Supporting Information on

Prediction of water transport properties on the anisotropic wetting surface via deep learning

Yuting Guo[a#], Haiyi Sun[a#], Meng An[b], Takuya Mabuchi[a,c], Yinbo Zhao[a], Gaoyang Li[a*]

[a] Institute of Fluid Science, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

[b] Department of Mechanical Engineering, Graduate School of Engineering, The University of Tokyo, 7-3-1, Bunkyo-ku, Tokyo113-8656, Japan

[c] Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, 2-1-1 Katahira Aoba-ku, Sendai, Miyagi 980-8577, Japan

* Corresponding author: Gaoyang Li

# These authors contributed equally to this work.

E-mail: li.gaoyang.e4@tohoku.ac.jp

S1. Generation of datasets and proposal of deep learning network

The generation of datasets and the proposal of network were the technical elements of deep learning research. As mentioned above, this study was devoted to the application of deep learning modeling multi-body system to realize the mapping of Pt/graphene substrate structure to water molecule transport landscape. Therefore, 1) point cloud (a collection of points in a specified space), as a flexible data format, was used to characterize the collection of nanoparticles in 3D Pt/graphene substrate, 2) two types of point clouds were extracted to represent the properties of Pt/graphene substrate and water molecules respectively and further construct datasets, and 3) a deep learning network with dual input and sampling channels was proposed.

LAMMPS software could directly mesh the MD simulation results and output high-resolution points, which was the geometric center points of each mesh. For all the cases, the orthogonal Cartesian grid method was used with mesh size 1 Å and the extracted mesh data was from a block with a size of 200 Å×200 Å×50 Å, corresponding to the evaluation region. The related MD insufficient sampling problem for velocity field and the proposal of nanoparticle tracking optimization strategy was detailed in Chapters 3.2 and 3.3. We constructed four deep learning datasets (water molecule adsorption or flow under uniform or random distribution of Pt). Each sample in each dataset (all points extracted from the MD simulation result of one case was called one sample) is composed of two types of point clouds, as shown in Figure 2. The solid point cloud characterizing the Pt/graphene structure could be expressed as $\{N_1 * P_1\}$, where N_1 represented the total number of points in the solid point cloud in a sample, and P_1 represented the properties of the solid. The liquid point cloud that characterizes adsorption or flow could be expressed as $\{N_2 * P_2\}$, where N_2 represented the total number of points in the liquid point cloud of water molecules in a sample, and P_2 represented the transportation properties of water molecules. The universality and scalability of the proposed deep learning algorithm mainly relied on the multidimensional attribute input of P_1 and P_2, see Chapter 3.5 for

further explanation. Here, ^{P_1} mainly included the spatial coordinates of each point that could represent the structure of the Pt/graphene structure, and molecular type labels (a Pt atom was assigned to a value of 0, and a graphene atom was assigned to a value of 1), and ^{P_2} mainly included the spatial coordinates of each mesh and transport properties of water molecules (density or velocity components in three directions). For the four datasets, each dataset is randomly divided into a training set and a test set at a ratio of 9:1 as shown in Table1.

S2. Design and advantages of our network

For the design and advantages of our network (including fully open source code), such as dual channel vs. single channel, fully connected layer shared weight vs. non-shared weight, the impact on prediction errors has been explained in detail by using control experiments in our previous work[1,2]. Here, we focused on the feature extraction and processing of point cloud by network. The network is shown in Figure 3 with two input and sampling channels to receive and process solid point cloud and liquid point cloud, respectively. For the sampling module, in order to enhance the correlation between the two types of point clouds and improve the network prediction performance, the first two feedforward fully connected layers (FC_1 and FC_2) shared the weight, that was, they shared the same underlying feature extraction method. The two types of point clouds entered two independent feedforward full connection layers (FC_3 and FC_4) after FC_2, which were used to extract the structural features of catalyst point cloud and the transportation features of liquid point cloud. After the processing of sampling module, the structural features and transportation features were abstracted into $\{N_1 * 512\}$ and $\{N_2 * 128\}$ dimensional vectors respectively and integrated into an $\{N_3 * 640\,(512 + 128)\}$ dimensional vector in the feature stitching module, followed by the output module, whose function was equivalent to the convolutional neural network decoding operation [3]. The $\{N_3 * 640\}$ dimensional vector could be decoded into $\{N_2 * P_2\}$ water molecule transport behavior.

For other details of the network, we added a Maxpooling layer as a symmetric function in

the sampling module, which could aid in resolving the disorder of the input point cloud [2]. The mean absolute error (MSE) was selected as the loss function. Adam was selected as the optimizer with hyper-parameters: learning rate=0.001, ε=0.001, ρ1=0.9, ρ2=0.999, and δ=1E−8 [4].

S3. Network training and testing process

The deep learning network was trained in the TensorFlow (2.0.0rc, Python3.6) environment on a Nvidia GeForce GTX 1080 Ti GPU. Batch size was set at 1 and epoch was set at 1000. The four datasets need to be trained separately using independent networks. During the training process, we saved the weight configuration with the minimum loss function value, and there were four trained networks for testing and performance evaluation. In the test process, the adsorption or flow properties of water molecules in P_2 could be predicted by inputting the structural attributes of the Pt/graphene substrate in P_1 and only the spatial coordinate information in P_2 into the trained networks.

S4. Discussion on the system size effect

The dependence of the static and dynamical properties on the system size has been studied widely. Patrick et al. indicated that the system become more locally organized for the small system size[5]. Yeh et al. indicated that for the water molecules and a Lennard-Jones (LJ) fluid, the diffusion coefficients increase as the system size increases, while the shear viscosities show no significant system-size dependences[6,7]. For thermal conductivity, minimum system size should be comparable to the largest mean-free paths[8]. In the present study, the system was in x and y periodic boundary conditions in order to minimize the finite-size effects. The size of system was much larger than that the minimum possible sizes that used in our previous study with the same water model[9]. Considering the correlation between velocity distribution and diffusivity coefficients, the possibility of system size dependence of the velocity distribution was high. Nevertheless, as we indicated that velocities of individual atoms are subject to error

and thermal fluctuations, the increase in the box volume could not completely exclude the possibility of a finite size effect on velocity distribution. The effect of the size of the present system on the velocity field in chaotic systems needs to be further elucidated in future work.

Reference

1 G. Li, H. Wang, M. Zhang, S. Tupin, A. Qiao, Y. Liu, M. Ohta and H. Anzai, *Communications Biology*, 2021, **4**, 1–12.

2 C. R. Qi, H. Su, K. Mo and L. J. Guibas, in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 77–85.

3 X. Guo, W. Li and F. Iorio, in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, vol. 13-17-Augu, pp. 481–490.

4 D. P. Kingma and J. L. Ba, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

5 P. Ganster, M. Benoit, W. Kob and J.-M. Delaye, *The Journal of Chemical Physics*, 2004, **120**, 10172–10181.

6 I.-C. Yeh and G. Hummer, *J. Phys. Chem. B*, 2004, **108**, 15873–15879.

7 S. H. Jamali, L. Wolff, T. M. Becker, A. Bardow, T. J. H. Vlugt and O. A. Moultos, *J. Chem. Theory Comput.*, 2018, **14**, 2667–2677.

8 D. P. Sellan, E. S. Landry, J. E. Turney, A. J. H. McGaughey and C. H. Amon, *Phys. Rev. B*, 2010, **81**, 214305.

9 T. Mabuchi and T. Tokumasu, *Journal of Polymer Science Part B: Polymer Physics*, 2019, **57**, 867–878.