UCLA Samueli Computer Science

The Thirty-Ninth Annual Conference on Neural Information Processing Systems

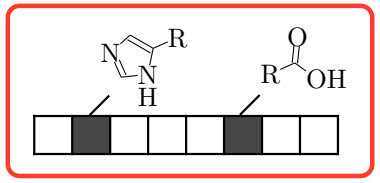
Symmetry-Preserving Conformer Ensemble Networks for Molecular Representation Learning

Yanqiao Zhu
yzhu@cs.ucla.edu
November 4, 2025

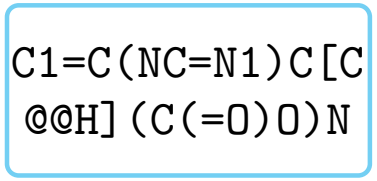
The Scalable Analytics Institute (ScAI)
Department of Computer Science
University of California, Los Angeles (UCLA)

Molecular representation learning

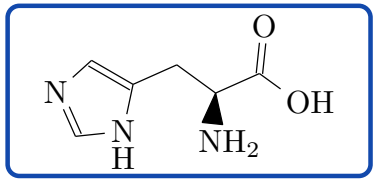
- MRL connects the chemical space and machine learning by learning meaningful molecular representations
- Current approaches: 1D/2D/3D models with efficiency vs. expressivity tradeoffs
  - 1D/2D useful without 3D structural info, more efficient
  - 3D captures key geometric effects but more computationally expensive
  - Key limitation: Most models treat molecules as rigid, static objects



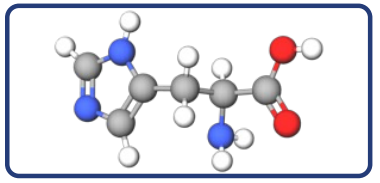
Fingerprint



SMILES String



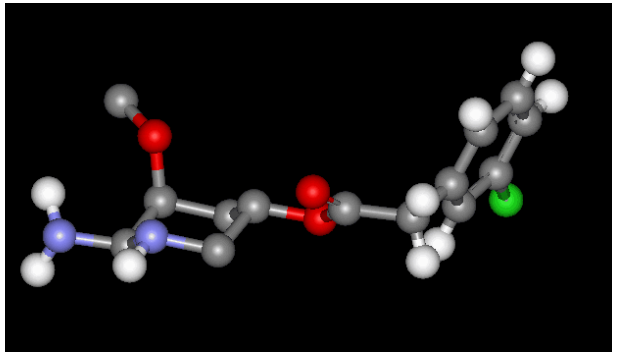
2D Topology


3D Geometry

The reality: Molecules are dynamic

- Molecules are not rigid objects but are flexible structures
  - Molecules continuously interconvert between these conformers by rotations around bonds and vibrational motions

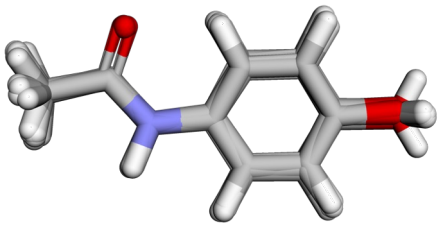
- Many properties depend on conformer ensembles, not single structures:
  - Protein-ligand binding: Conformational selection
  - Reaction mechanisms: Accessibility of geometric configurations
  - Drug activity: Distribution over accessible states



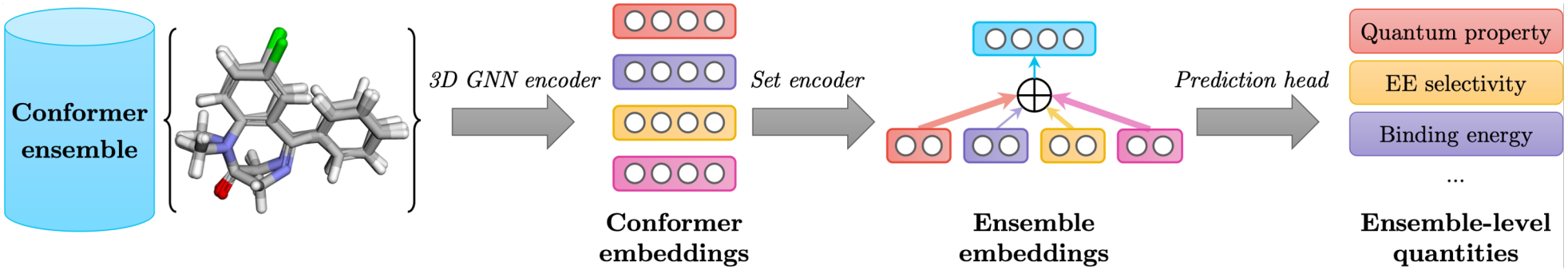
Video from
https://www.drugdesign.org/chapters/molecular-geometry/#conformers

Modeling conformer ensembles

- Question: Does explicitly encoding conformer ensemble improve MRL models compared to using only a single conformation?



- We developed the first Molecular Conformer Ensemble Learning (MARCEL) benchmark

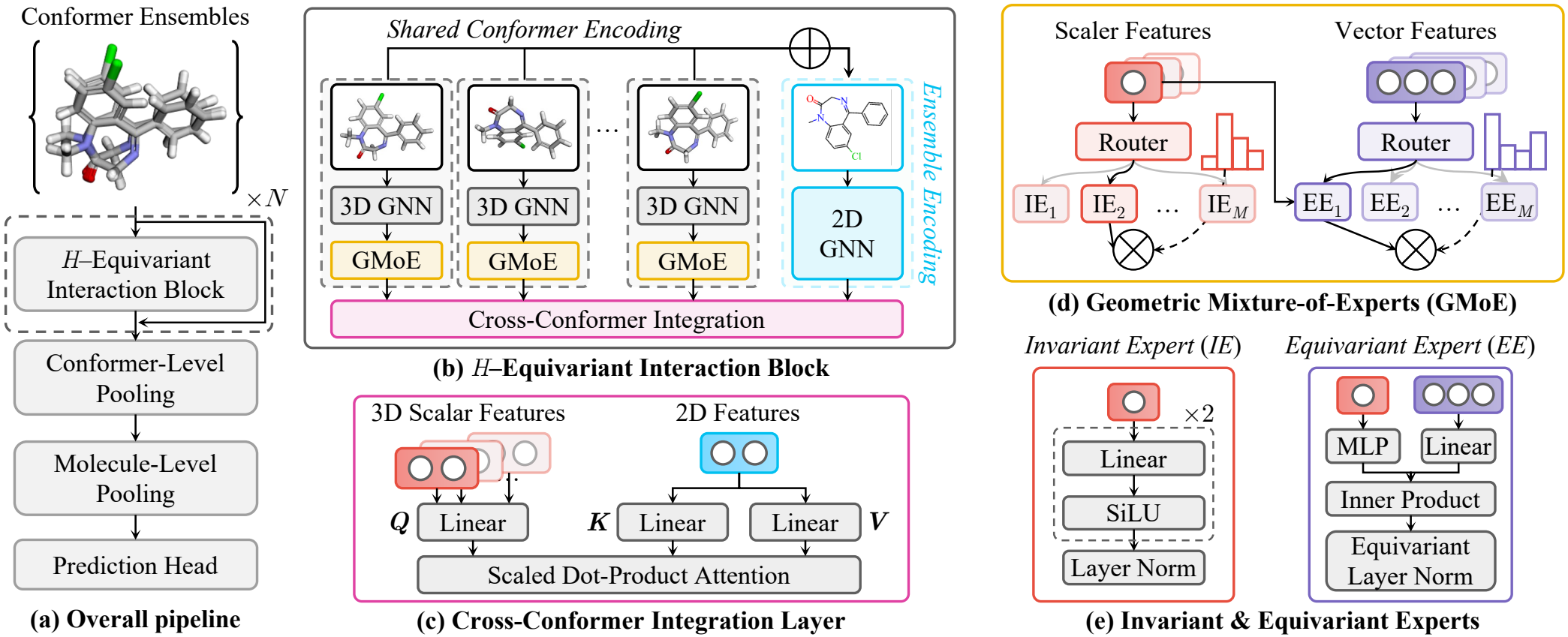


Y. Zhu et al., Learning Over Molecular Conformer Ensembles: Datasets and Benchmarks, in ICLR 2024

Our work: SPiCE

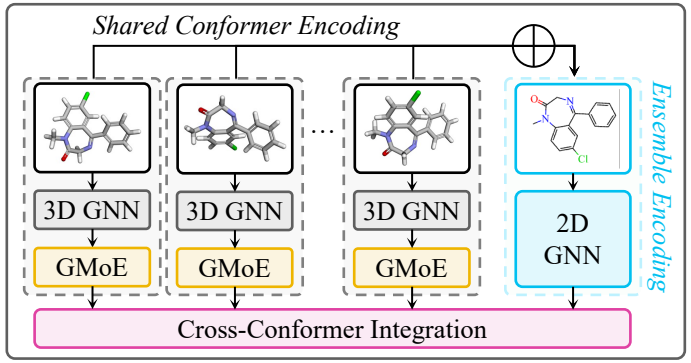
- SPiCE: **S**ymmetry-**P**reserv**i**ng **C**onformer **E**nsemble Networks
- Key design: Joint $S_n \times G^n$ equivariance
  - S_n: Conformer permutation
  - G^n: Geometric equivariance (e.g., rotation/translation)
- Novel components:
  - Geometric Mixture-of-Experts (GMoE): Separate routing for scalar vs. vector features
  - Hierarchical ensemble encoding: 2D+3D integration with cross-conformer attention
  - No alignment required: Processes raw conformer coordinates directly

SPiCE architecture overview



H-equivariant interaction blocks

- H-equivariant interaction blocks
  - Inspired by DSSNet [Maron et al., 2020]
  - Jointly preserve permutational equivariance and rotational/translational equivariance

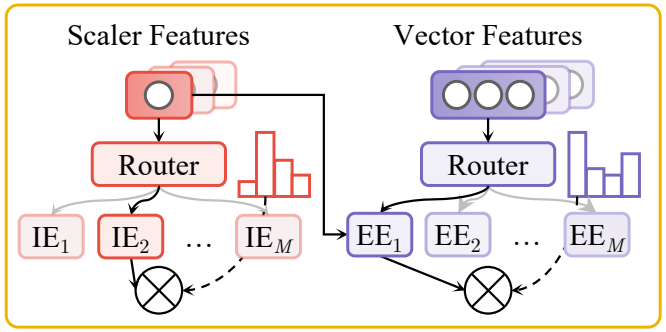


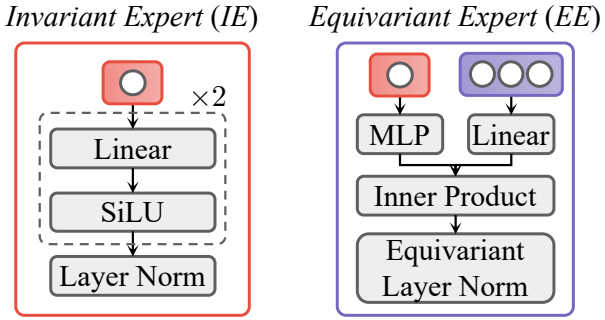
- Components:
  - Shared conformer encoding (weight-tied 3D GNNs)
  - Geometric Mixture-of-Experts (GMoE)
  - Ensemble encoding (2D GNN)
  - Cross-conformer integration (gated attention)

H. Maron et al., On Learning Sets of Symmetric Elements, in ICML 2020

Geometric Mixture-of-Experts (GMoE)

- Different atoms and structural features need specialized processing for different properties
- Key design:
  - Separate routing for scalar (type-0) vs. vector (type-1) features
    - Invariant experts (IE): Process scalar features via MLPs
    - Equivariant experts (EE): Process vector features via gated linear transformations
  - Router design: GCN-based (scalar) + inner product (vector)
    - Top-k selection with Gumbel-Sigmoid sampling





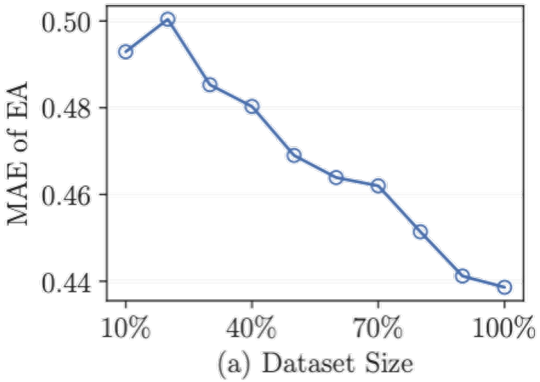
Result highlights

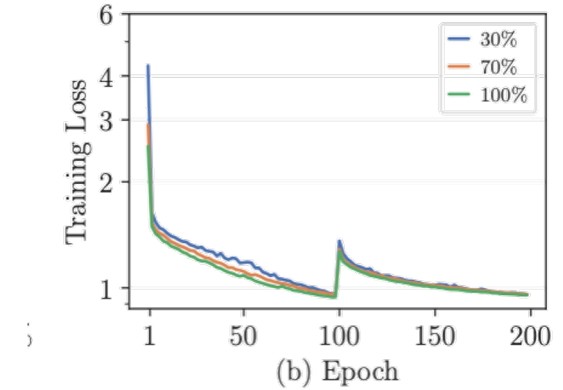
- SOTA across 34/36 configurations
- Regression (Drugs-7.5K, Kraken):
  - 3-9% MAE improvements over best baselines
- Classification (CoV2, 3CL):
  - High ROC-AUC
  - Robust to extreme class imbalance (1:60 ratio)

Backbone	Ensemble Strategy	Drugs-7.5K (MAE, ↓)			Kraken (MAE, ↓)				CoV2	3CL
		IP	EA	χ	B_5	L	$BurB_5$	BurL	ROC (↑)	ROC (↑)
► *1D String-based and 2D Topological Approaches*										
Fingerprint+RF [54]		0.5833	0.5277	0.3130	0.4760	0.4303	0.2758	0.1521	0.6071	0.9013
E3FP+RF [55]		0.6217	0.5774	0.3464	0.6249	0.5535	0.3692	0.1908	0.6046	0.7676
GIN [50]		0.5575	0.5116	0.2892	0.3128	0.4003	0.1719	0.1200	0.3708	0.5942
GIN-VN [56]		0.5398	0.5160	0.2937	0.3567	0.4344	0.2422	0.1741	0.4832	0.7387
GraphGPS [57]		0.5480	0.5054	0.2863	0.3450	0.4363	0.2066	0.1500	0.5601	0.8387
► *3D Single-Conformer Graph Neural Networks with Random Conformer Sampling [51]*										
PaiNN [37]		0.5557	0.5127	0.2924	0.3443	0.4471	0.2395	0.1673	0.2997	0.8368
ClofNet [13]		0.6316	0.6008	0.3615	0.4473	0.6369	0.3216	0.2426	0.5233	0.7562
Equiformer [34]		0.5471	0.4898	0.2887	0.2709	0.3759	0.2019	0.1526	0.4577	0.8035
ViSNet [58]		0.5393	0.4855	0.2985	0.3828	0.4495	0.2400	0.1755	0.5011	0.4774
► *Conformer Ensemble Approaches*										
ConfNet [60]		0.5760	0.5359	0.3057	0.4469	0.4680	0.2686	0.1657	0.5010	0.4930
ConAN-FGW [31]		0.5471	0.4945	0.2891	0.3242	0.5178	0.2026	0.1492	0.6340	0.9180
PaiNN [37]	Mean	0.5410	0.4966	0.2963	0.2877	0.3950	0.1817	0.1472	0.5722	0.8850
	DeepSets	0.5396	0.5091	0.2982	0.2225	0.3619	0.1693	0.1324	0.5802	0.6808
	Attention	0.6318	0.5985	0.3488	0.3496	0.4109	0.2123	0.1506	0.4179	0.6984
	SPiCE	0.5281	0.4929	**0.2792**	0.2178	0.3548	**0.1564**	0.1292	0.5910	0.8880
ClofNet [13]	Mean	0.5935	0.5441	0.3121	0.3986	0.5674	0.2857	0.2327	0.3900	0.7580
	DeepSets	0.5912	0.5533	0.3153	0.3314	0.5375	0.2532	0.1983	0.6208	0.7628
	Attention	0.6694	0.5949	0.3578	0.4979	0.6118	0.3353	0.2502	0.3707	0.8182
	SPiCE	0.5747	0.5283	0.3059	0.3193	0.4903	0.2477	0.1913	0.6730	**1.0000**
Equiformer [34]	Mean	0.5457	0.4932	0.2977	0.2303	0.3830	0.1680	0.1259	0.5601	0.8387
	DeepSets	0.5404	0.4888	0.2990	0.2564	0.3772	0.1782	0.1234	0.5125	0.7134
	Attention	0.5488	0.4923	0.2896	0.3187	0.4508	0.1673	0.1425	0.3882	0.7881
	SPiCE	0.5318	0.4830	0.2816	0.2241	**0.3456**	0.1611	**0.1229**	0.5650	0.8405
ViSNet [58]	Mean	0.5593	0.4927	0.2862	0.2811	0.3970	0.1874	0.1469	0.6035	0.7447
	DeepSets	**0.5280**	0.4987	0.2846	0.3104	0.4113	0.1716	0.1314	0.6626	0.4160
	Attention	0.5593	0.4988	0.2944	0.3755	0.4195	0.2384	0.1394	0.5262	0.7158
	SPiCE	0.5384	**0.4538**	0.2814	**0.2715**	0.3807	0.1657	0.1277	**0.6890**	0.7195

Model scaling

- Data size scaling:
  - Linear improvement from 7.5K to 75K molecules
  - MAE reduction from 0.4929 to 0.4386
  - Shows SPiCE effectively leverages additional training data without hitting capacity limits
- Training stability:
  - Consistent optimization patterns across dataset sizes
  - Demonstrates stable learning behavior suitable for varied data regimes





Key takeaways

- Technical contributions:
  - Maintains valid geometric symmetries throughout
  - Works with any equivariant 3D GNN
  - GMoE enables specialized feature handling
  - Hierarchical encoding captures ensemble relationships
- Future directions:
  - Incorporate thermodynamic priors and statistical mechanical principles
  - Extend to larger molecular systems and protein conformational ensembles
  - Develop more efficient attention mechanisms for real-time applications

Poster Session: December 3, 11 AM — 2 PM PST
Location: Exhibit Hall C,D,E
Code: https://github.com/DannieSYD/SPiCE