Characterizing the Uncertainties in Non-Equilibrium MD for Thermal Transport

Manav Vohra[†], Sankaran Mahadevan[†]

Collaborators: Seungha Shin[§], Ali Yousefzadi Nobakht[§], Alen Alexanderian[‡]

[†] Vanderbilt University

[§] University of Tennessee, Knoxville

[‡] North Carolina State University, Raleigh

January 22, 2018

BACKGROUND

- Classical MD is used to investigate heat transfer dominated by phonon-phonon interactions in material systems.
  - Commonly applied to study non-metallic systems like C, Si, and Ge.
- Typically conducted under equilibrium conditions characterized by thermodynamic ensembles like NVT, NVE, NPT, and μVT.
- Non-Equilibrium MD involves setting up thermostats in different regions to establish temperature gradients.
  - Thermostatting introduces errors.

WHY MD?

- Enables simulation of much larger systems compared to DFT in a reasonable amount of time.
- Trends from MD are useful despite possible errors in estimates.

PLAN

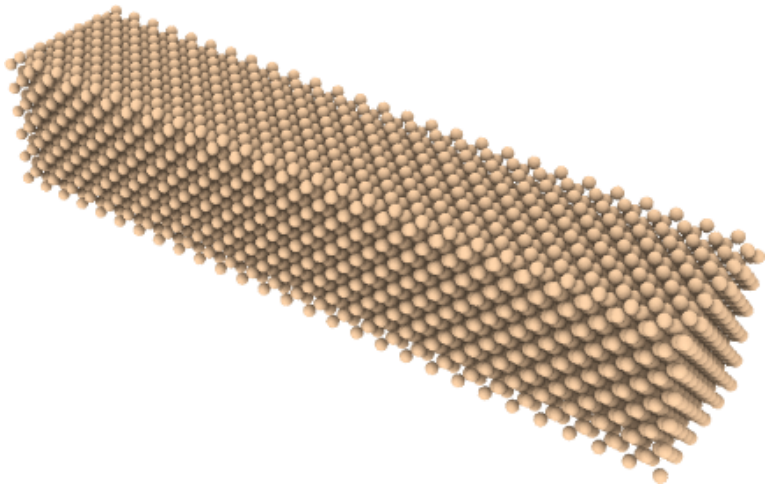
Part I: FORWARD PROBLEM:

- Investigate error in predictions due to size of the material system and fluctuations in thermal gradient.
  - Construct a response surface for the error.
- Characterize the impact of uncertainty in inter-atomic potential on predictions.
  - Dimension reduction using derivative-based sensitivity measures.
  - Construction of PCE in the reduced space.
  - Forward Propagation of uncertainty.

Part II: INVERSE PROBLEM:

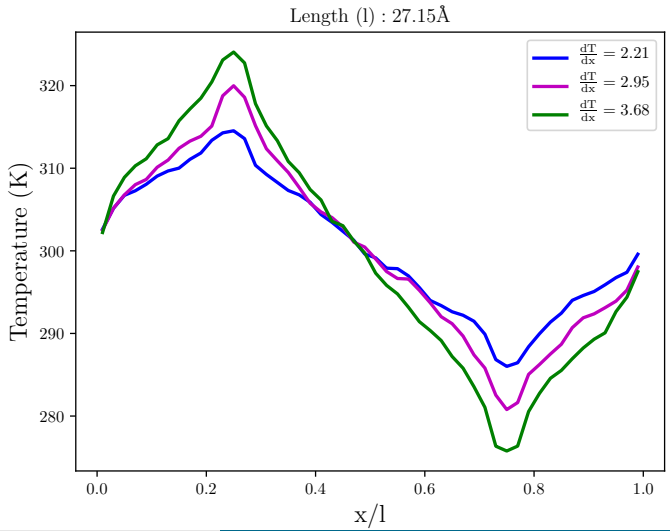
- Calibrate critical parameters associated with the potential function in a Bayesian setting using experimental data.

NEMD ON A SILICON BAR



$L = 50a$, $N \approx 200000$ atoms

Lattice Constant, a (Å)	5.43
W, H (Å)	$22a$, $22a$
Temperature (K)	300
Δt (ps)	0.0005
BC	Periodic
Structure	Diamond
Potential	Stillinger-Weber



DIRECT METHOD

NVT $\rightarrow$ NVE $\rightarrow$ NVE

[Eqilibrate system to 300 K] [Equilibrate thermostats] [Generate Data]

N: Number of Atoms V: Volume T: Temperature E: Energy

OBSERVABLE: Average energy exchange b/w thermostats (q)

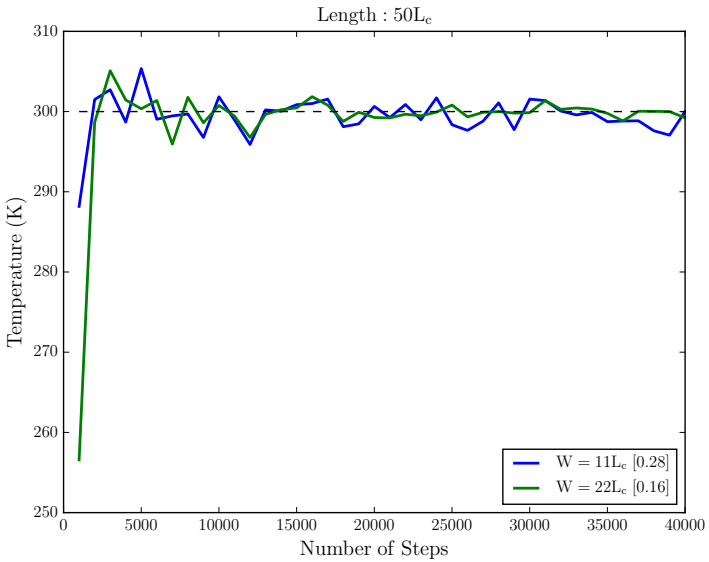
QOI: Bulk thermal conductivity (κ)

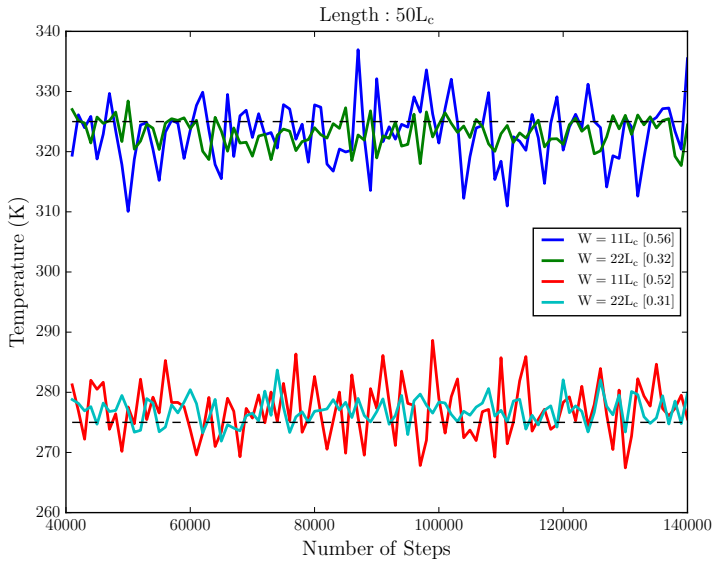
$$\kappa = \frac{q}{\left|\frac{dT}{dx}\right|}$$

INITIAL RUNS:

- Determine the time steps needed for equilibration.
- Select a reasonable width and height for the Si bar.
  - Small fluctuations due to changes in height and width.

SELECTION OF WIDTH

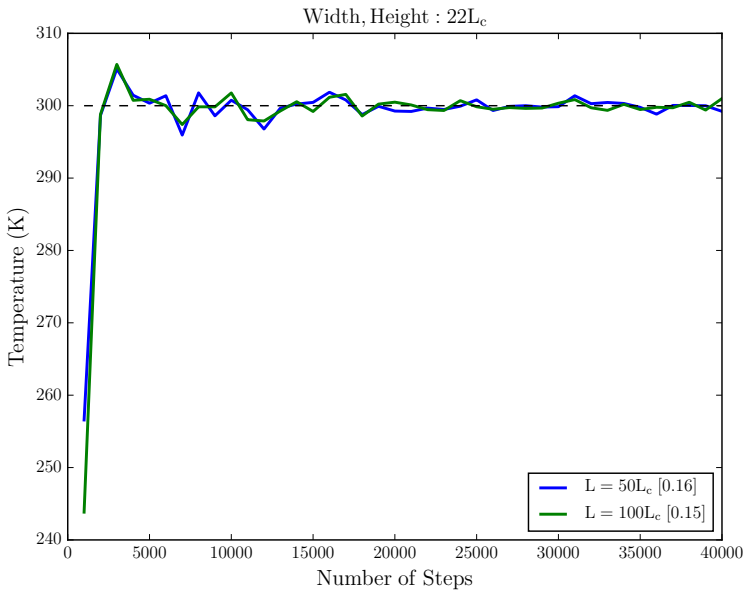


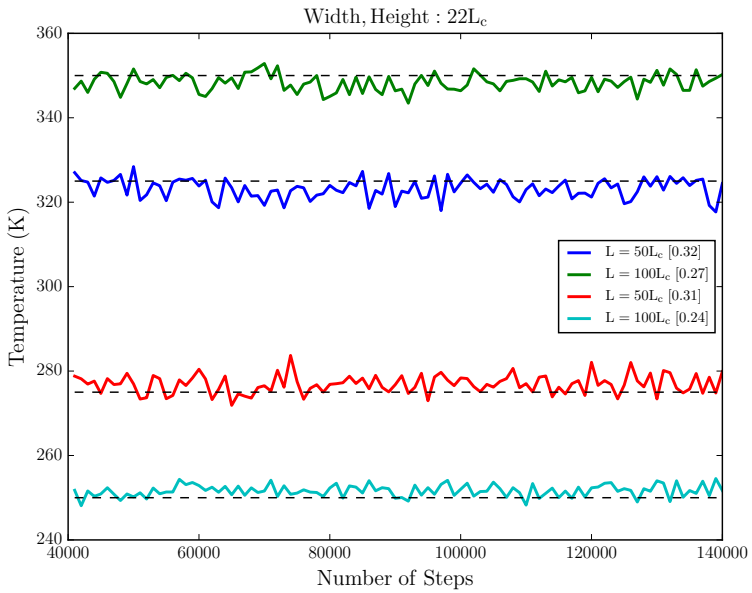


- Norm of the fluctuations (NF) is computed using:

$$NF = \frac{1}{N}\left[\sum_{k}\left(T_k - T_{\{nvt,nve\}}\right)^2\right]^{\frac{1}{2}}$$

SELECTION OF WIDTH

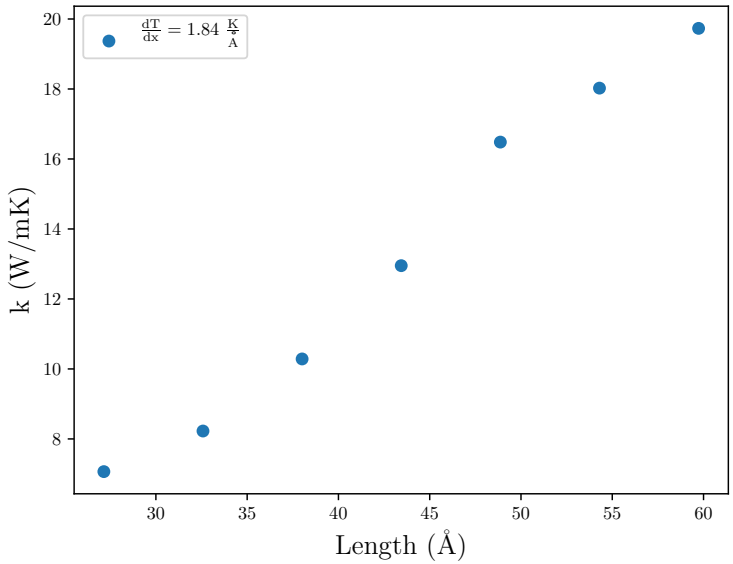




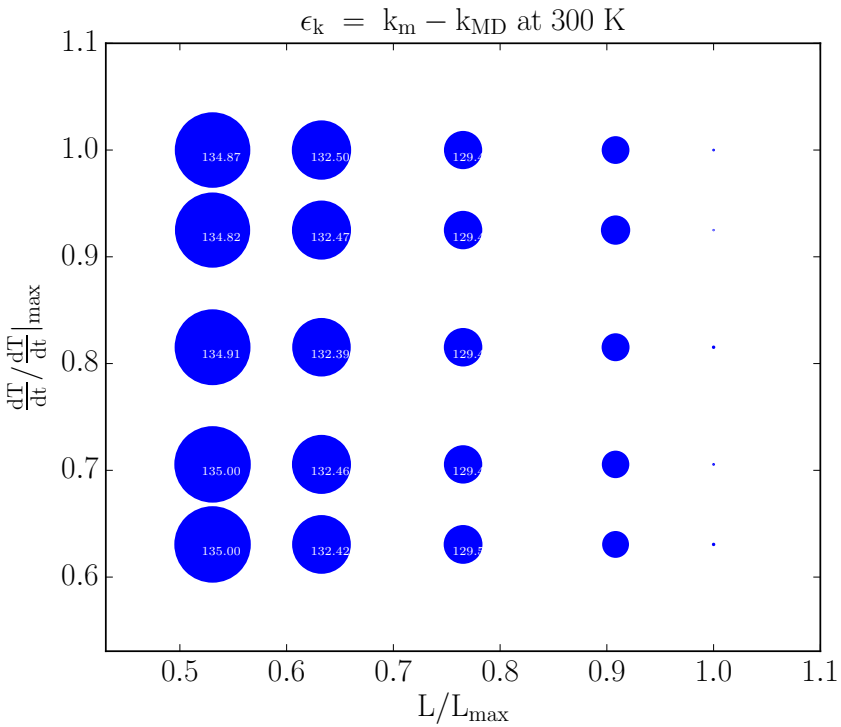
- At $W = 22L_c$, the effect of length on fluctuations seems negligible.

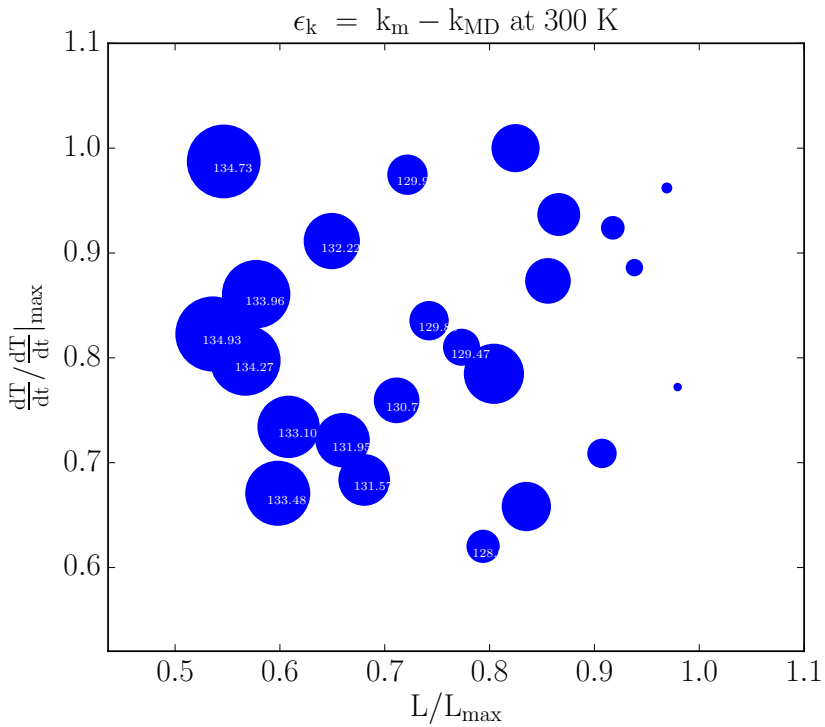
Need a Surrogate?

- Objective: Forward UQ, Sensitivity Analysis, calibration, Design
- Computational Effort: Simulations are computationally intensive.
- Accuracy: Can a surrogate represent the observable with reasonable accuracy in the domain of interest?



MODEL REALIZATIONS



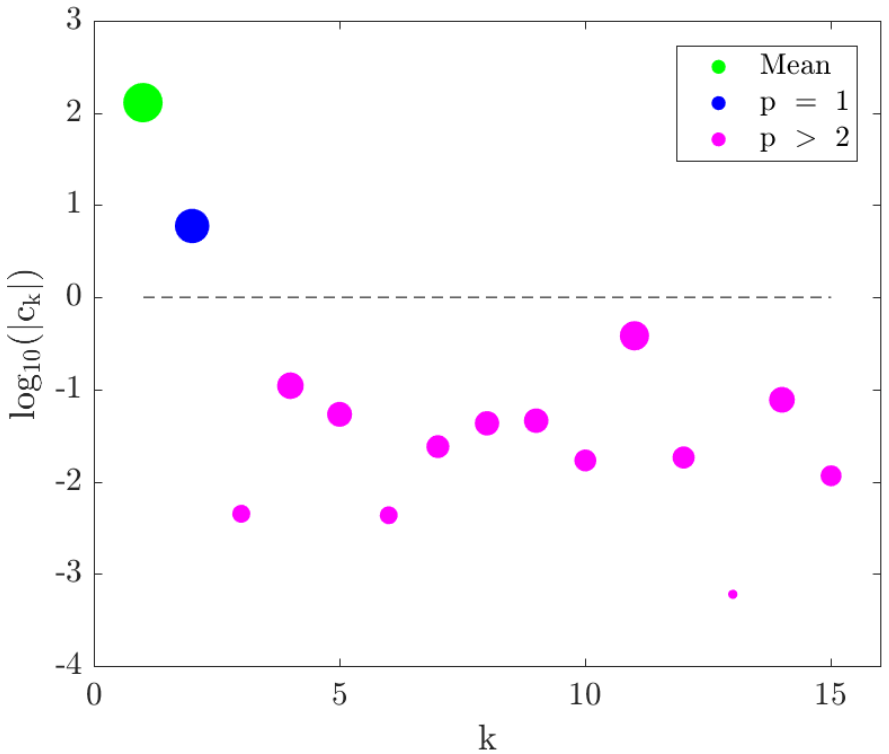


- Model realizations at Gauss-Legendre quadrature nodes are used to construct the PC surrogate.

PC EXPANSION

$$\epsilon(= |\kappa_m - \kappa_{\text{MD}}|) = \sum_j c_j \Psi_j(\xi_1, \xi_2)$$

κ_m: Measured, κ_{MD}: MD Prediction, j: Multi-index

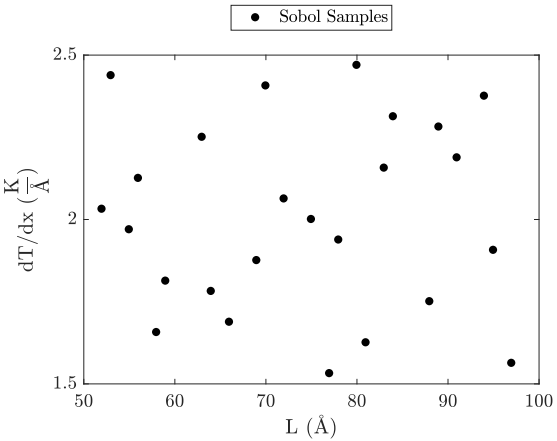


$$L: \mathcal{U}[50L_c, 100L_c]\ (\text{Å}) \quad \rightarrow \quad \xi_1 : \mathcal{U}[-1, 1]$$

$$\frac{dT}{dx}: \mathcal{U}[1.5/L_c, 2.5/L_c]\ \left(\frac{K}{\text{Å}}\right) \quad \rightarrow \quad \xi_2 : \mathcal{U}[-1, 1]$$

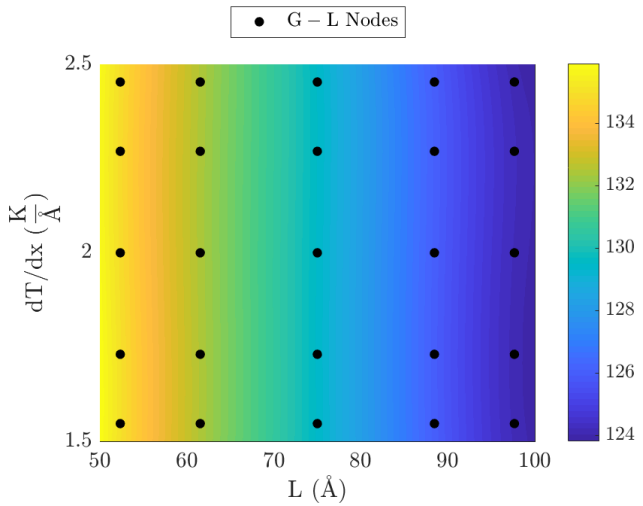
RESPONSE SURFACE: $\epsilon(L, \frac{dT}{dx})$

ACCURACY:

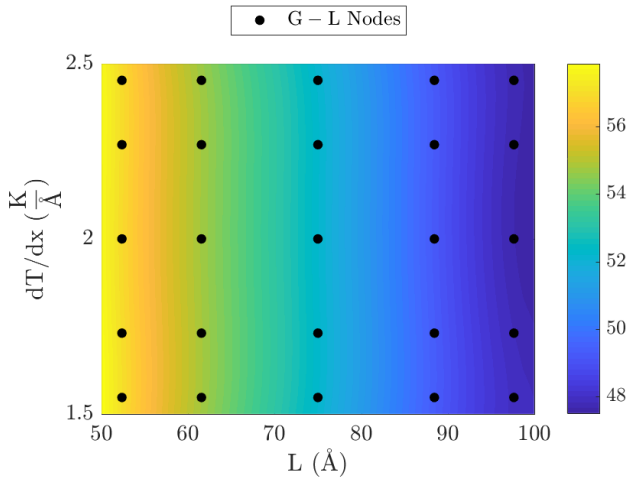


$$\epsilon = \frac{\left[\sum_j (\mathcal{G}_M - \mathcal{G}_{PCE}))^2\right]^{\frac{1}{2}}}{\left[(\mathcal{G}_M)^2\right]^{\frac{1}{2}}} \approx 1.8 \times 10^{-3}$$

$\mathcal{G}_M$: Model Output $\mathcal{G}_{PCE}$: PCE Estimate



(a) 300 K



(b) 500 K

Uncertainty in Inter-Atomic Potential

Objectives

- Relative importance of potential field parameters using sensitivity analysis.
- Assess variability in thermal conductivity estimates due to perturbations in the potential field (Forward Problem).

$$\Phi(A, B, p, q, a, \lambda, \gamma) \mapsto k$$

- Robust calibration of the potential field parameters in a Bayesian setting.

Challenges

- Global sensitivity analysis and the forward problem are computationally intractable.
- Explore the applicability of a derivative-based sensitivity measure to reduce the dimensionality of the problem.

Derivative-based Global Sensitivity Measures

Motivation

- Sensitivity analysis based on Sobol indices is commonly used to determine relative importance of the parameters.
- Sobol sensitivity indices are compute intensive:

$$\mathcal{T}(\theta_i) = \frac{\mathbb{E}_{\boldsymbol{\theta} \sim i}[\mathbb{V}_{\theta_i}(\mathcal{G}|\boldsymbol{\theta}_{\sim i})]}{\mathbb{V}(\mathcal{G})}$$

- Bounds on Sobol indices can be computed easily using DGSM and are shown to converge at a much faster rate.

Derivative-based Global Sensitivity Measures

Background

- DGSM for Randomly distributed parameters [Sobol and Kucherenko, 2009]:

$$\mu_i = \mathbb{E}\left[\left(\frac{\partial G(\boldsymbol{x})}{\partial x_i}\right)^2\right]$$

where,

$$\frac{\partial G(\boldsymbol{x}^*)}{\partial x_i} = \lim_{\delta \to 0} \frac{[G(x_1^*, \ldots, x_{i-1}^*, x_i^* + \delta, x_{i+1}^*, \ldots, x_d^*) - G(\boldsymbol{x}^*)]}{\delta}$$

- Total number of model realizations required to compute μ_i using N samples is $N(d+1)$.

Derivative-based Global Sensitivity Measures

Background

- Upper bound on Sobol Total Effect index (ST_i) [Sobol and Kucherenko, 2009]:

$$ST_i \leq \frac{\mathcal{C}_i \mu_i}{V} \; (\propto \hat{\mathcal{C}_i \mu_i})$$

$$\hat{\mathcal{C}_i \mu_i} = \frac{\mathcal{C}_i \mu_i}{\sum_i \mathcal{C}_i \mu_i}$$

$\mathcal{C}$: Poincaré Constant V: Variance

- The Poincaré Constant is specific to a given probability distribution:

$\mathcal{U}[a,b]$	$(b-a)^2/\pi^2$
$\mathcal{N}(\mu,\sigma^2)$	σ^2

Algorithm: Parameter Screening

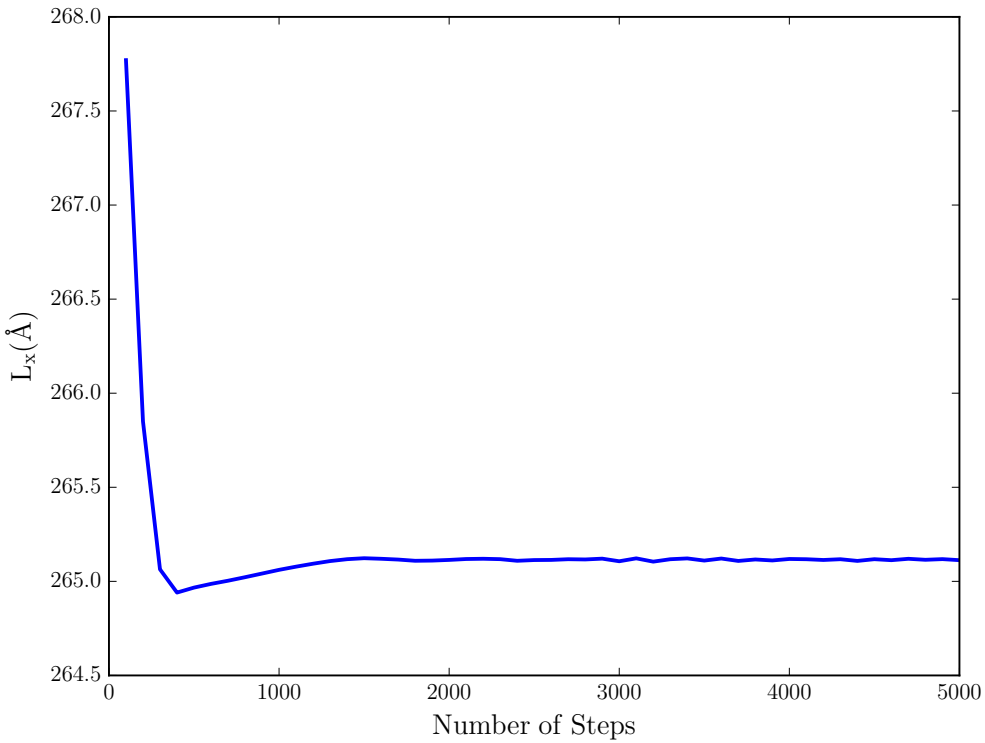
Generate n_1 points in $\mathbb{R}^d$;
Compute UB_i for parameters θ_i using n_1 points;
% $NF = n_1(d+1)$, NF: Number of model realizations;
Rank Parameters (θ_i) based on UB_i estimates $(\mathcal{R}^{old})$;
set $k = 1$ % k: Iteration counter;
repeat
    Generate βn_1 new points in $\mathbb{R}^d$ $(\beta n_1 \in \mathbb{Z})$;
    Compute UB_i^{new} using $(1+\beta k)n_1$ points;
    % $NF = (1+\beta k)n_1(d+1)$;
    Rank Parameters based on UB_i^{new} estimates $(\mathcal{R}^{new})$;
    if $(\mathcal{R}^{new} = \mathcal{R}^{old})$ **then**
        Compute: $r_i = \frac{UB_i^{new}}{max(UB_i^{new})}$;
        Construct a set $s = \{\theta_i \ni r_i < \alpha\}$;
        Exit the loop;
    end
    set $k = k + 1$;
until $\mathcal{R}^{new} \neq \mathcal{R}^{old}$;
Construct a validation set: $(\boldsymbol{\theta}_j, \mathcal{M}(\boldsymbol{\theta}_j))$, j=1,2,...,NF;

MD SIMULATION

NPT → NVT → NVE → NVE

[Relax System Length] [Equilibrate system to 300 K] [Equilibrate thermostats] [Generate Data]

N: Number of Atoms P: Pressure V: Volume T: Temperature E: Energy



PARAMETER SCREENING

$\mathcal{C}_i\hat{\mu}_i$: 10 Samples	$\mathcal{C}_i\hat{\mu}_i$: 15 Samples	$\mathcal{C}_i\hat{\mu}_i$: 20 Samples	$\mathcal{C}_i\hat{\mu}_i$: 25 Samples	$\mathcal{C}_i\hat{\mu}_i$: 30 Samples
γ: 0.2531	γ: 0.2001	γ: 0.1933	γ: 0.2075	γ: 0.2024
α: 0.2170	α: 0.1723	α: 0.1593	α: 0.1495	α: 0.1487
A: 0.2014	A: 0.1508	q: 0.1468	q: 0.1493	λ: 0.1424
B: 0.0964	B: 0.1381	λ: 0.1325	λ: 0.1392	q: 0.1404
p: 0.0916	p: 0.1253	A: 0.1261	A: 0.1217	A: 0.1395
q: 0.0741	q: 0.1232	B: 0.1242	B: 0.1175	B: 0.1142
λ: 0.0664	λ: 0.0900	p: 0.1177	p: 0.1153	p: 0.1125

PARAMETER SCREENING