分子シミュレーションと AIを組み合わせた マテリアルデザイン

慶應義塾大学 理工学部
泰岡顕治
yasuoka@mech.keio.ac.jp

Molecular Dynamics (MD) Simulation

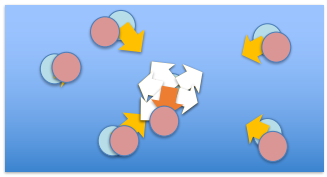
Simulating the motion of molecules by using computers, analyzing various phenomenon.

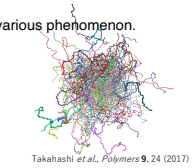
Newton's equation of motion $m\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} = \boldsymbol{F}$

position $\boldsymbol{r}(t)$, velocity $\boldsymbol{v}(t)$

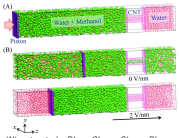
↓ force calculation $\boldsymbol{F}(t)$

↓ new pos. $\boldsymbol{r}(t+\Delta t)$, new vel. $\boldsymbol{v}(t+\Delta t)$





Takahashi *et al.*, *Polymers* **9**, 24 (2017)



Winarto et al., *Phys. Chem. Chem. Phys.*, **21**, 15431 (2019).

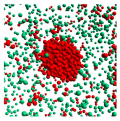
Contents

Machine learning for Molecular Dynamics simulations

- Acceleration of Molecular Dynamics simulations (Supervised Learning)
- Analysis of Molecular Dynamics data using machine learning (Unsupervised Learning)

Acceleration of MD Simulations

- **Problem of MD simulations:**
  The high computational costs of **large scale**, **long time** MD simulations
- **Acceleration of large scale（10^6～10^9 particles）MD simulations**
  → Divide the space and apply the massively parallel computing



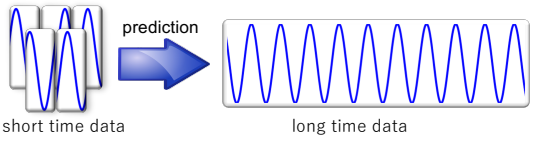
・Large scale simulation of droplet nucleation
Although the droplet nucleation is rare event, a lot of droplet can be observed in large scale simulation.
Simulation results showed good agreements with theory.
Ayuba *et al.*, *J. Chem. Phys.*, **149**, 044504 (2018)

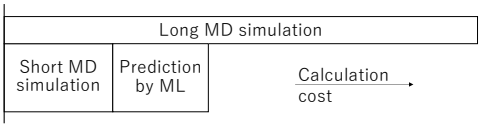
- **Acceleration of long time MD simulations**
  Time evolution cannot be separated and parallelized.
  → Multiple time step method (= still need long computational time)

A novel method for the acceleration of long time simulations is required.

Acceleration of MD Simulations by ML

- **Generation of long time data using short time data by ML**



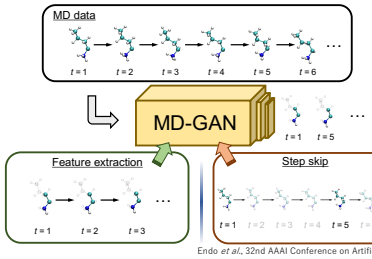


Endo *et al.*, 32nd AAAI Conference on Artificial Intelligence, 2192 (2018).

Multi-step Time Series Generator for MD

Efficient generation of MD time series data using **GANs***
→ Consider **the time evolution of a part of the whole system**
・Extraction of feature $\boldsymbol{y}$ from whole system $\boldsymbol{x}$
・Time step for $\boldsymbol{y}$ is larger than that for original MD (= step skip)

* GANs: Generative adversarial nets



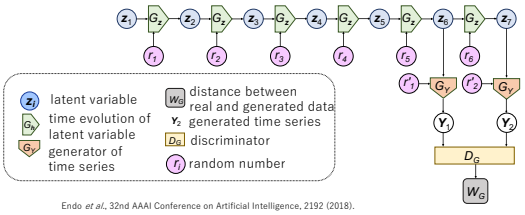
Endo *et al.*, 32nd AAAI Conference on Artificial Intelligence, 2192 (2018).

Architecture of MD-GAN

MD-GAN is based on Wasserstein GAN.

Time evolution of the part of system = **stochastic evolution**.
When we perform the stochastic evolution iteratively, we encounter the **exposure bias**.

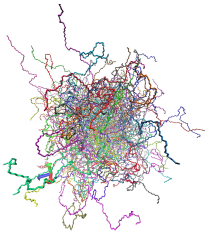
To mitigate the exposure bias, MD-GAN has two features.
- Indirect generation：time evolution of low-dimensional latent variable $\boldsymbol{z}$
- Stabilization of distribution：acquirement of appropriate structure for $\boldsymbol{z}$

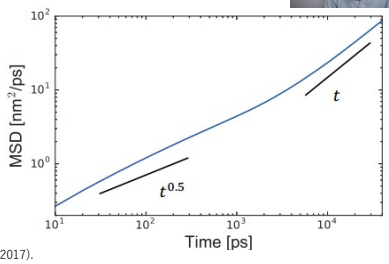


Endo *et al.*, 32nd AAAI Conference on Artificial Intelligence, 2192 (2018).

MSD of Polyethylene Polymer Chain

- MD simulation of 300 polyethylene polymer chains
  (M = 1405 g/mol, 500 K, TraPPE-UA force fields)
- Slope of MSD for the chain center anomaly changes[4]

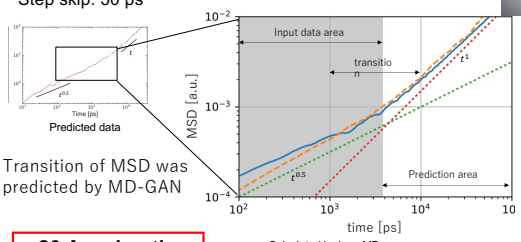




Takahashi *et al.*, *Polymers* **9**, 24 (2017).

MSD of Polyethylene Polymer Chain

Input for MD-GAN: 3840 ps of time series = $t^{0.5}$ area
Feature extraction: chain center positions
Step skip: 30 ps



Transition of MSD was predicted by MD-GAN

x20 Acceleration

Calculated by long MD
Prediction by MD-GAN

Endo *et al.*, 32nd AAAI Conference on Artificial Intelligence, 2192 (2018).

Contents

Machine learning for Molecular Dynamics simulations

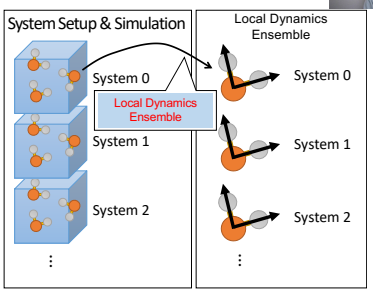
- Acceleration of Molecular Dynamics simulations
  (Supervised Learning)
- Analysis of Molecular Dynamics data using machine learning
  (Unsupervised Learning)

Local Dynamics Ensemble

- Solution for input → Local dynamics ensemble (LDE)

Ensemble:
Probability distribution over all possible states of the system.

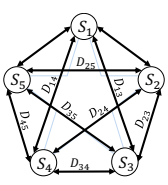
LDE:
Probability distribution of short trajectories for one molecule

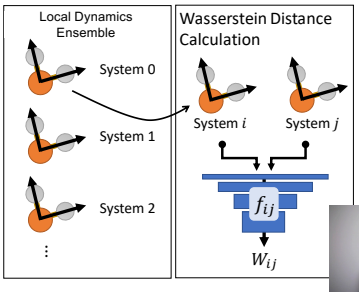


Endo et al., *Nanoscale*, **11**, 10064-10071 (2019).

Statistical distance between LDEs

- Comparison of systems using the statistical distance between LDEs.
- Wasserstein distance is easily calculated by DNNs.

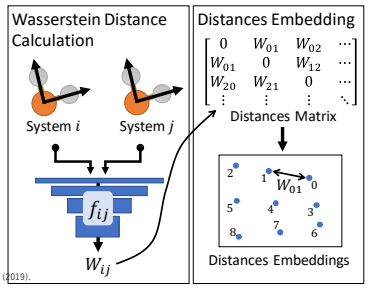




Endo et al., *Nanoscale*, **11**, 10064-10071 (2019).

Embeddings of LDEs

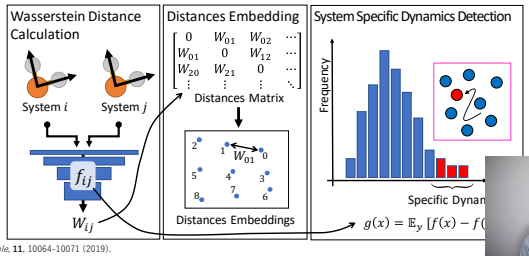
- To clearly illustrate the differences between systems, we obtain low-dimensional embeddings using the calculated distances.



Endo et al., *Nanoscale*, **11**, 10064-10071 (2019).

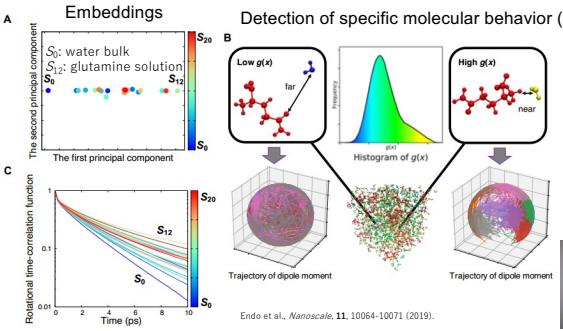
Molecular behavior contributing to distances

- As a result of trained function, the output (Wasserstein distance) is directly connected to the input (MD data).
- The molecular behavior that characterizes systems is detected using the trained function



Endo et al., *Nanoscale*, **11**, 10064-10071 (2019).

Amino acid solution (20 amino acid solution + 1 water bulk)
→ LDE: 64 steps of displacement for dipole moment



Endo et al., *Nanoscale*, **11**, 10064-10071 (2019).

Summary

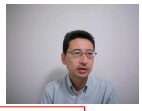
Research of MD using machine learning

・Acceleration of MD simulations
Using MD-GAN, physical properties of molecular systems can be predicted.
MD-GAN is powerful tool for the prediction of long time correlation.

・Analysis of MD data
Using Wasserstein distance calculations, we successfully detected the representative molecular behavior.
It is potentially useful for the investigation of new order parameters.

Future goal

Fast and automatic analysis tool for molecular dynamics data.
→ Development of material research



K. Endo, K. Tomobe, K. Yasuoka, 32nd AAAI Conference on Artificial Intelligence, 2192 (2018).
https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16477

K. Endo, D. Yuhara, K. Tomobe, K. Yasuoka, *Nanoscale*, **11**, 10064-10071 (2019).