Predicting Interaction Strength Of Potential Landscapes Using A Machine Learning Approach

Ahmad Haider[1]; Alan Liu[1]; Todd Sulchek[1]

[1] Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta

Motivation

- The Energy landscape (i.e. strength and range of interaction) fully describe basis of chemical and biological interactions[1,2]

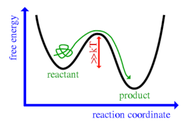
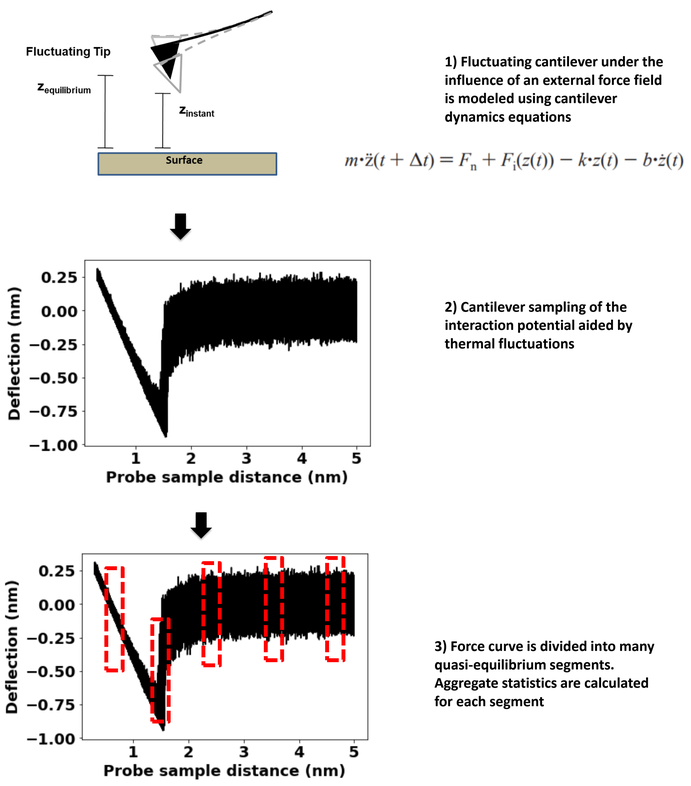


Fig. 1: Schematic of an Energy Landscape

- Energy landscapes are described by Boltzmann distributions
- Accurate construction of **energy landscapes** are dependent on sufficient Boltzmann sampling of all possible configurations of the system
- Current methods to compute energy landscapes are inaccurate because of insufficient sampling of portions of interactions with steep potential gradients or short reaction coordinates[3,4]
- AFM probes at fixed location or speed are tuned to accurately sample a certain range of potential energy gradients, but have low measurement accuracies for energies outside the range

GOAL: Classify energy landscape into regions of differential potential gradients so that probe speed can be dynamically adjusted to maximize overall sampling of the landscape

Methods and Materials – Data



1) Fluctuating cantilever under the influence of an external force field is modeled using cantilever dynamics equations

$$m \cdot \ddot{z}(t+\Delta t) = F_n + F_i(z(t)) - k \cdot z(t) - b \cdot \dot{z}(t)$$

2) Cantilever sampling of the interaction potential aided by thermal fluctuations

3) Force curve is divided into many quasi-equilibrium segments. Aggregate statistics are calculated for each segment

Fig. 2: Data generation and analysis approach

Methods and Materials - Process

- Aggregate statistics computed for each segment are
  - Max and min deflection
  - Mean, median, mode
  - Standard deviation, skew, kurtosis
  - Number of peaks, distance between peaks
  - Absolute energy, entropy
  - Gradient and intercepts
  - Percent of data points above and below mean
- Unsupervised ML algorithm (K-Means clustering) is applied to aggregate statistics data to categorize each segment into one of n = 5 clusters
- Cantilever (k=1000 pN/nm, f_0=25kHz, Q =3) is used for generating simulation force data and data is collected at 50kHz data sampling rate.
- 6-3 chemical interaction force is used to simulate external force field
- Simulation generated 1M data points divided into 1000 equal segments. 50 bins are allocated to each segment

Results: Classification Analysis

- Based on cluster analysis, 5 different classes are discovered which strongly correlate with the strength of interaction and direction of probe's motion
- The classes correspond to probe location and interaction strength as below

Probe location	Interaction strength	Color
Far from surface	Weak/None	Red
Approaching surface	Medium	Yellow
Near contact	Strong	Green
Contact	Medium	Blue
After contact	Weak/None	Purple

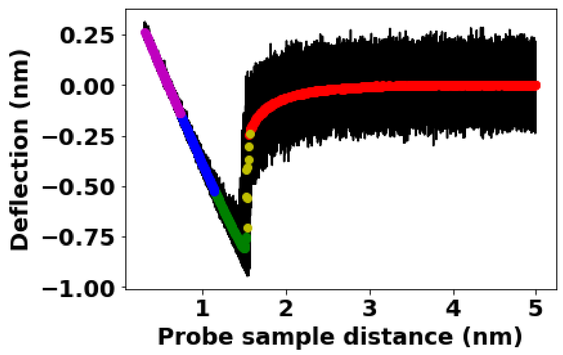


Fig. 3: AFM force curve (black) obtained through cantilever dynamics simulations when overlaid with machine learning based predictions of 5 classes of varying interaction strengths

Results – Cluster Optimization

- An optimal cluster count is determined by minimizing total sum of squared errors (SSE) or distortion for the data points against the cluster centroids

$$distortion = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_j^{(i)} - \mu_{c(i),j}\right)^2$$

- A sharp decrease in distortion across consecutive cluster count indicates that an optimum is reached
- Plot below shows 2 optimal points corresponding to 2 clusters and 5 clusters.
- Two clusters are not relevant for analysis since they indicate only general trends, such as interaction potential can be classified as far away and contact
- Five clusters captures the complete probe behavior as it samples over the entire energy landscape

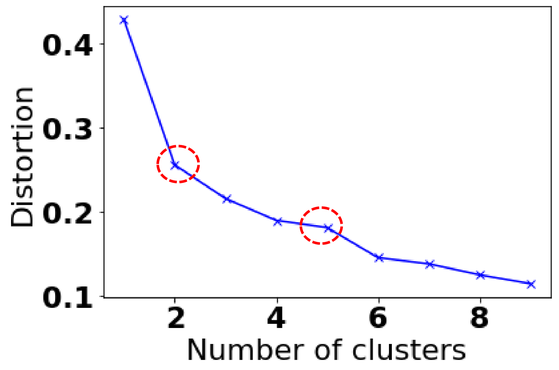


Fig. 4: Choosing the optimal number of clusters based on minimization of SSE

Conclusions

Boltzmann analysis calculates the potential energy landscapes which is dependent on perfect sampling of the landscape

To account for insufficient sampling time in regions of sharp potential gradients, it is advisable to adjust probe speed to match the gradient strength

Based on characteristics of deflection data, machine learning approaches can identify regions of energy landscapes with differential gradients, thereby allowing for adjustment of probe speeds for varying energy gradients

Flexibility of machine learning allows us to rapidly experiment with tunable parameters to identify best solution space for interaction sampling

Contact

Ahmad Haider
Georgia Institute of Technology
Email: ahmadhaider@ieee.org
Phone:404-368-4134

References

[1] Ashby P.D., Lieber C. L., *J Am. Chem. Soc.*, 126:16973-16980, 2004

[2] Willemsen, O.H., Kuipers , L. et al., *J Langmuir*, 16:4339-4347, 2000

[3] Haider, A., Potter. D., Sulchek T.A., *Proc. Nat. Acad. Sc.,* 113(50): 14213-14218, 2016

[4] Haider, A., Potter. D., Sulchek T.A., *J. Phys. Chem. C*, *121* (13), 7248–7258, 2017