Identification of Optimal Metal-Organic Frameworks by Machine Learning: Structure Decomposition, Feature Integration, and Predictive Modeling

Zihao Wang[a], Yageng Zhou[b], Teng Zhou[a,b,*], Kai Sundmacher[a,b]

[a] Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany

[b] Process Systems Engineering, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany

[*] Corresponding Author: zhout@mpi-magdeburg.mpg.de (Teng Zhou)

Supporting Information

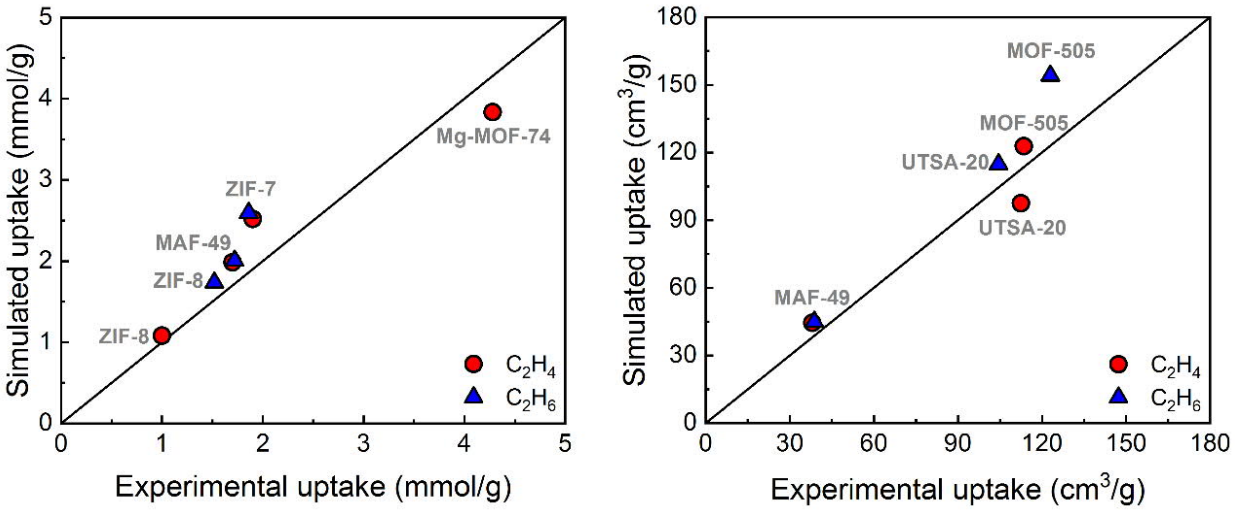

Figure S1. Comparison between experimental and GCMC simulated single-component C_2H_4 and C_2H_6 uptakes at 1 bar (296 K for MOF-505 [1] and UTSA-20 [1], 298 K for Mg-MOF-74 [2] and ZIF-7 [3], 303 K for ZIF-8 [4], and 316 K for MAF-49 [5]).

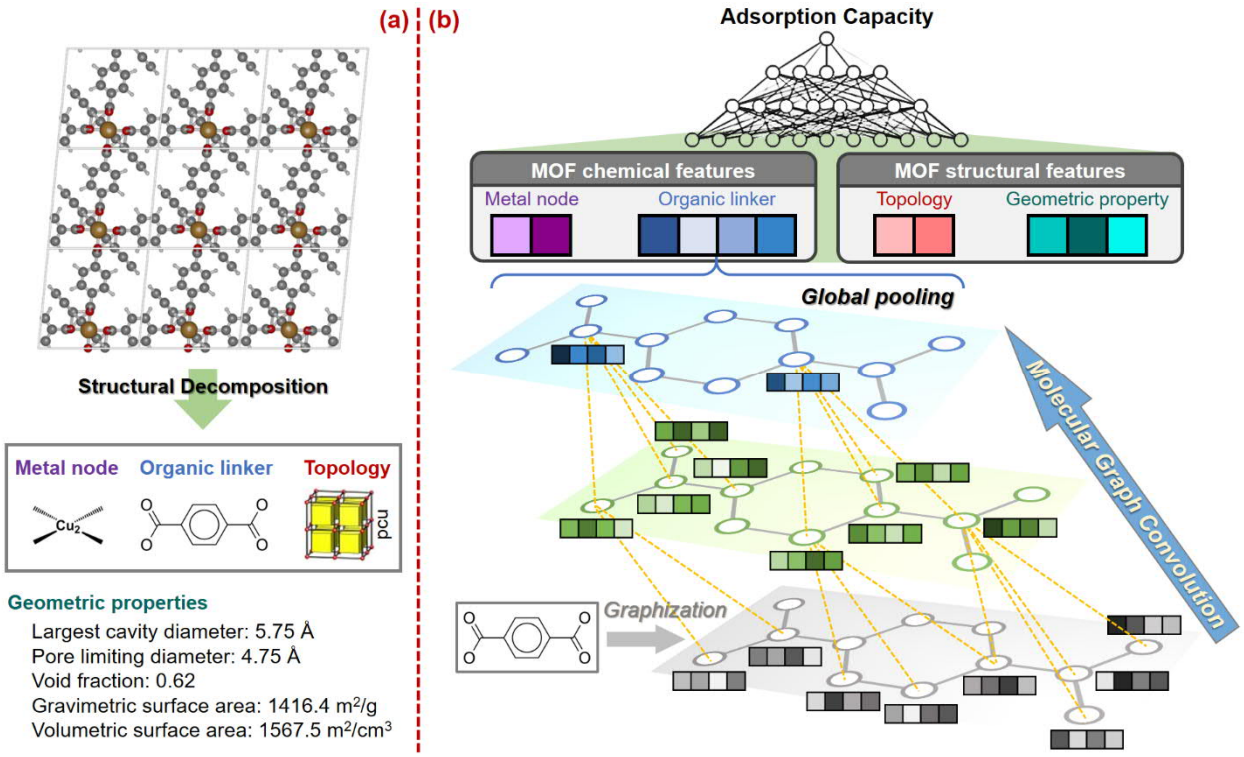

Figure S2. An example of MOF structure decomposition, feature integration, and predictive modeling.

- Atoms in organic linker, metal node, and topology are separately encoded by the word embedding;
- We did not pre-assign any physical features for each atom such as atom weight, aromaticity, etc. Instead, after distinguishing atoms by their atomic and bond types, we first assign a fixed-length vector of arbitrary features for each type of atom, and these features are later updated or optimized during model training to minimize ML prediction error;
- Three graph convolution layers are used to update atomic features;
- The global pooling strategy, Set2Set in PyTorch Geometric, is used to obtain the feature of the entire organic linker;
- The embedded metal node, embedded topology, organic linker features obtained from the global pooling, and geometric properties are concatenated into a one-dimensional vector for each MOF;
- The concatenated vector is used as MOF input feature for the final prediction of adsorption capacities by an FNN with three hidden layers.

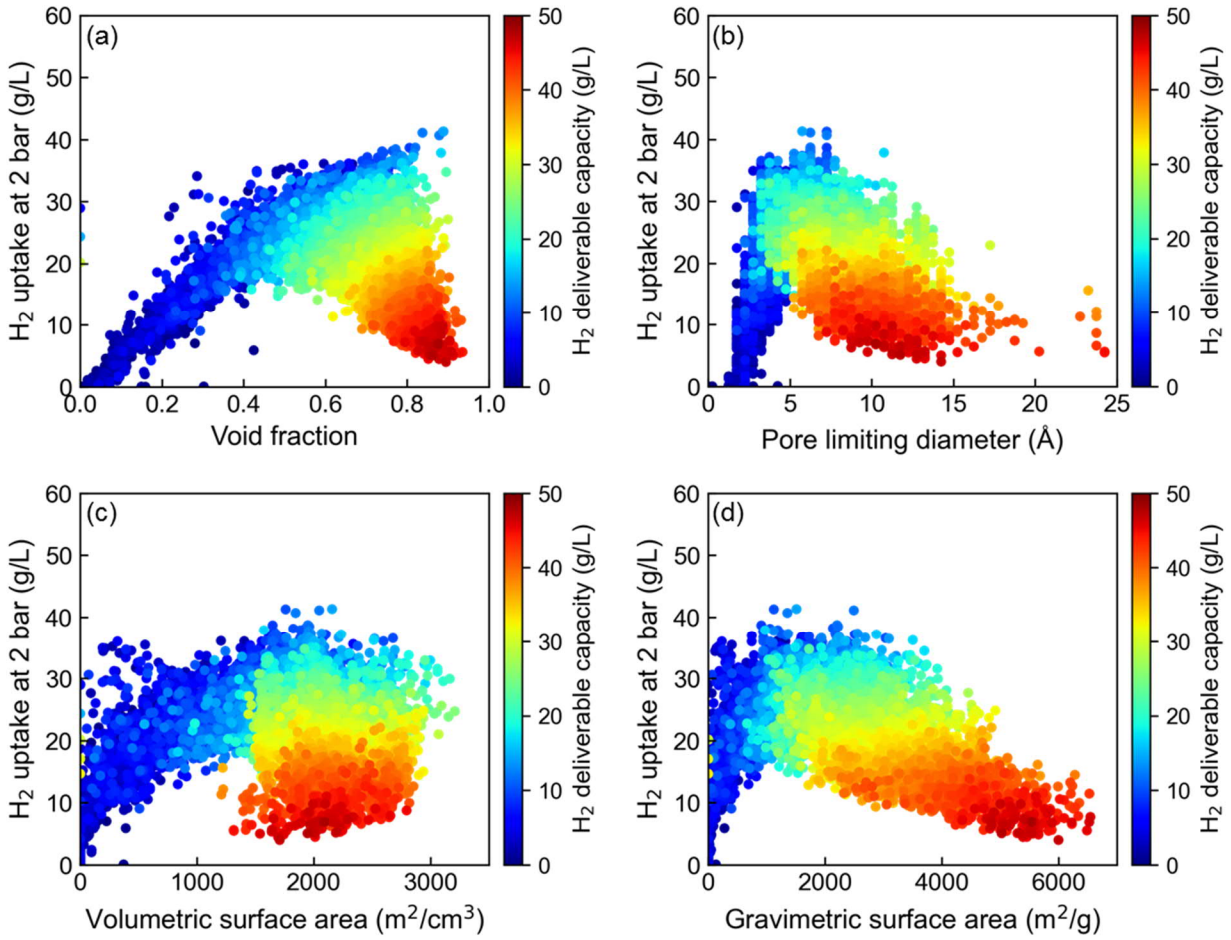

Figure S3. Relations between MOF geometric properties and the H_2 uptake at 2 bar and 77K: (a) void fraction, (b) pore limiting diameter, (c) volumetric surface area, and (d) gravimetric surface area. Data points are colored by the H_2 deliverable capacity.

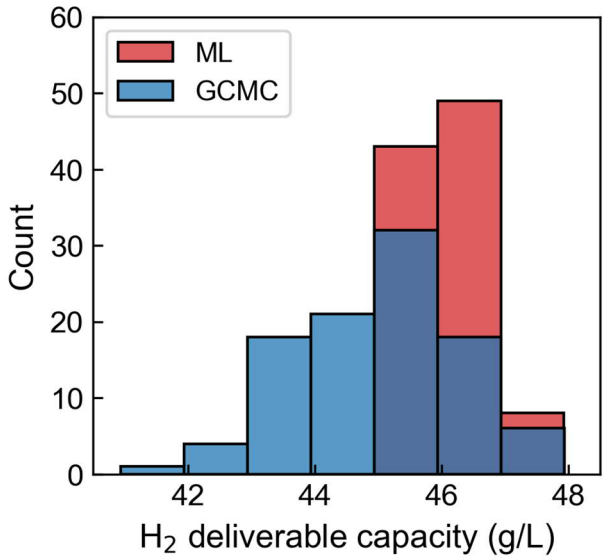

Figure S4. Comparison between ML predictions and GCMC simulations for the top 100 MOFs with the highest H_2 deliverable capacity identified by the ML-assisted large-scale screening.

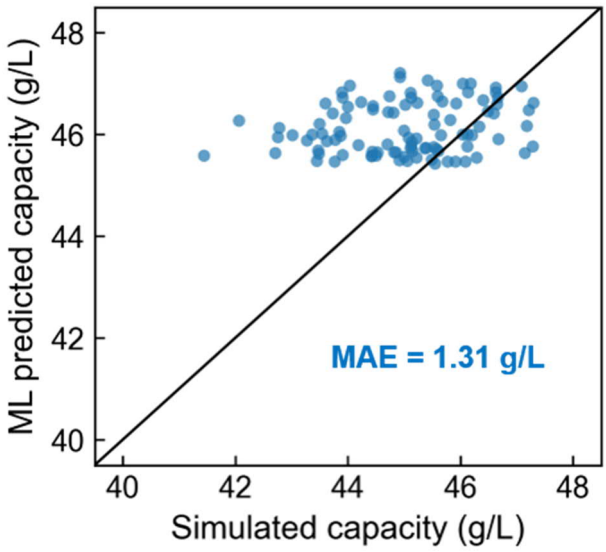

Figure S5. Comparison of GCMC-based and ML-predicted H_2 deliverable capacity for the top 100 MOFs.

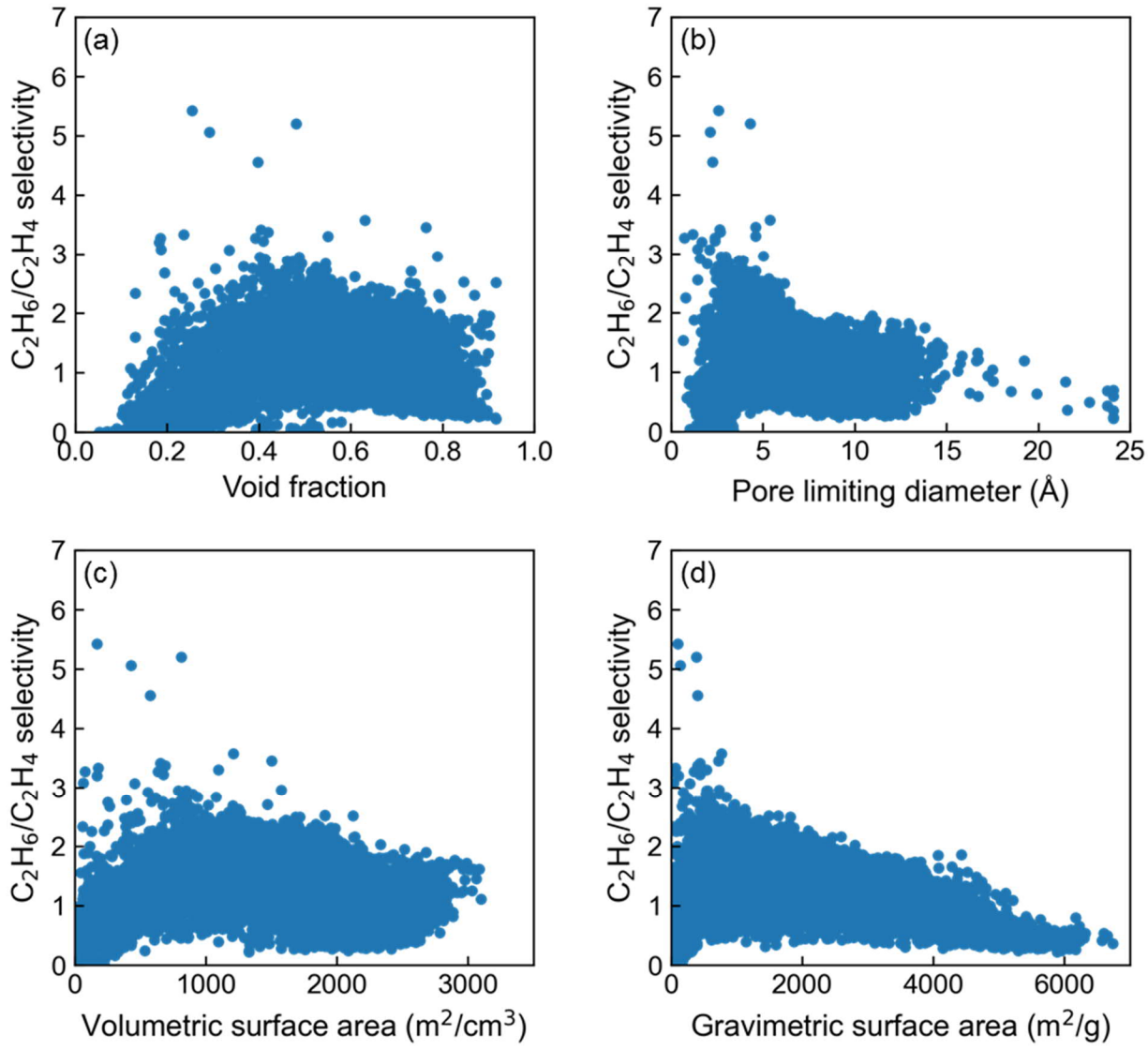

Figure S6. Relations between MOF geometric properties and the C_2H_6/C_2H_4 selectivity at 1 bar and 298 K: (a) void fraction, (b) pore limiting diameter, (c) volumetric surface area, and (d) gravimetric surface area.

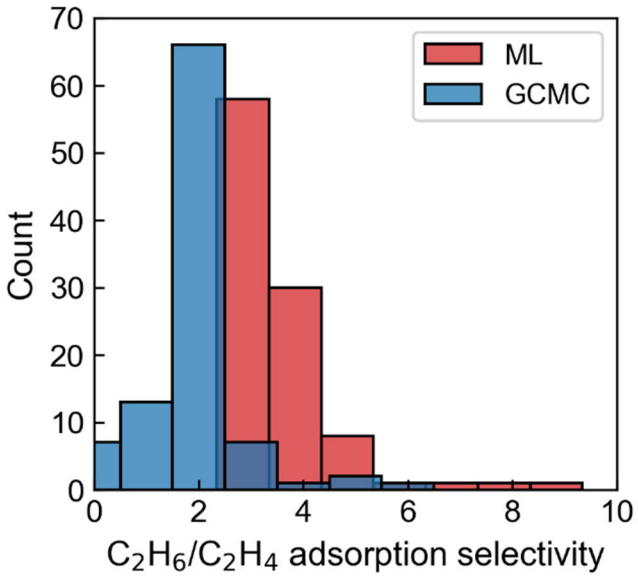

Figure S7. Comparison between ML predictions and GCMC simulations for top 100 MOFs with the highest C_2H_6/C_2H_4 selectivity identified by the ML-assisted large-scale screening.

Table S1. Lennard-Jones parameters for framework atoms.

Atom	C	O	H	N	F	Cl	Br	Zn	Cu	V	Zr
σ (Å)	3.47	3.03	2.85	3.26	3.09	3.52	3.52	2.46	3.11	2.80	2.78
ε/k_B (K)	47.86	48.16	7.65	38.95	36.48	142.56	186.19	62.40	2.52	8.05	34.72

Table S2. Hyper-parameters considered for the ML model configuration.

Hyper-parameter	Setting
Hidden layer width	8, 16, 24, 32
Activation function	Tanh, ELU, ReLU, Sigmoid, Softplus
Batch size	64, 128, 256
Graph convolution method [6]	GINConv, GCNConv, AGNNConv, ClusterGCNConv, GATConv, GraphConv, LEConv, MFConv, SAGEConv

Table S3. Optimal hyper-parameter combinations for ML models.

Case	Prediction target	Optimal hyper-parameters
Case 1 – H_2 storage	H_2 uptake at 100 bar/77 K	32, Sigmoid, 64, GCNConv
	H_2 uptake at 2 bar/77 K	16, Sigmoid, 256, ClusterGCNConv
Case 2 – C_2H_4/C_2H_6 separation	C_2H_4 uptake at 1 bar/298 K	16, Tanh, 256, GINConv
	C_2H_6 uptake at 1 bar/298 K	16, ELU, 256, GINConv

Table S4. Model performance in the prediction of H_2 uptakes.

Prediction target	Model type	Dataset	MAE (g/L)	R^2
H_2 uptake at 100 bar/77 K	w/ chemical features	Training	1.04	0.984
		Validation	1.24	0.975
		Test	1.25	0.976
	w/o chemical features	Training	1.35	0.972
		Validation	1.41	0.968
		Test	1.33	0.974
H_2 uptake at 2 bar/77 K	w/ chemical features	Training	1.03	0.961
		Validation	1.29	0.927
		Test	1.29	0.928
	w/o chemical features	Training	1.88	0.873
		Validation	1.98	0.852
		Test	2.02	0.842

Table S5. Optimal hyper-parameter combinations for ML models without chemical features.

Case	Prediction target	Optimal hyper-parameters*
Case 1 – H_2 storage	H_2 uptake at 100 bar/77 K	16, Softplus, 64, -
	H_2 uptake at 2 bar/77 K	16, Tanh, 128, -
Case 2 – C_2H_4/C_2H_6 separation	C_2H_4 uptake at 1 bar/298 K	32, Softplus, 64, -
	C_2H_6 uptake at 1 bar/298 K	24, Sigmoid, 128, -

* The hyper-parameter "graph convolutional layer" is not applied.

Table S6. Model performance in the prediction of C_2H_4 and C_2H_6 uptakes.

Prediction target	Model type	Dataset	MAE (cm^3/g)	R^2
C_2H_4 uptake at 1 bar/298 K	w/ chemical features	Training	5.00	0.926
		Validation	6.03	0.887
		Test	5.79	0.896
	w/o chemical features	Training	8.63	0.776
		Validation	9.33	0.739
		Test	9.00	0.742
C_2H_6 uptake at 1 bar/298 K	w/ chemical features	Training	0.62	0.942
		Validation	0.80	0.899
		Test	0.77	0.897
	w/o chemical features	Training	1.12	0.796
		Validation	1.21	0.768
		Test	1.17	0.772

References

[1] He, Y., Krishna, R., Chen, B., 2012. Metal-organic frameworks with potential for energy-efficient adsorptive separation of light hydrocarbons. Energy Environ. Sci. 5 (10), 9107-9120.

[2] Liao, Y., Zhang, L., Weston, M.H., Morris, W., Hupp, J.T., Farha, O.K., 2017. Tuning ethylene gas adsorption via metal node modulation: Cu-MOF-74 for a high ethylene deliverable capacity. Chem. Commun. 53 (67), 9376-9379.

[3] Gucuyener, C., Van Den Bergh, J., Gascon, J., Kapteijn, F., 2010. Ethane/ethene separation turned on its head: selective ethane adsorption on the metal-organic Framework ZIF-7 through a gate-opening mechanism. J. Am. Chem. Soc. 132 (50), 17704-17706.

[4] Mondal, S.S., Hovestadt, M., Dey, S., Paula, C., Glomb, S., Kelling, A., Schilde, U., Janiak, C., Hartmann, M., Holdt, H.J., 2017. Synthesis of a partially fluorinated ZIF-8 analog for ethane/ethene separation. CrystEngComm 19 (39), 5882-5891.

[5] Liao, P., Zhang, W., Zhang, J., Chen, X., 2015. Efficient purification of ethene by an ethane-trapping metal-organic framework. Nat. Commun. 6 (1), 1-9.

[6] PyTorch Geometric. https://pytorch-geometric.readthedocs.io/