
Hierarchical Representation Learning for Metal-Organic Frameworks

Michał Sanocki

Multiscale Modeling of Fluid Materials
TUM School of Engineering and Design
Technical University of Munich, Germany
michal.sanocki@tum.de

Julija Zavadlav

Multiscale Modeling of Fluid Materials
TUM School of Engineering and Design
Technical University of Munich, Germany
julija.zavadlav@tum.de

Abstract

Metal-organic framework (MOF) property prediction is challenging because labeled datasets remain sparse relative to the vast chemical design space, despite substantial reuse of chemically similar building blocks across structures. We introduce a three-stage framework comprising structural segmentation (MOFSEG), representation learning (MOFENC), and property modeling (MOFMOD). MOFMOD is a hierarchical multiresolution graph neural network that represents MOFs simultaneously through atoms, linker motifs, organic linkers, and inorganic nodes. It combines periodic geometry with component representations pretrained by MOFENC and exchanges information across chemically defined resolutions. This design enables representations learned from recurring MOF components to be reused across distinct crystal structures while retaining their local and periodic environments. MOFSEG provides the structural hierarchy required by the model and successfully processes 93.3% of 159,289 structures across three MOF datasets WS24, hMOF, and QMOF. On QMOF band-gap prediction, MOFMOD achieves an MAE of 0.211 eV, compared with 0.321 eV when pretrained component representations are replaced by trainable embeddings. On hMOF CO₂ adsorption, the same architecture achieves an MAE of 0.375 mol kg⁻¹ on the full processed dataset. These results demonstrate the potential of hierarchical component-level representation learning for MOF property prediction and provide a framework for transferring chemical information across the MOF design space.

1 Introduction

Metal-organic frameworks (MOFs) are porous crystals assembled from inorganic nodes and organic linkers Yuan et al. [2022], Burner et al. [2020]. Their modular chemistry enables broad control of topology, porosity, and functionality, supporting applications in separation, storage, and catalysis Li et al. [2024], Jia et al. [2022], Bavykina et al. [2020]. The same modularity produces a combinatorial design space that precludes exhaustive experimental or first-principles screening Burner et al. [2023], Liu et al. [2024], Sriram et al. [2024]. Machine learning can accelerate MOF screening, but labels remain task-specific. Popular quantum-chemical property datasets such as QMOF contain roughly 20,000 structures Rosen et al. [2021], whereas experimental datasets like WS24 contain only about 1,000 experimentally derived water-stability labels Terrones et al. [2024]. on the other hand, organic molecule datasets such as QM9 and ZINC cover substantially larger chemical spaces Ramakrishnan et al. [2014], Tingle et al. [2023], Gómez-Bombarelli et al. [2018]. This mismatch motivates our attempt to develop models that learn reusable representations of recurring MOF components and transfer them from substructure-level pretraining to material-level property prediction.

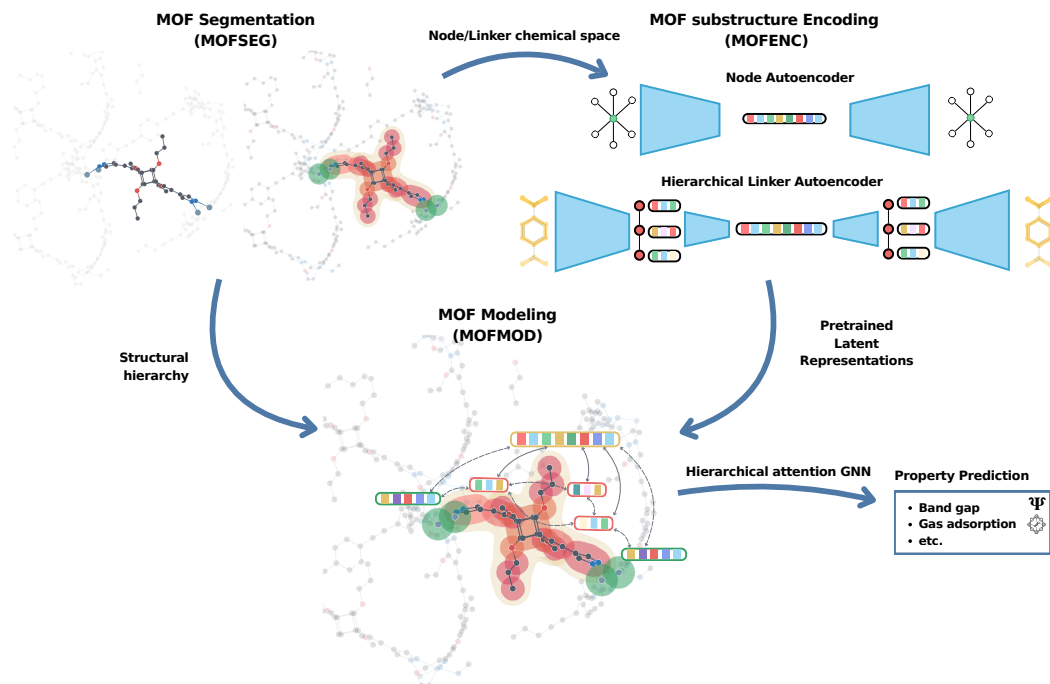


Figure 1: **Hierarchical workflow.** MOFSEG identifies atom-mapped nodes, linkers, and linker motifs. MOFENC embeds these components. MOFMOD combines the learned representations with periodic geometry for property prediction.

We exploit the compositional hierarchy of MOFs through three hierarchical modules: MOFSEG, MOFENC, and MOFMOD (Fig. 1). MOFSEG maps MOFs to chemically defined substructures while preserving atom correspondence. MOFENC learns latent representations of organic and inorganic components using autoencoders. MOFMOD performs property prediction with coupled atom-, motif-, linker-, and node-level graphs. Unlike string-based decompositions such as MOFid Bucior et al. [2019], the representation retains explicit structural mapping and can therefore combine component pretraining with MOF geometry. The design enforces three properties needed for transfer across MOF datasets: complete atom assignment, chemically interpretable coarse units, and reversible mappings between coarse tokens and the parent crystal. These constraints allow molecular-scale representations to be reused without discarding periodic geometry at prediction.

2 Hierarchical MOF representation

MOFSEG. MOFSEG represents each MOF as a periodic atomistic graph and separates inorganic node fragments from organic linkers using connectivity and chemical priors. Linkers are unwrapped across periodic boundaries, reconstructed as RDKit molecular graphs Landrum [2025], and partitioned into a complete, non-overlapping set of ring, functional-group, and residual motifs. Every linker atom belongs to exactly one motif, and all assignments retain indices in the original periodic structure (see Appendix A for details). MOFSEG processes 148,681 of 159,289 structures (93.3%) across WS24, QMOF, and hMOF, producing 208,409 linker records in the final combined corpus (Table A1). Structures are accepted only when periodic reconstruction, valence checks, atom mapping, and partition completeness are consistent. MOFs for which linkers could not be transformed to viable SMILES strings were discarded. Consequently, the resulting component records can be traced back to the original crystal and reused across representation-learning and supervised tasks.

MOFENC. MOFENC uses separate autoencoders for organic and inorganic components. For organic linkers, it employs a hierarchical autoencoder that first propagates atom-level features, then pools atoms into chemically defined motifs, and finally pools motif representations into a linker latent. The motif-level latent representations learned within this hierarchy provide reusable embeddings of

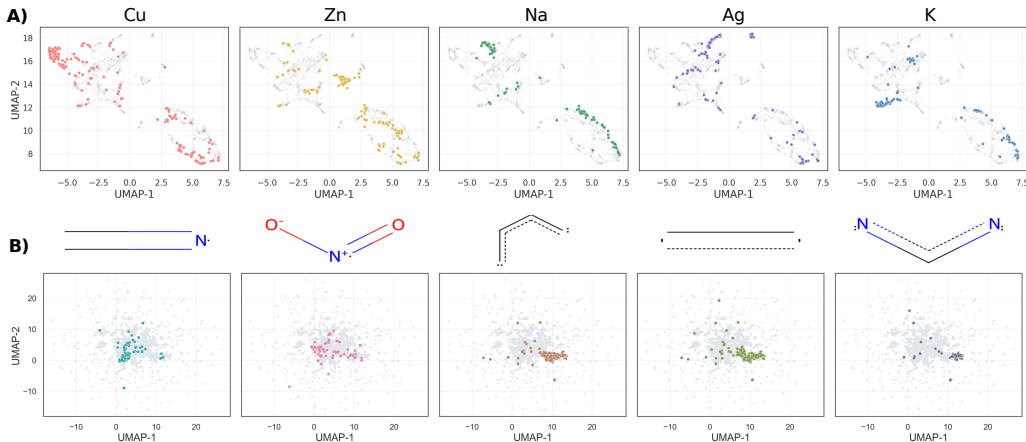


Figure 2: **MOFENC latent spaces.** UMAP projections of QMOF node embeddings (left) and hierarchical linker embeddings (right), colored by representative node elements and linker motifs.

recurring chemical functionality. The linker decoder reconstructs motif identity, motif attachments, charge, and auxiliary molecular properties. A separate graph autoencoder embeds inorganic nodes and reconstructs node composition, connectivity, coordination, and atom presence. The organic and inorganic encoders use independent parameters and latent spaces because the two component classes have distinct chemistry and reconstruction targets. Linker encoders are trained on broader unlabeled molecule datasets spanning QMOF, hMOF, and WS24, whereas node encoders use inorganic fragments extracted from the MOF datasets (see Figure 1). This separation enables component-level representation learning before fitting a crystal-level property model. The resulting latent spaces exhibit organization by linker and node composition (Fig. 2), full architecture descriptions are given in Appendix B. Since MOFENC represents components through graph encoders rather than identity-specific lookup tables, a linker or inorganic node not encountered during supervised property-model training can be embedded from its molecular graph using the pretrained encoder, provided its atom and motif types are supported by the learned vocabulary. The present random-split experiments do not explicitly evaluate generalization to entirely unseen component classes or chemical vocabularies.

MOFMOD. MOFMOD constructs three periodic graphs: atoms, motifs plus inorganic nodes, and linkers plus inorganic nodes. Atom tokens use learned element embeddings, while higher-level tokens use precomputed MOFENC representations. Resolution-specific directed CGCNN layers Xie and Grossman [2018] encode radial and three-body angular information. MOFSEG membership maps define bidirectional atom-motif/node and motif-linker exchange, avoiding inference of hierarchy from distance alone. Motif, linker, and node states are pooled separately and combined with a rotation-invariant unit-cell representation for readout (see Figure 1). The three resolutions serve complementary roles: atom graphs preserve local chemistry, motif-node graphs connect functional groups to coordination environments, and linker-node graphs provide a coarse communication path across the framework. Upward and downward hierarchy updates are parameterized independently, so information transferred from atoms to motifs need not be identical to information returned from motifs to atoms. Cross-resolution communication is defined by MOFSEG membership rather than geometric proximity. Upward and downward transformations use independent parameters, allowing coarse tokens to aggregate atomistic information and subsequently return context-dependent information to their constituents. The property head therefore receives both geometric features learned within each resolution and chemically constrained information exchanged between resolutions (see Appendix C for implementation details).

3 Results

We evaluate MOFMOD on QMOF band-gap prediction Rosen et al. [2021]. Structures passing MOFSEG checks are split randomly into 80% training, 10% validation, and 10% test sets. MOFMOD uses two directed three-body CGCNN layers at each resolution, with cutoffs of 4, 8, and 15 Å for

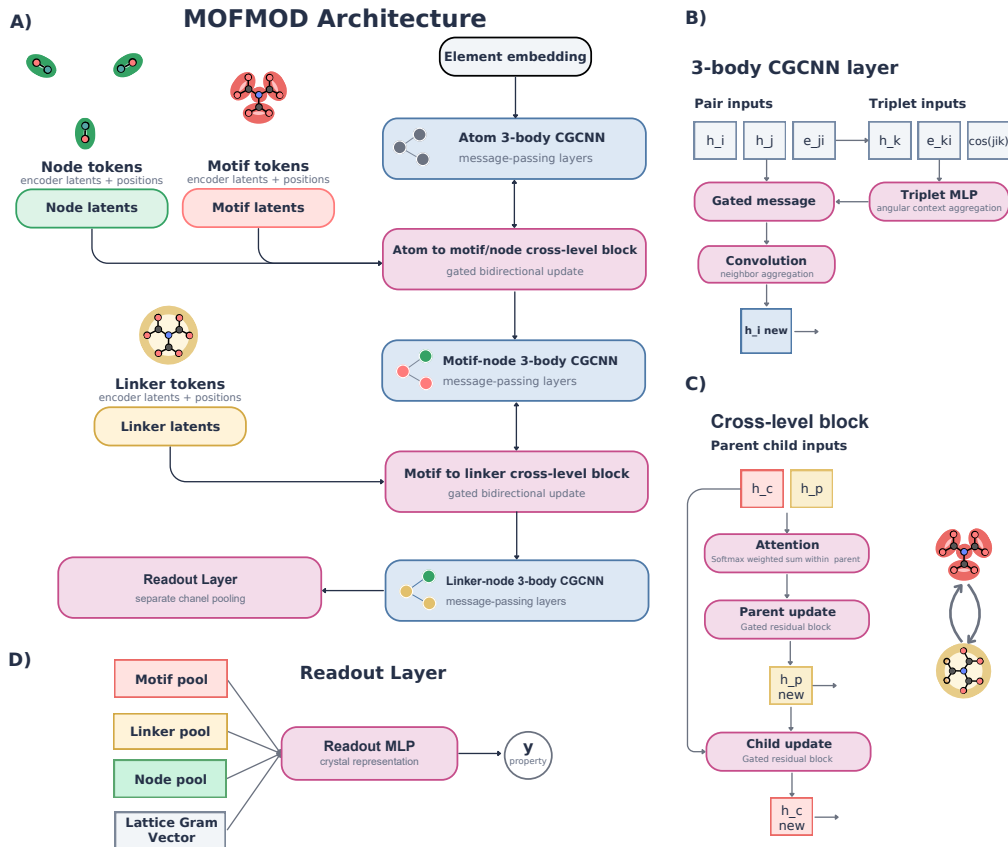


Figure 3: **MOFMOD architecture.** Three resolution-specific geometric graphs are coupled by bidirectional hierarchy blocks. Separate motif, linker, and node pools are combined with unit-cell geometry for readout.

the atom, motif-node, and linker-node graphs. MOFMOD obtains a test MAE of **0.211 eV** (Table 1). Replacing the pretrained MOFENC representations with trainable embeddings, while retaining the hierarchical architecture and cross-resolution exchange, increases the MAE to 0.321 eV, supporting the contribution of component-level pretraining. Literature results are included to establish the performance scale, but differences in filtering, featurization, and split construction prevent controlled model ranking.

Performance is also assessed for MOFMOD on CO₂ adsorption at 298 K and 2.5 bar using the hMOF dataset Wilmer et al. [2012]. With 5,610 segmented structures and an 80/10/10 random split, MOFMOD obtains an MAE of **0.580 mol,kg⁻¹**. Expanding the dataset to 128,370 successfully segmented hMOFs reduces the MAE to **0.375 mol,kg⁻¹**, showing that the hierarchical model continues to benefit from additional property-labeled data. Reported results from established hMOF models Park et al. [2023], Choudhary et al. [2022], Cao et al. [2023] are provided only as contextual reference because dataset sizes and evaluation protocols differ across studies (Table 2).

4 Discussion

MOFMOD represents MOFs at chemically defined resolutions rather than only as atom graphs or global sequences. This separates reusable component identity from crystal-specific context and enables efficient information exchange across atoms, motifs, linkers, and nodes. The MOFENC latent spaces show chemically structured organization, while the QMOF ablation shows that replacing pretrained component representations increases the band-gap MAE from 0.211 to 0.321 eV. On hMOF, MOFMOD trained on only 5,610 structures achieves an MAE of 0.580 mol,kg⁻¹, comparable to literature models trained on more than 130,000 structures, while the full-data model reaches

Table 1: QMOF band-gap prediction. Literature values are reproduced from the cited studies and are provided only as performance context because filtering, featurization, and data splits differ. MOFMOD values correspond to single training runs. In MOFMOD w/o pretrained latents, pretrained MOFENC representations are replaced by trainable embeddings while the hierarchical architecture and cross-resolution exchange are retained. This comparison therefore evaluates component pretraining rather than hierarchy itself.

Model	MAE (eV)
CGCNN Rosen et al. [2021]	0.274 ± 0.008
MOFTransformer Scratch Kang et al. [2023]	0.270
MOFTransformer Pretrained Kang et al. [2023]	0.224
MGT Anselmi et al. [2024]	0.240
EdgeConv-GANN Yang et al. [2025]	0.224 ± 0.006
MOFMOD	0.211
MOFMOD w/o pretrained latents	0.321

Table 2: CO₂ adsorption prediction at 298 K and 2.5 bar on hMOF. Literature results use different filtering and evaluation protocols and are therefore included only as performance context. MOFMOD uses an 80/10/10 random split.

Model	Architecture	hMOFs used	MAE (mol kg ⁻¹)
ALIGNN Choudhary et al. [2022]	Angular GNN	137,953	0.480
CGCNN _{scratch} Cao et al. [2023]	Crystal GNN	137,652	0.818 ± 0.050
CGCNN _{pretrain} Cao et al. [2023]	Pretrained GNN	137,652	0.645 ± 0.003
SOAP Cao et al. [2023]	Descriptor + ML	137,652	0.666 ± 0.003
MOFormer _{pretrain} Cao et al. [2023]	Transformer	137,652*	0.982 ± 0.011
MOFMOD _{5k}	Hierarchical GNN	5,610	0.580
MOFMOD_{full}	Hierarchical GNN	128,370	0.375

*MOFormer uses only structures for which a valid MOFid representation is available; its effective training set is therefore smaller than the complete 137,652-structure hMOF dataset.

0.375 mol.kg⁻¹, substantially below the reported literature values. MOFSEG additionally yields reusable atom-mapped decompositions for downstream analysis and makes the coarse representation inspectable at the level of explicit chemical fragments.

A central motivation for this factorization is data reuse. Property labels are attached to complete MOFs, whereas linker and node chemistry recur across structures and datasets. Component pretraining therefore decouples representation learning from a specific target property and provides a common interface between molecular corpora and periodic materials. Our experiments establish the feasibility of hierarchical, component-pretrained representations for MOF property prediction on both QMOF band-gap prediction and hMOF CO₂ adsorption. More broadly, the same hierarchical strategy can be adapted to other modular porous materials, including covalent organic frameworks and zeolites. Future work will assess data efficiency in low-data regimes, pretraining on large organic chemical spaces like ZINC and QM9, learning across multiple properties, and generalization to entirely unseen linkers and nodes. Overall, this framework provides a general and extensible route toward reusable chemical representations for property prediction across diverse porous materials.

Acknowledgments and Disclosure of Funding

Funded by the European Union. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. This work was funded by the ERC (StG SupraModel), grant 101077842.

References

Marco Anselmi, Greg Slabaugh, Rachel Crespo-Otero, and Devis Di Tommaso. Molecular graph transformer: stepping beyond alignn into long-range interactions. *Digital Discovery*, 3(5):1048–1057, 05 2024. ISSN

- 2635-098X. doi: 10.1039/d4dd00014e. URL <https://doi.org/10.1039/d4dd00014e>.
- Anastasiya Bavykina, Nikita Kolobov, Il Son Khan, Jeremy A Bau, Adrian Ramirez, and Jorge Gascon. Metal-organic frameworks in heterogeneous catalysis: recent progress, new trends, and future perspectives. *Chemical reviews*, 120(16):8468–8535, 2020.
- James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Yash Katariya, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL <http://github.com/jax-ml/jax>.
- Benjamin J Bucior, Andrew S Rosen, Maciej Haranczyk, Zhenpeng Yao, Michael E Ziebel, Omar K Farha, Joseph T Hupp, J Ilja Siepmann, Alán Aspuru-Guzik, and Randall Q Snurr. Identification schemes for Metal-Organic frameworks to enable rapid search and cheminformatics analysis. *Crystal Growth & Design*, 19(11):6682–6697, November 2019.
- Jake Burner, Ludwig Schwiedrzik, Mykhaylo Krykunov, Jun Luo, Peter G Boyd, and Tom K Woo. High-Performing deep learning regression models for predicting Low-Pressure CO₂ adsorption properties of Metal-Organic frameworks. *J. Phys. Chem. C*, 124(51):27996–28005, December 2020.
- Jake Burner, Jun Luo, Andrew White, Adam Mirmiran, Ohmin Kwon, Peter G. Boyd, Stephen Maley, Marco Gibaldi, Scott Simrod, Victoria Ogden, and Tom K. Woo. Arc-mof: A diverse database of metal-organic frameworks with dft-derived partial atomic charges and descriptors for machine learning. *Chemistry of Materials*, 35(3):900–916, Feb 2023. ISSN 0897-4756. doi: 10.1021/acs.chemmater.2c02485. URL <https://doi.org/10.1021/acs.chemmater.2c02485>.
- Zhonglin Cao, Rishikesh Magar, Yuyang Wang, and Amir Barati Farimani. MOFormer: Self-Supervised transformer model for Metal-Organic framework property prediction. *J. Am. Chem. Soc.*, 145(5):2958–2967, February 2023.
- Kamal Choudhary, Taner Yildirim, Daniel W. Siderius, A. Gilad Kusne, Austin McDannald, and Diana L. Ortiz-Montalvo. Graph neural network predictions of metal organic framework co₂ adsorption properties. *Computational Materials Science*, 210:111388, 2022. ISSN 0927-0256. doi: <https://doi.org/10.1016/j.commatsci.2022.111388>. URL <https://www.sciencedirect.com/science/article/pii/S092702562200163X>.
- Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 01 2018. ISSN 2374-7943. doi: 10.1021/acscentsci.7b00572. URL <https://doi.org/10.1021/acscentsci.7b00572>.
- John J. Irwin, Teague Sterling, Michael M. Mysinger, Erin S. Bolstad, and Ryan G. Coleman. Zinc: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, 52(7):1757–1768, 05 2012. ISSN 1549-9596. doi: 10.1021/ci3001277. URL <https://doi.org/10.1021/ci3001277>.
- Tao Jia, Yifan Gu, and Fengting Li. Progress and potential of metal-organic frameworks (mofs) for gas storage and separation: A review. *Journal of Environmental Chemical Engineering*, 10(5):108300, 2022.
- Yeonghun Kang, Hyunsoo Park, Berend Smit, and Jihan Kim. A multi-modal pre-training transformer for universal transfer learning in metal-organic frameworks. *Nature Machine Intelligence*, 5(3):309–318, March 2023.
- Greg Landrum. Rdkit: Open-source cheminformatics software. 2025. URL https://github.com/rdkit/rdkit/releases/tag/Release_2025_09_3.
- Dongxiao Li, Anurag Yadav, Hong Zhou, Kaustav Roy, Pounraj Thanasekaran, and Chengkuo Lee. Advances and applications of metal-organic frameworks (mofs) in emerging technologies: a comprehensive review. *Global Challenges*, 8(2):2300244, 2024.
- Shanping Liu, Romain Dupuis, Dong Fan, Salma Benzaria, Mickaele Bonneau, Prashant Bhatt, Mohamed Eddaoudi, and Guillaume Maurin. Machine learning potential for modelling h₂ adsorption/diffusion in mofs with open metal sites. *Chem. Sci.*, 15:5294–5302, 2024. doi: 10.1039/D3SC05612K. URL <http://dx.doi.org/10.1039/D3SC05612K>.
- Hyunsoo Park, Yeonghun Kang, and Jihan Kim. Enhancing Structure-Property relationships in porous materials through transfer learning and Cross-Material Few-Shot learning. *ACS Appl. Mater. Interfaces*, 15(48):56375–56385, December 2023.
- Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, August 2014.

- Andrew S. Rosen, Shaelyn M. Iyer, Debmalya Ray, Zhenpeng Yao, Alán Aspuru-Guzik, Laura Gagliardi, Justin M. Notestein, and Randall Q. Snurr. Machine learning the quantum-chemical properties of metal organic frameworks for accelerated materials discovery. *Matter*, 4(5):1578–1597, May 2021. ISSN 2590-2393. doi: 10.1016/j.matt.2021.02.015. URL <https://doi.org/10.1016/j.matt.2021.02.015>.
- Anuroop Sriram, Sihoon Choi, Xiaohan Yu, Logan M Brabson, Abhishek Das, Zachary Ulissi, Matt Uyttendaele, Andrew J Medford, and David S Sholl. The open DAC 2023 dataset and challenges for sorbent discovery in direct air capture. *ACS Cent. Sci.*, 10(5):923–941, May 2024.
- Gianmarco G Terrones, Shih-Peng Huang, Matthew P Rivera, Shuwen Yue, Alondra Hernandez, and Heather J Kulik. Metal–Organic framework stability in water and harsh environments from Data-Driven models trained on the diverse WS24 data set. *J. Am. Chem. Soc.*, 146(29):20333–20348, July 2024.
- Benjamin I Tingle, Khanh G Tang, Mar Castanon, John J Gutierrez, Munkhzul Khurelbatar, Chinzorig Dandarchuluun, Yurii S Moroz, and John J Irwin. ZINC-22-A free Multi-Billion-Scale database of tangible compounds for ligand discovery. *J. Chem. Inf. Model.*, 63(4):1166–1176, February 2023.
- Christopher E Wilmer, Michael Leaf, Chang Yeon Lee, Omar K Farha, Brad G Hauser, Joseph T Hupp, and Randall Q Snurr. Large-scale screening of hypothetical metal–organic frameworks. *Nature Chemistry*, 4(2): 83–89, February 2012.
- Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.*, 120:145301, Apr 2018. doi: 10.1103/PhysRevLett.120.145301. URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.145301>.
- Zexi Yang, Qi Yu, Yapeng Zhan, and Jiying Liu. Incorporating edge convolution and correlative self-attention into graph neural network for material properties prediction. *Machine Learning: Science and Technology*, 6(1):015020, jan 2025. doi: 10.1088/2632-2153/ad9fcf. URL <https://doi.org/10.1088/2632-2153/ad9fcf>.
- Shuai Yuan, Jiayu Peng, Bin Cai, Zhehao Huang, Angel T Garcia-Esparza, Dimosthenis Sokaras, Yirui Zhang, Livia Giordano, Karthik Akkiraju, Yun Guang Zhu, René Hübner, Xiaodong Zou, Yuriy Román-Leshkov, and Yang Shao-Horn. Tunable metal hydroxide–organic frameworks for catalysing oxygen evolution. *Nature Materials*, 21(6):673–680, June 2022.

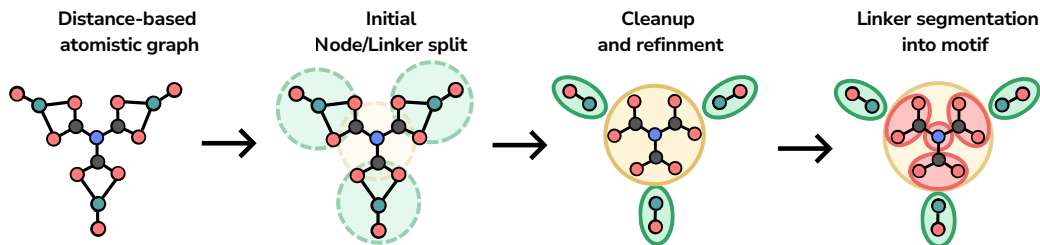


Figure A1: **Overview of the MOFSEG workflow.** MOFSEG represents each MOF as a periodic atomistic graph and segments it into atom-mapped inorganic nodes and organic linkers. The initial node/linker partition is refined by reassigning small or isolated species and merging nearby node clusters. Each linker is then converted into a molecular graph and decomposed into chemical motifs, producing a hierarchical representation of linkers for downstream embedding and property-prediction models.

A MOFSEG: periodic segmentation and motif assignment

MOFSEG starts from a periodic atomistic graph (see A1). Node seeds are taken from QMOF node descriptors when available; otherwise, metals and selected elements excluded from linkers define the initial node set. All other atoms are assigned provisionally to linkers. Periodic edges are inferred from minimum-image distances and covalent radii from `pymatgen`. An edge between atoms i and j is proposed when

$$d_{ij} \leq \alpha (r_i^{\text{cov}} + r_j^{\text{cov}}), \quad (\text{A1})$$

with $\alpha = 1.30$, reduced to 1.25 for oxygen bonded to elements other than C, N, S, or P. Maximum parent distances are 1.20 Å for H and 1.70, 2.10, 2.20, and 2.60 Å for F, Cl, Br, and I. Terminal atoms are assigned to at most one heavy-atom parent. H–H and O–O edges are removed; oxygen preferentially retains its nearest C, N, S, or P neighbor. Maximum degrees for B, C, N, O, F, P, S, Cl, Br, I, and H are 4, 4, 3, 2, 1, 5, 6, 1, 1, 1, and 1.

Connected components of the pruned linker graph define provisional linkers. Single-atom components are reassigned to bonded nodes or linkers when possible, otherwise to the nearest periodic neighbor. Each linker is unwrapped by graph traversal using minimum-image displacements. RDKit bond orders and formal charges are tested in the order 0, $-1, \dots, -8, +1, \dots, +4$; the first graph passing sanitization, valence, atom-map, and connectedness checks is retained. Molecular reconstruction uses maximum parent distances of 1.35, 1.80, 2.20, 2.40, and 2.60 Å for H, F, Cl, Br, and I. Up to three long non-terminal edges may be removed to resolve overvalence if connectivity is preserved. Water- or hydroxide-like oxygen species are reassigned to nodes. Structures failing reconstruction or partition validation are excluded.

Node fragments are connected components of the final node-induced graph. Metal-containing nodes use element-labeled single-bond topology over the assigned atoms. For metal-free nodes, bond orders are inferred by testing charges 0, $+1, \dots, +6, -1, \dots, -6$. Linker rings are assigned before non-ring motifs. Isolated rings are labeled by aromaticity, size, composition, formal charge, and cyclic topology. Overlapping rings are split into disjoint ring-derived segments and shared junctions. Functional groups are assigned from local bond order and coordination; remaining atoms receive alkyl or element labels. Connected atoms with the same label define one motif instance. Hydrogens are used for validation and assigned to their heavy-atom parent’s motif but are omitted from exported motif labels. Centers of mass are computed after graph-based periodic unwrapping and wrapped into the reference cell.

B MOFENC architecture

MOFENC contains a hierarchical variational autoencoder for organic linkers and a separate graph autoencoder for inorganic nodes. The models use independent parameters, latent spaces, and reconstruction objectives.

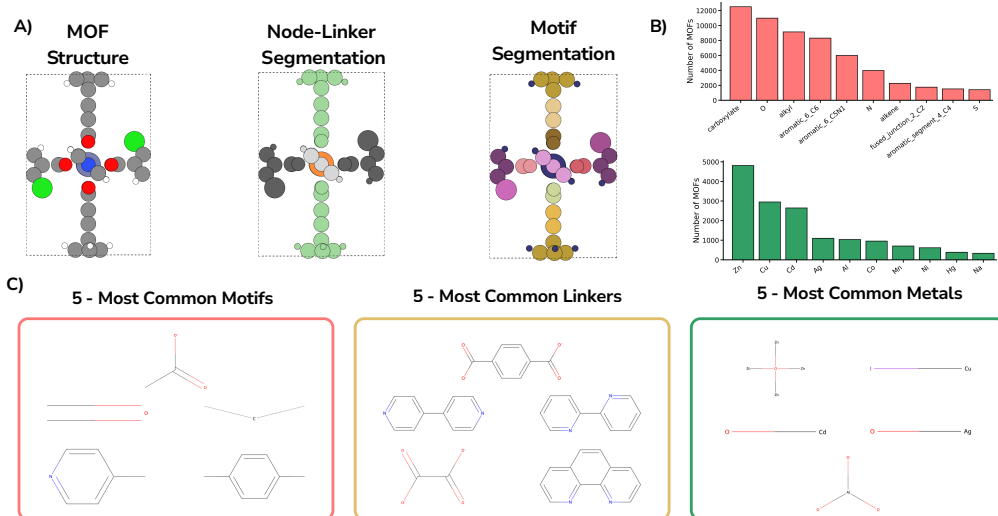


Figure A2: **QMOF segmentation analysis**. Example atom-, node/linker-, and motif-level representations, distributions of common node metals and linker motifs, and representative frequent motifs, linkers, and node environments.

Table A1: MOFSEG processing summary. The total linker count reports the final merged molecules.

Dataset	Attempted	Successful	Rate	Linkers
WS24	911	810	88.9%	577
hMOF	137,953	128,370	93.1%	176,623
QMOF	20,425	19,501	95.5%	10,472
Total	159,289	148,681	93.3%	208,409

B.1 Hierarchical linker representation

For linker L , the atom graph $\mathcal{G}_L^a = (\mathcal{V}_L^a, \mathcal{E}_L^a)$ stores atomic number, hydrogen count, and directed bond type. The motif graph $\mathcal{G}_L^m = (\mathcal{V}_L^m, \mathcal{E}_L^m)$ contains one node per MOFSEG motif and edges induced by inter-motif atomic bonds. The map $\pi : \mathcal{V}_L^a \rightarrow \mathcal{V}_L^m$ links the two resolutions.

Atom message passing produces states $\mathbf{h}_i^a$. Atoms assigned to motif u are pooled and combined with a motif-type embedding before motif-level message passing:

$$\tilde{\mathbf{h}}_u^m = \text{AttnPool}_a(\{\mathbf{h}_i^a\}_{i \in \mathcal{A}_u}), \quad (\text{A2})$$

$$\{\mathbf{h}_u^m\} = \text{MPN}_m(\mathcal{G}_L^m, \{W_{a \rightarrow m} \tilde{\mathbf{h}}_u^m + W_{\text{type}} E_M(c_u)\}). \quad (\text{A3})$$

The linker state concatenates attention-pooled motifs with mean-pooled atom and motif states. It parameterizes a diagonal Gaussian latent:

$$\mathbf{h}_L^{\text{pool}} = \text{LN}[\text{Concat}(\mathbf{h}_L^{\text{attn}}, \bar{\mathbf{h}}_L^a, \bar{\mathbf{h}}_L^m)], \quad (\text{A4})$$

$$\boldsymbol{\mu}_L = \text{MLP}_\mu(\mathbf{h}_L^{\text{pool}}), \quad \log \sigma_L^2 = \text{MLP}_{\log \sigma^2}(\mathbf{h}_L^{\text{pool}}), \quad (\text{A5})$$

$$\mathbf{z}_L = \boldsymbol{\mu}_L + \sigma_L \odot \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (\text{A6})$$

with $\mathbf{z}_L = \boldsymbol{\mu}_L$ at inference. An auxiliary head predicts normalized LogP, QED, and synthetic-accessibility values. The decoder traverses motifs and predicts motif type, attachment, total charge, and auxiliary properties. The linker objective is

$$\mathcal{L}_L = \gamma_{\text{type}} \mathcal{L}_{\text{type}} + \gamma_{\text{pair}} \mathcal{L}_{\text{pair}} + \gamma_{\text{charge}} \mathcal{L}_{\text{charge}} + \gamma_{\text{prop}} \mathcal{L}_{\text{prop}} + \gamma_{\text{KL}} \mathcal{L}_{\text{KL}}. \quad (\text{A7})$$

B.2 Inorganic-node representation

The node encoder processes element-labeled node atoms with dense-adjacency message passing. A learned graph token is connected bidirectionally to all valid node atoms; its final state defines the

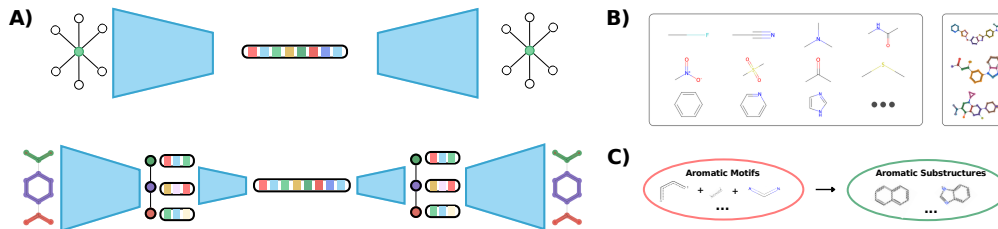


Figure A3: **MOFENC**. The node autoencoder embeds inorganic fragments. The hierarchical linker autoencoder maps atoms to motifs and motifs to a linker latent; its decoder reconstructs motif sequence and attachment. The motif vocabulary provides a complete, non-overlapping linker partition.

latent

$$\mathbf{z}_N = \text{MLP}_{z,N} \left(\left[\text{MPN}_N \left(\mathcal{G}_N^+, \{\mathbf{h}_i^N\} \cup \{\mathbf{h}_{\text{token}}^{(0)}\} \right) \right]_{\text{token}} \right). \quad (\text{A8})$$

The decoder reconstructs element identity, atom-slot presence, adjacency, and coordination classes using

$$\mathcal{L}_N = \lambda_{\text{elem}} \mathcal{L}_{\text{elem}} + \lambda_{\text{edge}} \mathcal{L}_{\text{edge}} + \lambda_{\text{coord}} \mathcal{L}_{\text{coord}} + \lambda_{\text{pres}} \mathcal{L}_{\text{pres}}. \quad (\text{A9})$$

Element identity uses masked categorical cross-entropy; adjacency uses positive-weighted binary cross-entropy over the upper triangle; coordination and presence use binary cross-entropy.

B.3 Low-level encoder and decoder definitions

B.3.1 Linker atom encoder

Linker atom and bond features are embedded as

$$\mathbf{x}_i^a = \text{LN} \left[\text{MLP}_a \left(E_Z(Z_i) \| E_H(n_i^H) \right) \right], \quad (\text{A10})$$

$$\mathbf{e}_{ji}^a = \text{MLP}_b \left(E_B(b_{ji}) \right). \quad (\text{A11})$$

Each atom layer forms directed messages from receiver, sender, and bond states,

$$\mathbf{m}_{j \rightarrow i}^{(\ell)} = \phi_m^{(\ell)} \left(\mathbf{h}_i^{(\ell)} \| \mathbf{h}_j^{(\ell)} \| \mathbf{e}_{ji}^a \right), \quad (\text{A12})$$

followed by a normalized residual update:

$$\mathbf{h}_i^{(\ell+1)} = \text{LN} \left[\mathbf{h}_i^{(\ell)} + \phi_u^{(\ell)} \left(\sum_{j \in \mathcal{N}(i)} \mathbf{m}_{j \rightarrow i}^{(\ell)} \right) \right]. \quad (\text{A13})$$

The resulting states $\mathbf{h}_i^a$ enter atom-to-motif pooling.

Atom-to-motif and motif-to-linker pooling use the learned weighted average

$$\text{AttnPool}_w(\{\mathbf{h}_i\}_{i \in I}) = \sum_{i \in I} \frac{\exp(\mathbf{w}^T \mathbf{h}_i)}{\sum_{j \in I} \exp(\mathbf{w}^T \mathbf{h}_j)} \mathbf{h}_i, \quad (\text{A14})$$

with independent scoring parameters at each hierarchy level.

B.3.2 Hierarchical linker decoder

The decoder traverses motifs rather than generating atoms sequentially. At step t , u_t and $\mathbf{z}_L$ define the next-motif distribution:

$$p(c_{t+1} | \mathbf{h}_{u_t}^m, \mathbf{z}_L) = \text{softmax} \left[\text{MLP}_{\text{type}} \left(\mathbf{h}_{u_t}^m \| \mathbf{z}_L \right) \right]. \quad (\text{A15})$$

For candidate attachment (i, j) , the pair representation is

$$\mathbf{r}_{ij}^{(t)} = \mathbf{h}_i \| \mathbf{h}_j \| (\mathbf{h}_i \odot \mathbf{h}_j) \| \mathbf{h}_i - \mathbf{h}_j \| \| \mathbf{h}_{u_t}^m \| \mathbf{a}_{ij} \| \mathbf{z}_L. \quad (\text{A16})$$

Scores are normalized over candidate pairs:

$$p_{ij}^{(t)} = \frac{\exp[\text{MLP}_{\text{pair}}(\mathbf{r}_{ij}^{(t)})]}{\sum_{(k,l) \in \mathcal{P}_t} \exp[\text{MLP}_{\text{pair}}(\mathbf{r}_{kl}^{(t)})]}, \quad (\text{A17})$$

where $\mathcal{P}_t$ is the candidate set. Equivalent correct pairs are marginalized in the loss:

$$\mathcal{L}_{\text{pair}}^{(t)} = -\log \left(\sum_{(i,j) \in \mathcal{P}_t^+} p_{ij}^{(t)} \right), \quad (\text{A18})$$

where $\mathcal{P}_t^+$ contains equivalent correct attachments.

B.3.3 Inorganic-node encoder and decoder

Node atoms are initialized from element types and processed by dense-adjacency message passing:

$$\mathbf{x}_i^N = \text{LN} [\text{MLP}_N (E_N(c_i^N))], \quad (\text{A19})$$

$$\{\mathbf{h}_i^N\} = \text{MPN}_N (\mathcal{G}_N, \{\mathbf{x}_i^N\}). \quad (\text{A20})$$

The decoder expands $\mathbf{z}_N$ to K_N candidate atom slots:

$$\{\mathbf{r}_k^{N,0}\}_{k=1}^{K_N} = \text{Expand}_N(\mathbf{z}_N), \quad (\text{A21})$$

The slots are refined on a complete graph without self-edges:

$$\{\mathbf{r}_k^N\} = \text{MPN}_{\text{dec}} (\mathcal{G}_{K_N}^{\text{complete}}, \{\mathbf{r}_k^{N,0}\}). \quad (\text{A22})$$

Independent heads predict element identity and slot presence:

$$\mathbf{p}_k^{\text{elem}} = \text{softmax} [\text{Linear}_{\text{elem}}(\mathbf{r}_k^N)], \quad (\text{A23})$$

$$p_k^{\text{pres}} = \sigma [\text{Linear}_{\text{pres}}(\mathbf{r}_k^N)]. \quad (\text{A24})$$

Adjacency reconstruction first forms a soft element embedding

$$\tilde{\mathbf{e}}_i = \sum_z p_{i,z}^{\text{elem}} E_N(z), \quad (\text{A25})$$

which is combined with the slot state to form query/key projections. The symmetric edge logit is

$$\ell_{ij}^E = \frac{1}{2} (\mathbf{q}_i^T \mathbf{k}_j + \mathbf{q}_j^T \mathbf{k}_i) + b_i + b_j + b_E, \quad p_{ij}^E = \sigma(\ell_{ij}^E). \quad (\text{A26})$$

Diagonal entries are suppressed; a separate head predicts global coordination classes.

B.4 Latent-space cluster profiles

The UMAP projections in Fig. 2 show qualitative chemical organization. Independent cluster profiles provide a complementary view: linker clusters differ in motif composition and molecular descriptors, whereas node clusters differ in elemental composition, coordination, size, and connectivity (Fig. A4). These profiles indicate that the learned latents encode chemically and structurally distinct subspaces rather than only fragment identity.

C MOFMOD architecture

MOFMOD represents each segmented crystal at three resolutions: atoms; motifs plus inorganic nodes; and linkers plus inorganic nodes. Atomic tokens retain atomic coordinates. Motif, linker, and node tokens are placed at MOFSEG centers of mass. Each resolution has an independent periodic neighbor graph and independent convolution parameters.

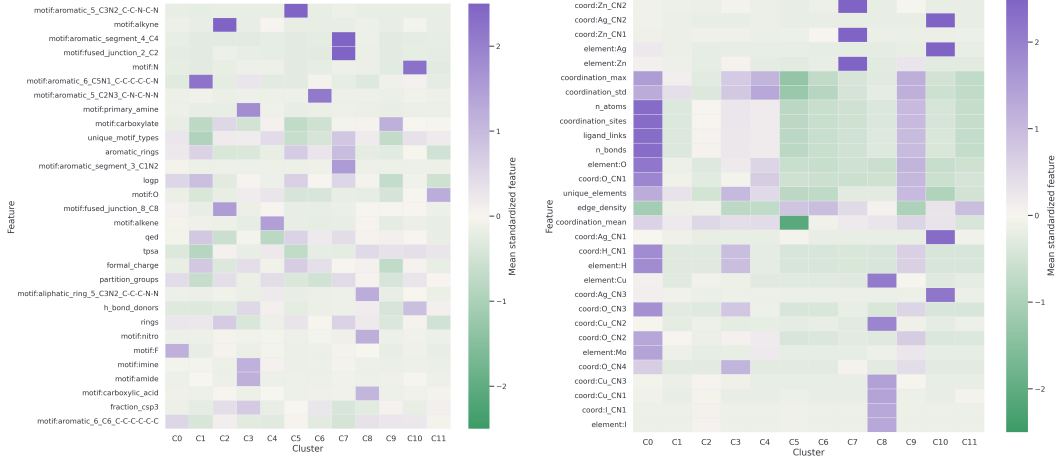


Figure A4: **Latent-space cluster profiles.** Mean standardized features for linker (left) and node (right) clusters. Distinct enrichment patterns reflect differences in chemical motifs, molecular descriptors, elemental composition, and coordination environments.

C.1 Multiresolution update

Atoms use element embeddings. Higher-level token u uses its MOFENC representation $\mathbf{x}_u$ and component class $\tau(u)$:

$$\mathbf{h}_i^{a,0} = E_Z(Z_i), \quad \mathbf{h}_u^0 = \Phi_{\tau(u)}(\mathbf{x}_u, E_{\tau}(\tau(u))). \quad (\text{A27})$$

Resolution-specific directed CGCNN layers encode radial and three-body angular information. Cross-level exchange follows the MOFSEG membership map. For child c , parent $p = \pi(c)$, and children $\mathcal{C}(p)$,

$$\alpha_{cp} = \frac{\exp s(\mathbf{h}_c, \mathbf{h}_p)}{\sum_{c' \in \mathcal{C}(p)} \exp s(\mathbf{h}_{c'}, \mathbf{h}_p)}, \quad (\text{A28})$$

$$\mathbf{r}_p = \sum_{c \in \mathcal{C}(p)} \alpha_{cp} \mathbf{h}_c. \quad (\text{A29})$$

Information then moves in both directions:

$$\mathbf{h}_p^+ = \text{LN}(\mathbf{h}_p + \mathbf{g}_p \odot \text{MLP}_{\uparrow}(\mathbf{r}_p)), \quad (\text{A30})$$

$$\mathbf{h}_c^+ = \text{LN}(\mathbf{h}_c + \mathbf{g}_c \odot \text{MLP}_{\downarrow}(\mathbf{h}_{\pi(c)}^+)). \quad (\text{A31})$$

The sequence is atom convolution $\rightarrow$ atom-to-motif/node exchange $\rightarrow$ motif-node convolution $\rightarrow$ motif-to-linker exchange $\rightarrow$ linker-node convolution $\rightarrow$ linker-to-motif feedback. This preserves known membership while increasing the effective receptive field through coarse representations.

Global cell geometry is encoded by the normalized lattice Gram matrix

$$\tilde{G} = \frac{AA^T}{V^{2/3}}, \quad V = |\det A|, \quad (\text{A32})$$

where A contains the lattice vectors. A two-layer MLP embeds the six independent components into $\mathbf{h}^{\text{lat}}$. The final crystal state is

$$\mathbf{h}^{\text{MOF}} = \text{Concat}(\bar{\mathbf{h}}^l, \bar{\mathbf{h}}^n, \bar{\mathbf{h}}^m, \mathbf{h}^{\text{lat}}), \quad \hat{y} = \text{Readout}(\mathbf{h}^{\text{MOF}}). \quad (\text{A33})$$

Motifs, linkers, and nodes are pooled separately so that the larger motif count does not dominate coarser representations.

C.2 Implementation details

C.2.1 Hierarchy maps

MOFSEG supplies deterministic cross-level maps. Atoms in linkers map to motifs; inorganic atoms map directly to node tokens:

$$\pi_{a \rightarrow mn} : \mathcal{V}^a \rightarrow \mathcal{V}^m \cup \mathcal{V}^n. \quad (\text{A34})$$

Motifs map to their parent linker:

$$\pi_{m \rightarrow l} : \mathcal{V}^m \rightarrow \mathcal{V}^l. \quad (\text{A35})$$

These maps define cross-resolution exchange; periodic neighbor graphs define within-level interactions.

C.2.2 Class-conditioned token projection

Higher-level tokens use a class-specific gated residual projection:

$$\mathbf{v}_u = W_{\tau(u)} \mathbf{x}_u, \quad (\text{A36})$$

with feature-wise gate

$$\mathbf{g}_u^{(0)} = \sigma [\phi_{g, \tau(u)} (\mathbf{v}_u \| E_{\tau}(\tau(u)))]. \quad (\text{A37})$$

The initialized state is

$$\mathbf{h}_u^0 = \mathbf{v}_u + \mathbf{g}_u^{(0)} \odot \phi_{r, \tau(u)} (\mathbf{v}_u \| E_{\tau}(\tau(u))). \quad (\text{A38})$$

C.2.3 Cross-level attention and gates

Cross-level attention uses the compatibility score

$$s(\mathbf{h}_c, \mathbf{h}_p) = \mathbf{w}_{\text{att}}^T \text{softplus} [W_{\text{att}} (\mathbf{h}_c \| \mathbf{h}_p \| \mathbf{h}_c \odot \mathbf{h}_p) + \mathbf{b}_{\text{att}}] + b'_{\text{att}}. \quad (\text{A39})$$

Upward and downward gates are independent:

$$\mathbf{g}_p = \sigma [W_g^{\uparrow} (\mathbf{h}_p \| \mathbf{r}_p) + \mathbf{b}_g^{\uparrow}], \quad (\text{A40})$$

$$\mathbf{g}_c = \sigma [W_g^{\downarrow} (\mathbf{h}_c \| \mathbf{h}_{\pi(c)}^+) + \mathbf{b}_g^{\downarrow}]. \quad (\text{A41})$$

Compatibility, message, and gate parameters are block- and direction-specific.

C.3 Three-body directed crystal graph convolution

All resolutions use the same convolutional form with independent parameters and cutoffs. Receiver i retains up to M periodic neighbors within r_c ; the zero-translation self-edge is excluded.

C.3.1 Radial and angular features

Distance d_{ji} is expanded in Q Gaussian radial basis functions:

$$e_{ji,q} = \exp \left[-\frac{(d_{ji} - \mu_q)^2}{\Delta^2} \right], \quad q = 1, \dots, Q, \quad (\text{A42})$$

where μ_q are uniformly spaced centers and Δ is the spacing.

For incoming edges $j \rightarrow i$ and $k \rightarrow i$, the angle is expanded as

$$\mathbf{c}(\theta_{jik}) = [\cos(\theta_{jik}), \dots, \cos(K_{\theta} \theta_{jik})]. \quad (\text{A43})$$

For $k \neq j$, the triplet state is

$$\mathbf{q}_{jik} = W_i \mathbf{h}_i + W_n \mathbf{h}_j + W_n \mathbf{h}_k + W_r \mathbf{e}_{ji} + W_r \mathbf{e}_{ki} + W_{\theta} \mathbf{c}(\theta_{jik}) + \mathbf{b}, \quad (\text{A44})$$

$$\mathbf{t}_{jik} = \text{softplus} [W_o \text{softplus}(\mathbf{q}_{jik}) + \mathbf{b}_o]. \quad (\text{A45})$$

Distances use the cosine cutoff

$$f_c(d) = \frac{1}{2} \left[\cos \left(\frac{\pi d}{r_c} \right) + 1 \right], \quad d \leq r_c, \quad (\text{A46})$$

with zero weight outside r_c . The angular context for $j \rightarrow i$ is

$$\mathbf{a}_{ji} = \frac{\sum_{k \in \mathcal{N}(i) \setminus \{j\}} w_{jik} \mathbf{t}_{jik}}{\max [1, \sum_{k \in \mathcal{N}(i) \setminus \{j\}} w_{jik}]}, \quad w_{jik} = f_c(d_{ji}) f_c(d_{ki}). \quad (\text{A47})$$

Table A2: Training hyperparameters.

Component	Setting	Value
Linker encoder	atom/motif/latent dim.	128 / 64 / 64
	message-passing layers	4 per resolution
	optimizer / epochs / batch	Adam / 80 / 8
	initial learning rate	2×10^{-4}
Node encoder	hidden / latent dim.	128 / 128
	optimizer / epochs / batch	Adam / 40 / 12
	initial learning rate	5×10^{-4}
MOFMOD	graph layers	2 per resolution
	cutoffs (atom/motif/linker)	4 / 8 / 15 Å
	max. neighbors	36 / 24 / 24
	optimizer / epochs / batch	AdamW / 200 / 4
	initial learning rate	5×10^{-4}
	weight decay	10^{-5}

C.3.2 Directed message update

Sender, receiver, radial, and angular states are projected to filter and signal channels:

$$\begin{bmatrix} \mathbf{f}_{ji} \\ \mathbf{s}_{ji} \end{bmatrix} = \text{LN}_e [W_e (\mathbf{h}_i \parallel \mathbf{h}_j \parallel \mathbf{e}_{ji} \parallel \mathbf{a}_{ji}) + \mathbf{b}_e]. \quad (\text{A48})$$

The directed message is

$$\mathbf{m}_{j \rightarrow i} = \sigma(\mathbf{f}_{ji}) \odot \text{softplus}(\mathbf{s}_{ji}), \quad (\text{A49})$$

and the receiver update is

$$\mathbf{h}'_i = \text{softplus} \left[\mathbf{h}_i + \text{LN}_v \left(\sum_{j \in \mathcal{N}(i)} \mathbf{m}_{j \rightarrow i} \right) \right]. \quad (\text{A50})$$

Here LN_e and LN_v denote layer normalization over valid edges and receiver objects.

D Training configuration

Two linker encoders were trained on increasingly broad unlabeled corpora: QMOF linkers; QMOF+hMOF+WS24 linkers Rosen et al. [2021], Wilmer et al. [2012], Terrones et al. [2024], Gómez-Bombarelli et al. [2018], Irwin et al. [2012]. Duplicate molecular identities were merged before combined splits; mapped representations retained atom correspondence within MOFSEG records. Motif labels were harmonized across datasets. Node encoders were trained on QMOF nodes and on the combined QMOF+hMOF+WS24 node set.

For band-gap prediction, QMOF structures passing segmentation are split 80:10:10 into training, validation, and test sets. Motif, linker, and node vectors are precomputed; linker tokens use the posterior mean μ_L . MOFMOD is optimized with mean-squared error, warm-up followed by cosine learning-rate decay, and validation-loss checkpoint selection. Experiments use JAX Bradbury et al. [2018].