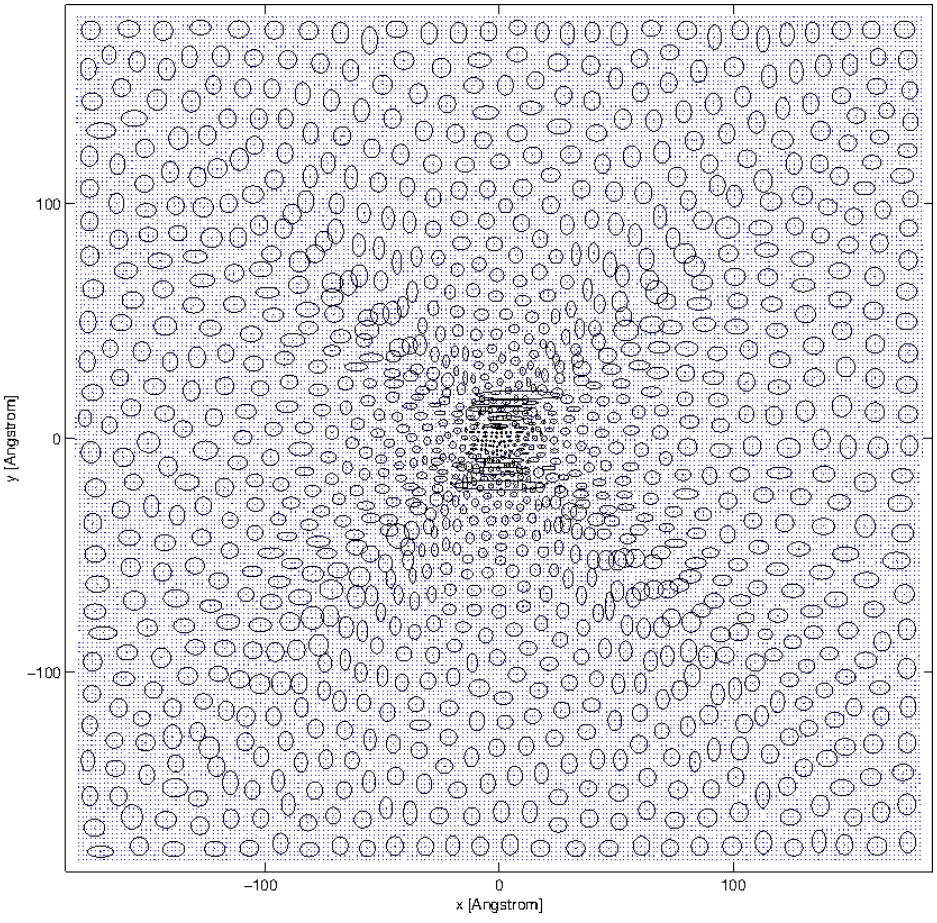
Location of components

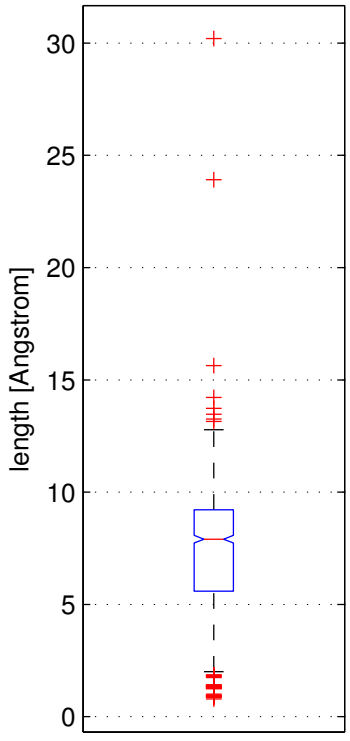




Numerical example: edge dislocation in fcc copper

Location of components

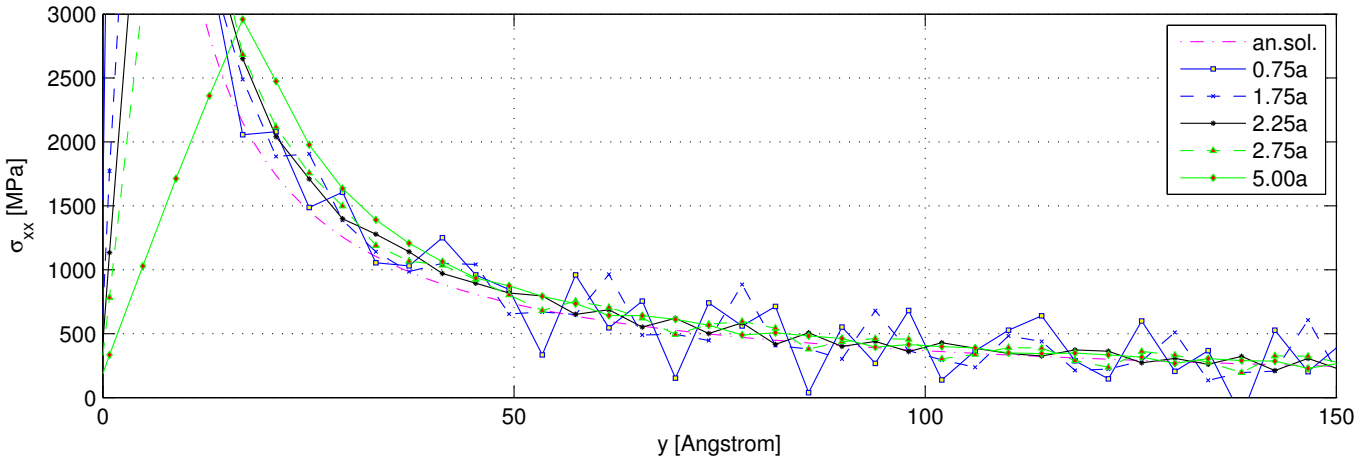


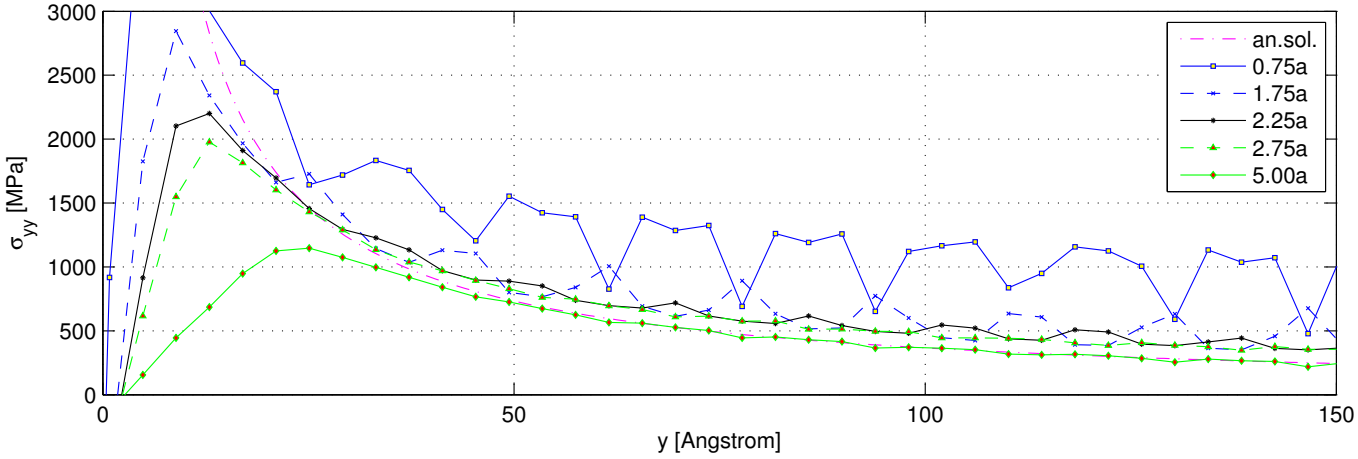


Median = 7.906 Å

Numerical example: edge dislocation in fcc copper

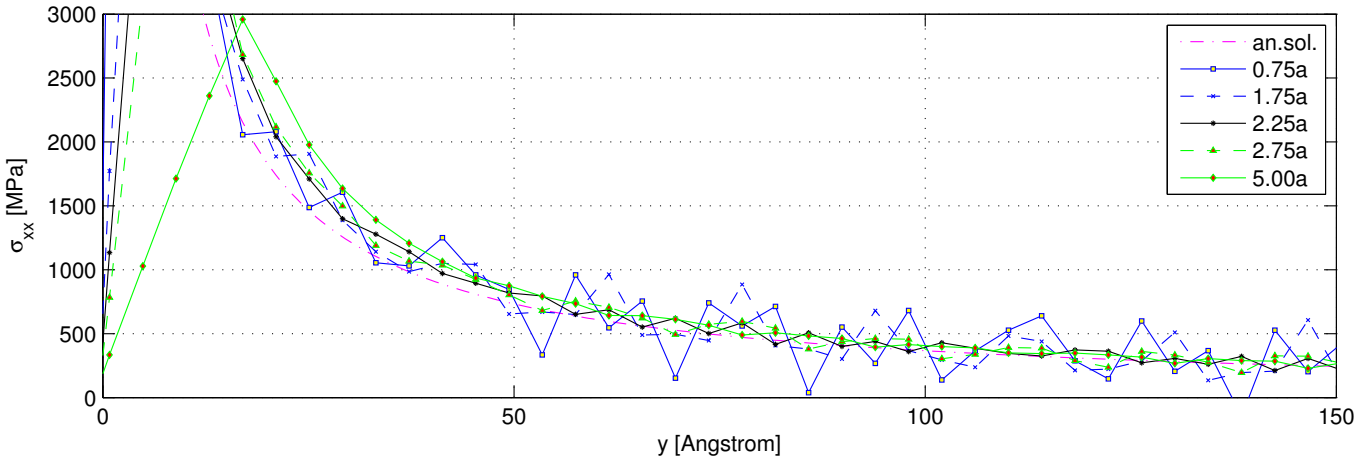
Hardy stress with LAMMPS

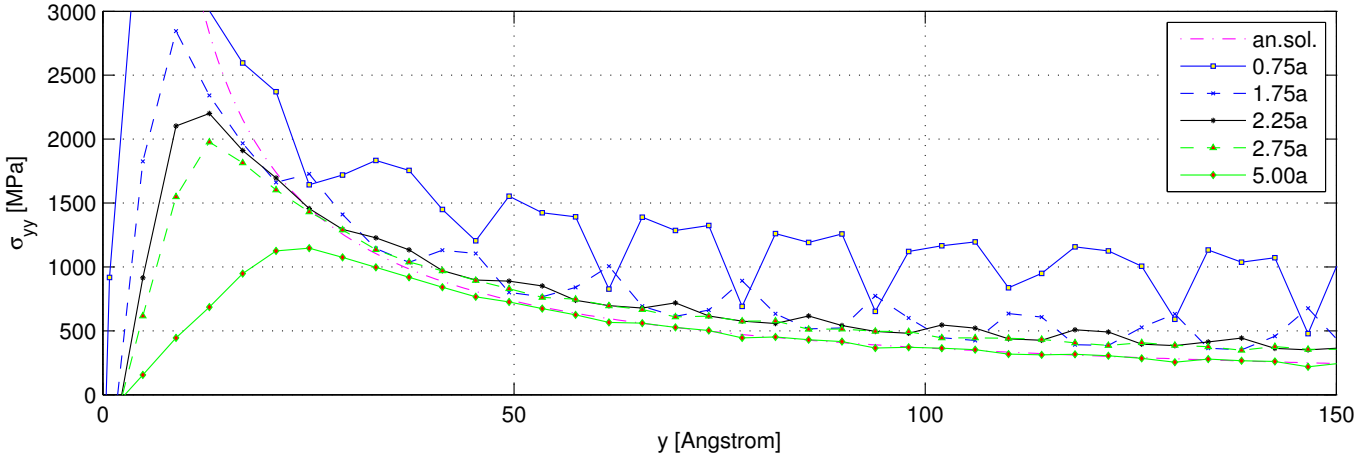




Numerical example: edge dislocation in fcc copper

Hardy stress with LAMMPS

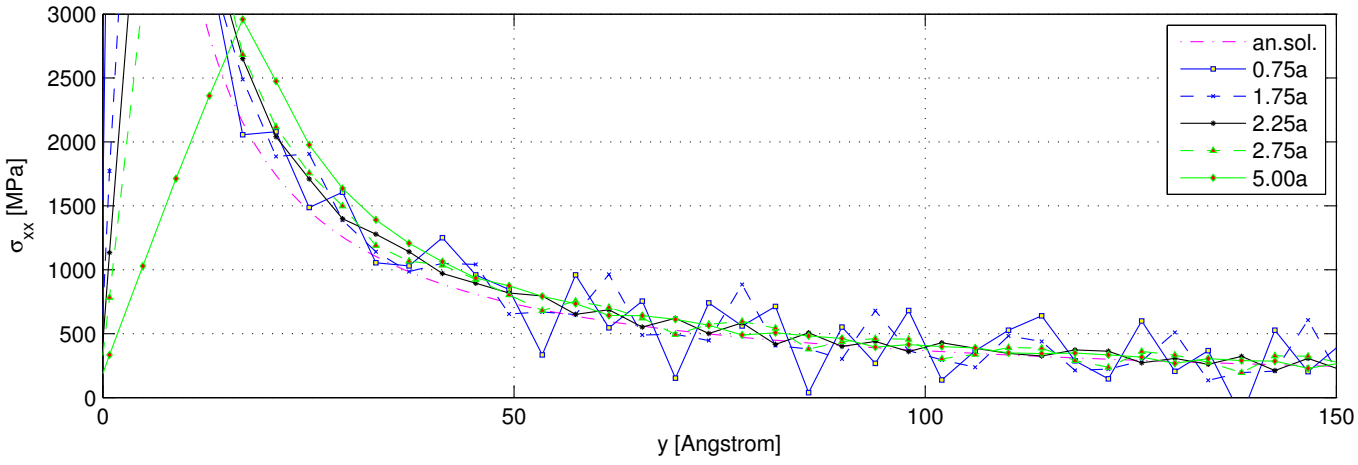


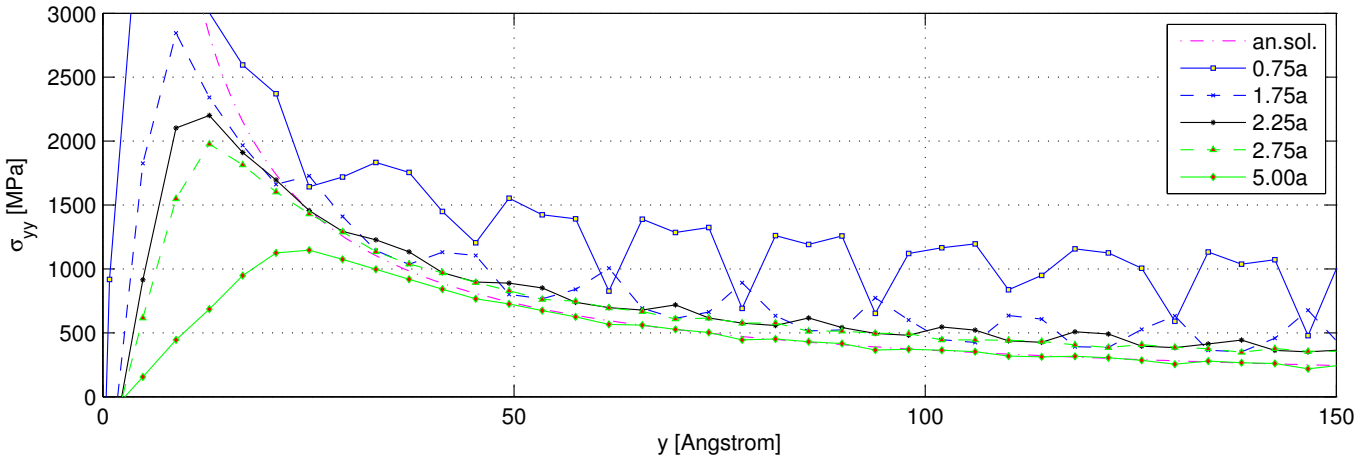


Convergence to analytical solution at appr. 2.25a.

Numerical example: edge dislocation in fcc copper

Hardy stress with LAMMPS



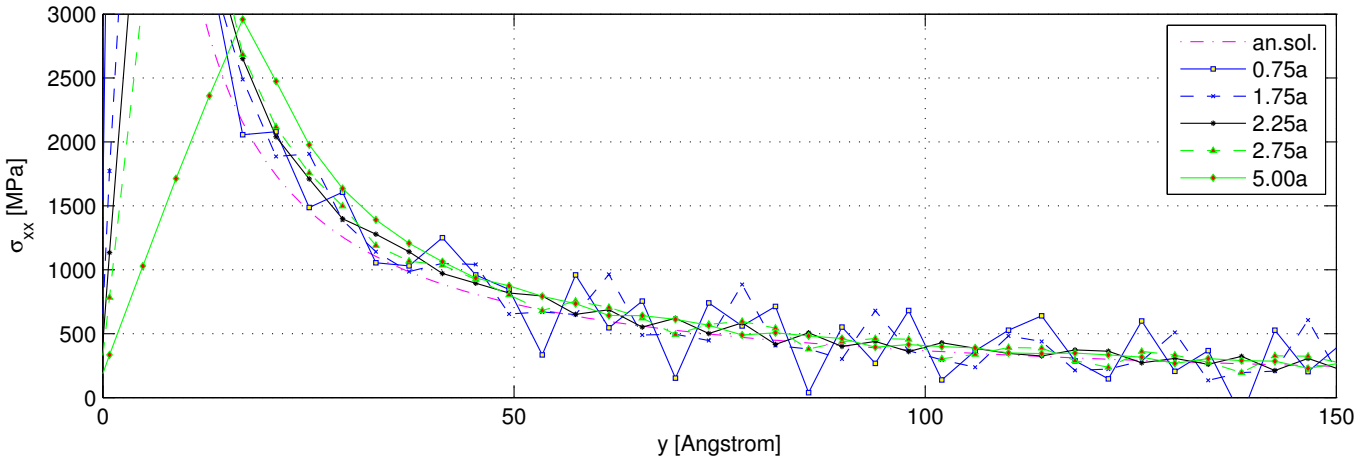


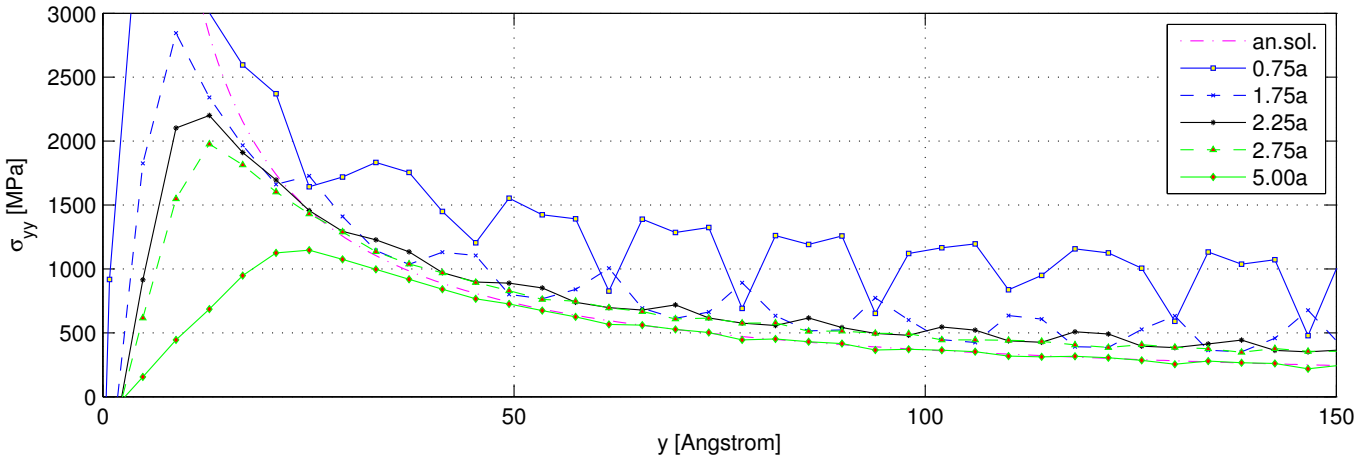
Convergence to analytical solution at appr. 2.25a.

Atomistic simulation: 8.134 Å

Numerical example: edge dislocation in fcc copper

Hardy stress with LAMMPS





Convergence to analytical solution at appr. 2.25a.

Atomistic simulation: 8.134 Å

GMM: 7.906 Å

Thank you for your attention.