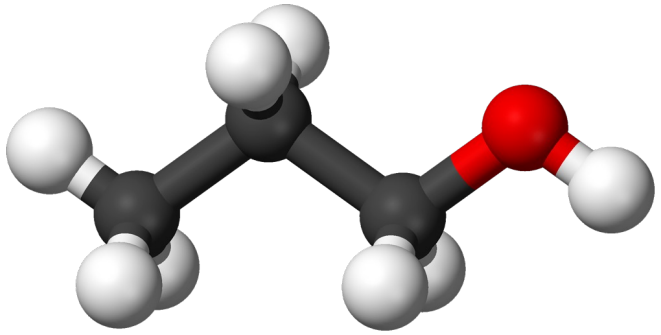
Chemomile :
An Explainable Multi-Level GNN Model for Combustion Property Prediction

Beomgyu Kang and Bong June Sung*

Department of Chemistry, Sogang University, Seoul 04107, Republic of Korea

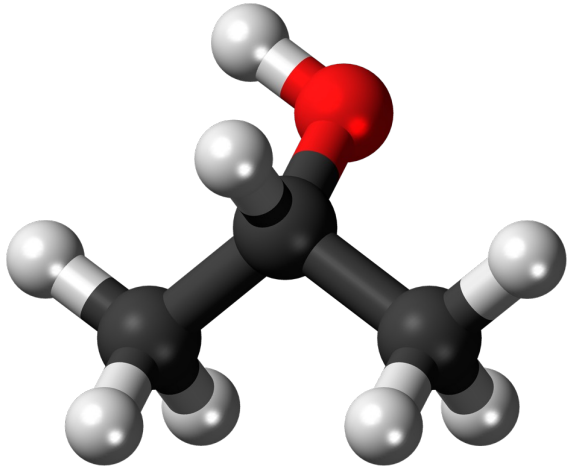
E-mail : bjsung@sogang.ac.kr

Can you ***read*** it ?



1-propanol (C_3H_8O)

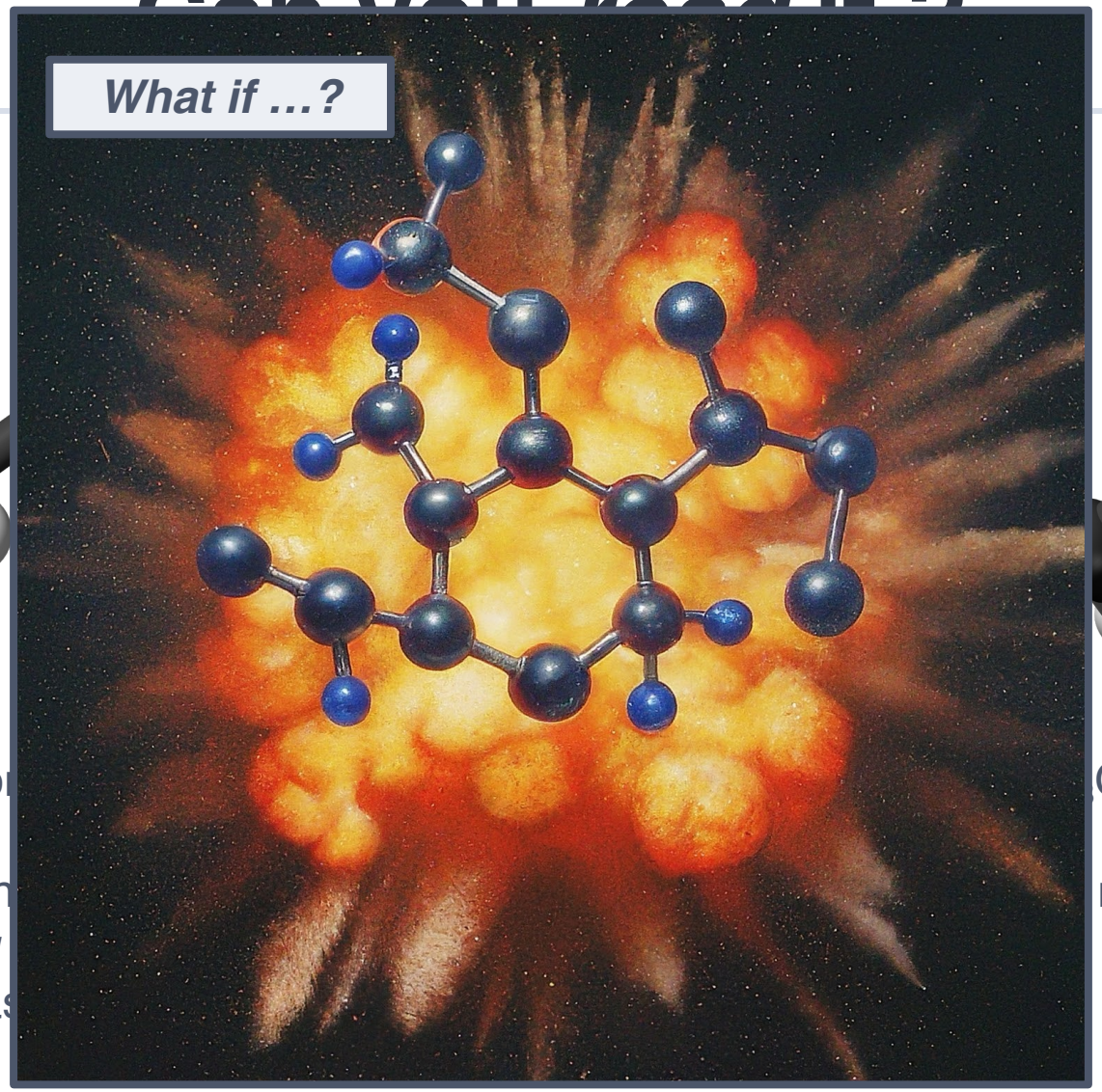
- density : 0.803 g mL^{-1}
- *log P* : 0.329
- Flash Point : 295 K



2-propanol (C_3H_8O)

- density : 0.786 g mL^{-1}
- *log P* : -0.16
- Flash Point : 284 K

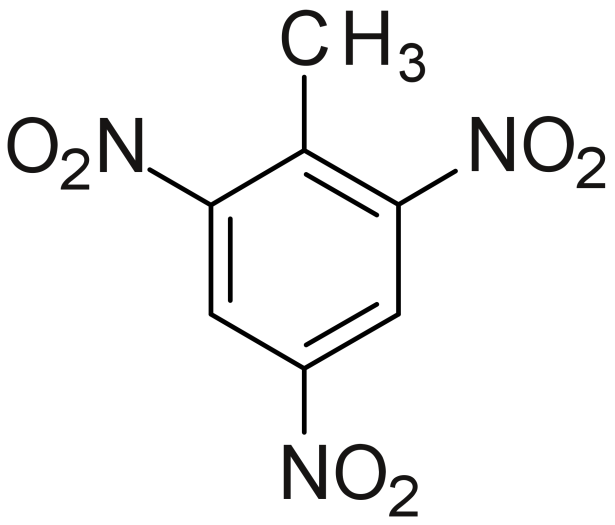
Can you *read* it ?



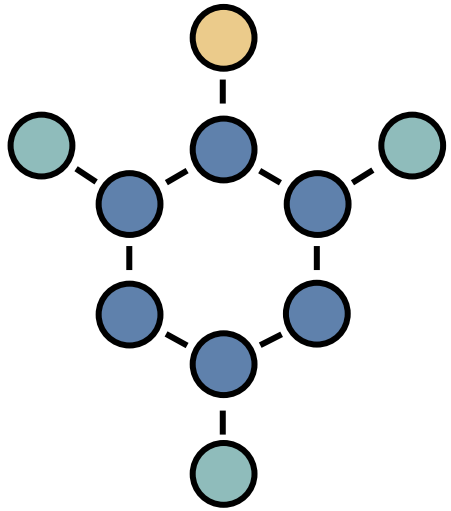
A picture from Gemini

- den mL^{-1}
- *log*
- Flas K

Molecule as a Graph



as



- ✓ We can think of a molecular graph $G = \{V,\ E\}$
  : a node(V) → an atom & an edge(E) → a bond
- ✓ However, a simple graph representation is not enough

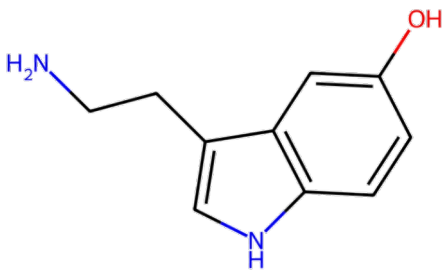
Chemomile

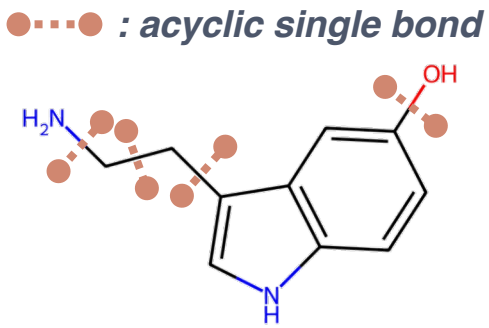
: **CHEM**ical **O**riented **M**achine **I**ntelligence **L**earning **E**ngine

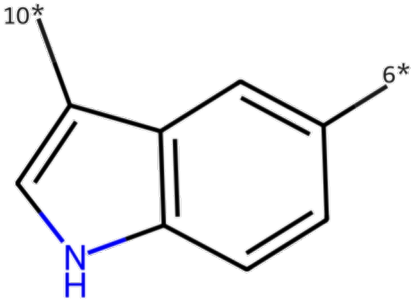
- ✓ 3D Geometry + Fragmentation
- ✓ Multi-Level Graph Representation
- ✓ Explainability via atom-wise contribution

Fragmentation

C1=CC2=C(C=C1O)C(=CN2)CCN

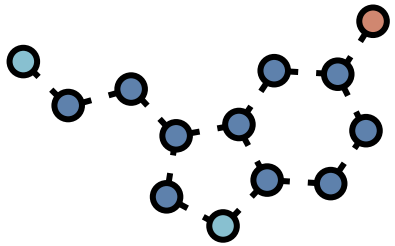




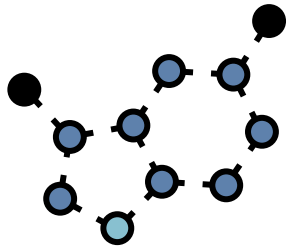


SMILES *Molecule* *Fragmentation* *Fragment*

- ✓ Multiple sub-graphs can be obtained
  - A **Molecule-Level** Subgraph
  - Several **Fragment-Level** Subgraphs
  - A **Junction Tree-Level** Subgraph

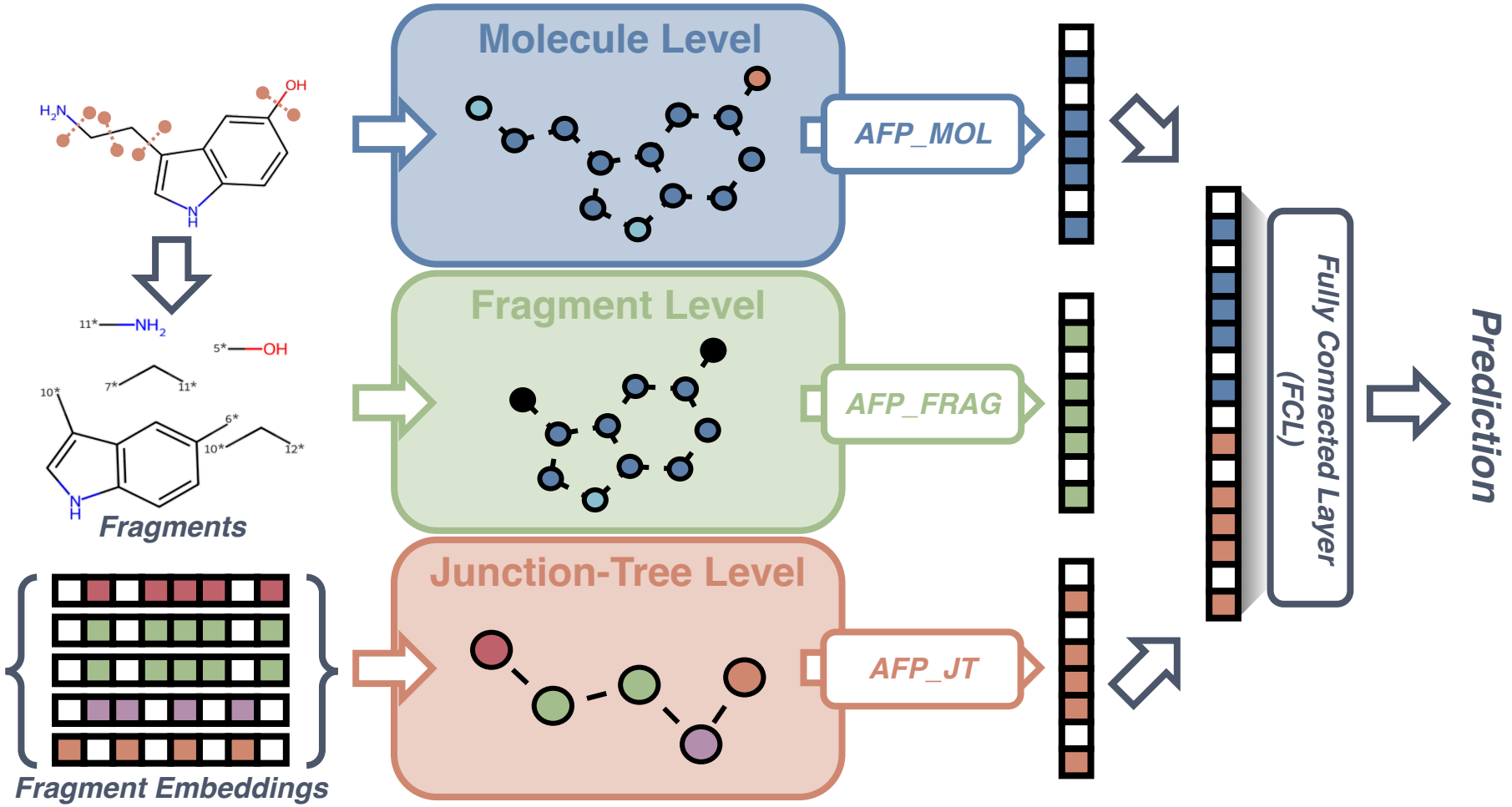


Molecule-level Graph



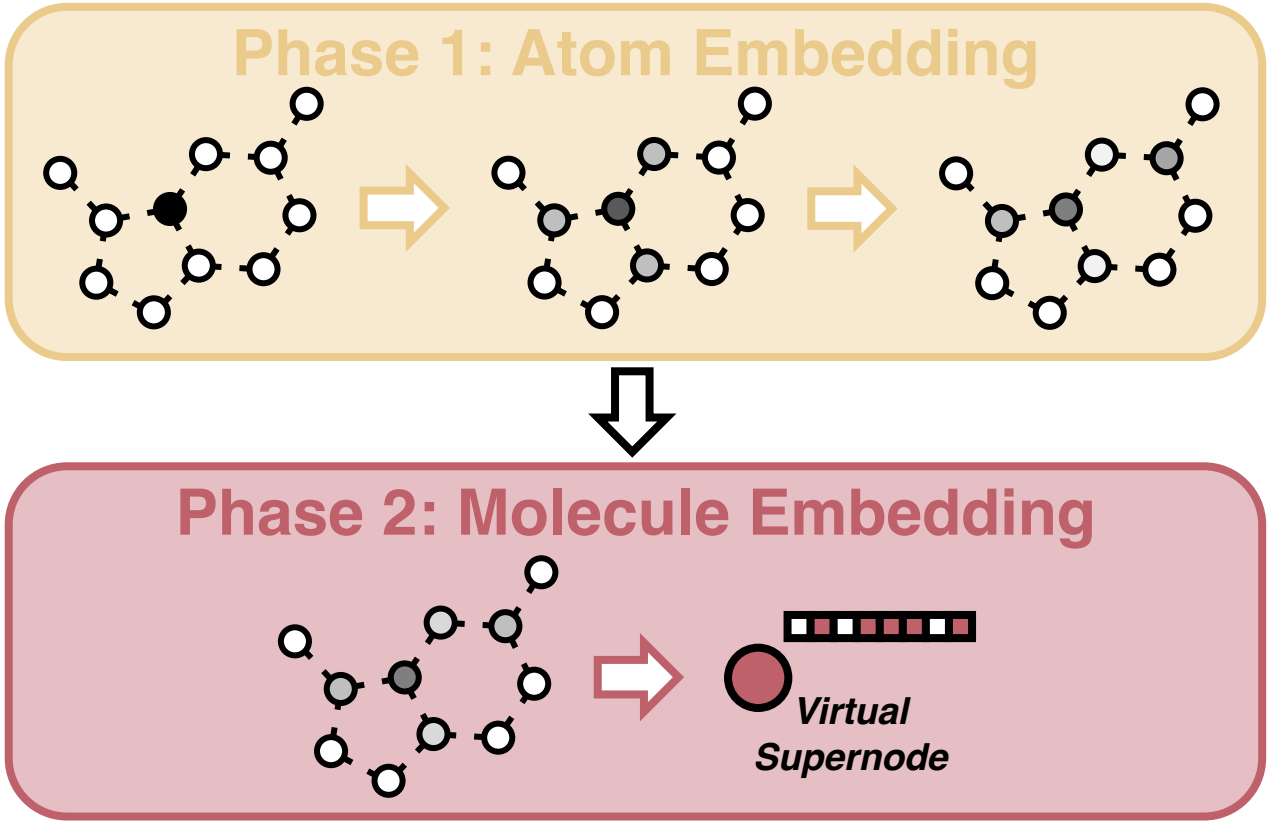
Fragment-level Graph

Chemomile Network



AttentiveFP

✓ AttentiveFP is based on Graph Attention Network (GAT)



- **Alignment:**

$$e_{vu} = leaky\_relu(W \cdot [h_v, h_u])$$

- **Weighting:**

$$a_{vu} = softmax(e_{vu})$$

- **Context Generation:**

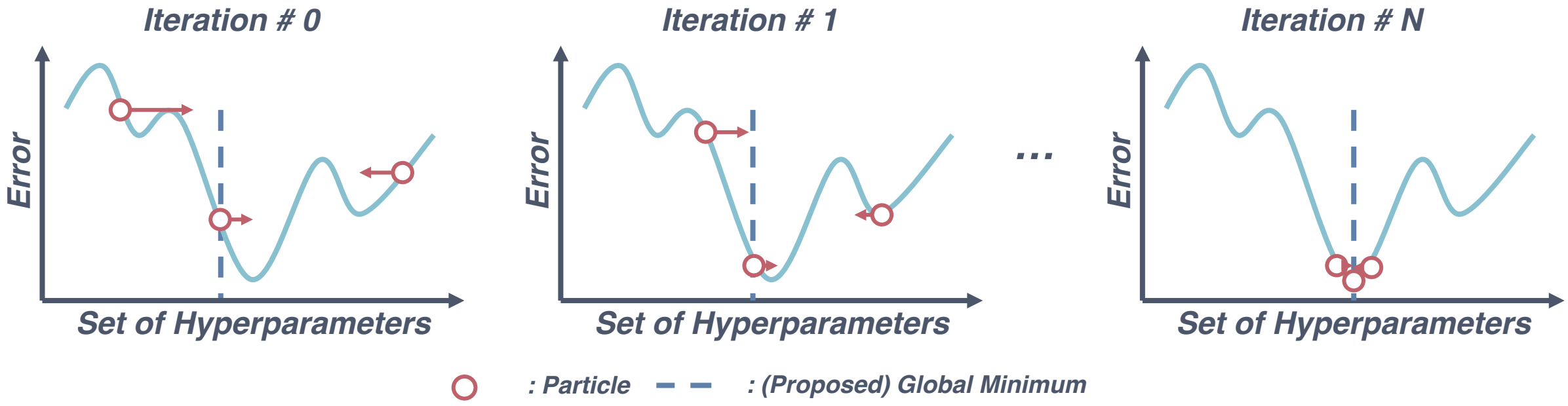
$$C_v = elu\left(\sum_{u \in N(v)} a_{vu} \cdot W \cdot h_u\right)$$

- **State Update:**

$$h_v^k = GRU^{k-1}(C_v^{k-1}, h_v^{k-1})$$

Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; others
Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism.
Journal of medicinal chemistry 2019, 63, 8749– 8760.

Particle Swarm Optimization (PSO)



- ✓ PSO mimics the social behavior of animals
- ✓ Each “agent” decides its next move considering **Personal Best**, **Social Best**, and **Inertia**

Miranda, L. J. V.
PySwarms, a research-toolkit for Particle Swarm Optimization in Python.
Journal of Open Source Software 2018, 3 .

Dataset

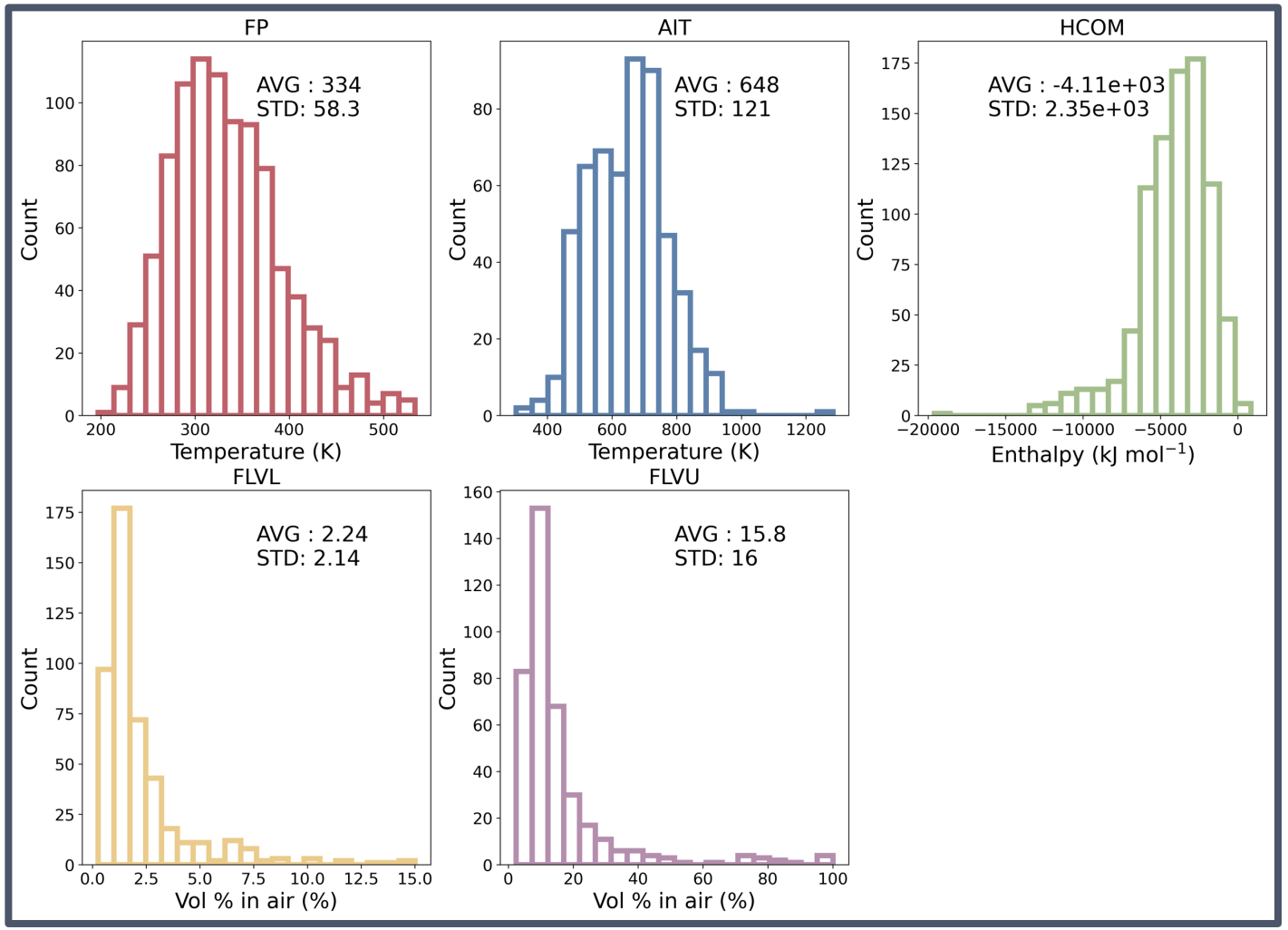
- ✓ DIPPR* 801 is managed by AIChE**
  * Design Institute for Physical Properties
  ** American Institute of Chemical Engineers

- ✓ Combustion Properties
  - Flash Point (FP)
  - Autoignition Temperature (AIT)
  - Heat of Combustion (HCOM)
  - Lower/Upper Flammability Limit (FLVL/FLVU)

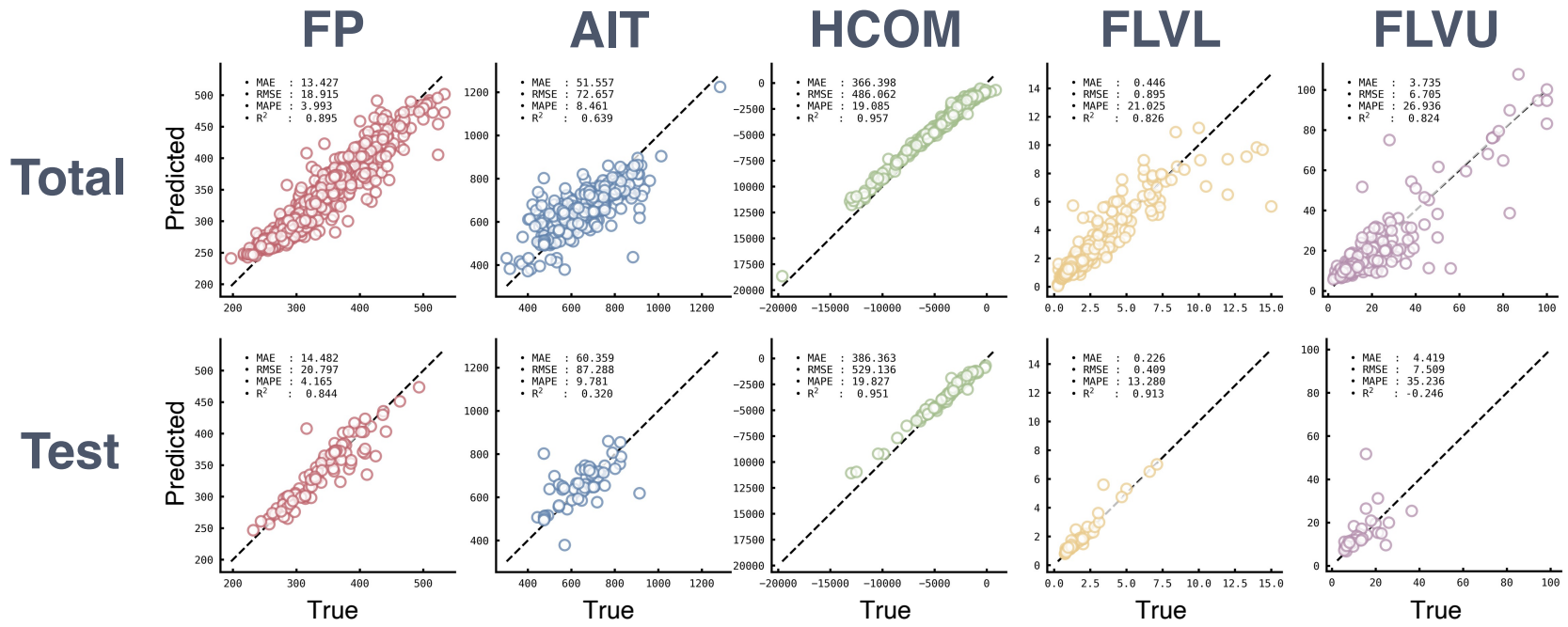
- ✓ Chemomile predicts Z-score

$$Z = \frac{X - \bar{X}}{s}$$

- ✓ Error Metrics
  - MAE : Mean Absolute Error
  - RMSE : Root Mean Squared Error
  - MAPE : Mean Absolute Percentage Error
  - R^2



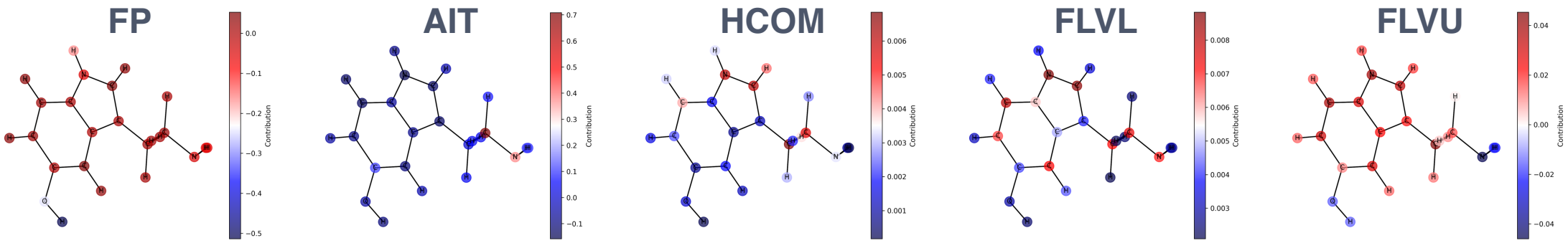
Results



Target	Total				Test			
	MAE	RMSE	MAPE	R^2	MAE	RMSE	MAPE	R^2
FP	13.427 (K)	18.915 (K)	3.993	0.895	14.482 (K)	20.797 (K)	4.165	0.844
AIT	51.557 (K)	72.567 (K)	8.461	0.639	60.359 (K)	87.288 (K)	9.781	0.320
HCOM	366.398 (kJ mol^{-1})	486.062 (kJ mol^{-1})	19.085	0.957	386.363 (kJ mol^{-1})	529.136 (kJ mol^{-1})	19.827	0.951
FLVL	0.446 (%)	0.895 (%)	21.025	0.826	0.226 (%)	0.409 (%)	13.280	0.913
FLVU	3.375 (%)	6.705 (%)	26.936	0.824	4.419 (%)	7.509 (%)	36.236	-0.246

Perturbation-based Explanation

$$Contrib.(\bigcirc) = Pred.(\quad) - Pred.(\quad)$$



- ✓ Atom-wise contribution can be quantified
- ✓ Different contribution for different properties
- ✓ Explanation relies on model's performance
- ✓ Explanation agrees with existing chemical knowledge (Functional Groups)

Conclusion

- ✓ Inaccurate prediction of combustion properties may lead to hazards
- ✓ A molecule can be represented as a graph form
- ✓ Chemomile utilizes multi-level graph representation combined with 3D geometry of a molecule
- ✓ Chemomile shows compatible performance to existing GNN models
- ✓ Atom-wise contribution can be quantified, providing an explanation of predicted value
- ✓ This knowledge can be a valuable tool for researchers and engineers

Beomgyu Kang and Bong June Sung
Chemomile: An Explainable Multi-Level GNN Model for Combustion Property Prediction
(Manuscript in preparation)