Modelling of Cellulose Materials Using Graph-Based Interatomic Potentials

Aneta D. Niklas[1], Volker L. Deringer[1]
[1]Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford, UK

aneta.niklas@sjc.ox.ac.uk

Biomass-derived carbon materials are emerging as sustainable alternatives for energy storage, catalysis, and carbon capture [1]. Cellulose, an abundant polymer with high carbon and oxygen content, is a promising precursor for the synthesis of functional carbon materials [2].

Despite extensive experimental studies, the transformation from amorphous biomass to graphitic carbon remains poorly understood, as experiments provide structural insight but limited access to the atomistic dynamics. Computational approaches offer complementary understanding, but density functional theory is restricted to small systems and short timescales [3], while classical force fields such as ReaxFF often struggle to describe reactive CHO chemistry accurately [4]. Machine-learned interatomic potentials (MLIPs) provide a route to overcome these limitations, combining near DFT accuracy with the efficiency required for large-scale molecular dynamics (MD) simulations. Graph neural network (GNN) based potentials have shown strong potential for modelling complex reactive CHO systems [5].

Here we compare MLIP models obtained either by direct training or by fine-tuning the pre-trained MACE-MH-1 model [6] for cellulose carbonisation and graphitisation. We evaluate their performance across chemically diverse CHO environments relevant to carbonisation, from cellulose and its decomposition products to amorphous carbon structures, using energy and force errors. We examine how dataset composition affects performance and find that models trained directly on mixed datasets show strong dependence on the data, with large errors for amorphous carbon configurations despite good performance on more ordered structures. In contrast, fine-tuning pre-trained models improves robustness and reduces errors across all domains.

We further find that model performance depends strongly on the choice of pre-trained head used for fine-tuning. Fine-tuning from different heads of the same base model leads to clear differences in accuracy, with some yielding lower energy and force errors and showing less sensitivity to dataset composition. Together, these results show that consistent performance across the different stages of cellulose carbonisation depends on both the choice of training data and the pre-trained head used for fine-tuning.

Using MLIP-driven MD simulations, we plan to explore the ability of these models to capture molecular decomposition and the emergence of extended carbon structures across temperatures and pressures relevant to carbonisation.

References

[1] S. Park, J. Song, W. C. Lee, S. Jang, J. Lee, J. Kim, H. K. Kim, K. Min, Chemical Engineering Journal, 470 (2023), 144234
[2] Z. Chen, H. Zheng, J. Yi, T. Liu, H. Lai, S. Zhang, W. Huang, Y. Yin, X. Huang, Y. Tong, D. Liang, R. Li, L. Zhong, C. Zhang, H. Zhang, Resources Chemicals and Materials, 4 (2025), 100120
[3] Q. Wang, H. Song, S. Pan, N. Dong, X. Wang, S. Sun, Scientific Reports, 10 (2020), 3626
[4] Y. Zhang, M. S. Ahmad, W. Zhou, Y. Li, R. Cao, B. Shen, Chemical Engineering Journal, 533 (2026), 174660
[5] P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. V. Hoesel, H. Schopmans, T. Sommer, P. Friederich, Communications Materials, 3 (2022), 93
[6] I. Batatia, C. Lin, J. Hart, E. Kasoar, A. M. Elena, S. Walton Norwood, T. Wolf, G. Csanyi, “Cross Learning between Electronic Structure Theories for Unifying Molecular, Surface, and Inorganic Crystal Foundation Force Fields”, arXiv preprint arXiv:2510.25380, 2025