Molecular Property Prediction: A Multilevel Quantum Interactions Modeling Perspective

Chengqiang Lu†, Qi Liu†*, Chao Wang†, Zhenya Huang†, Peize Lin‡, Lixin He‡

†Anhui Province Key Lab. of Big Data Analysis and Application, University of S&T of China

‡China Key Laboratory of Quantum Information, University of S&T of China

AAAI 2019

CONTENTS

01 Introduction

02 Related Work

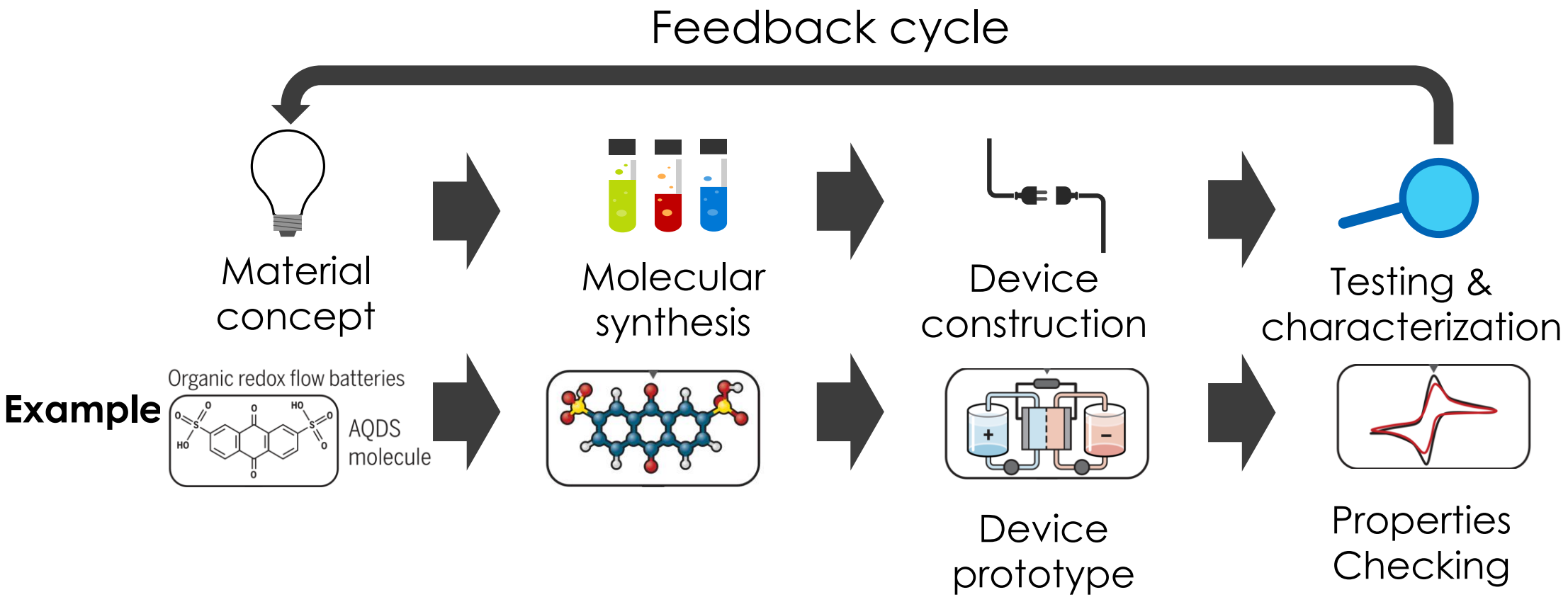
03 MGCN

04 Experiment

01

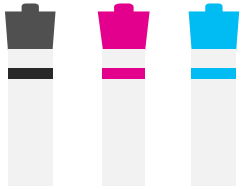
Introduction

Material Discovery Paradigms



Science 2018. Sanchez-Lengeling, et al. "Inverse molecular design using machine learning: Generative models for matter engineering."

Application



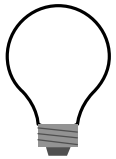
Material
Discovery

Medicine
Design

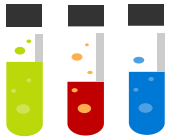
Food
Development

01 Introduction

The Most Time-consuming Step



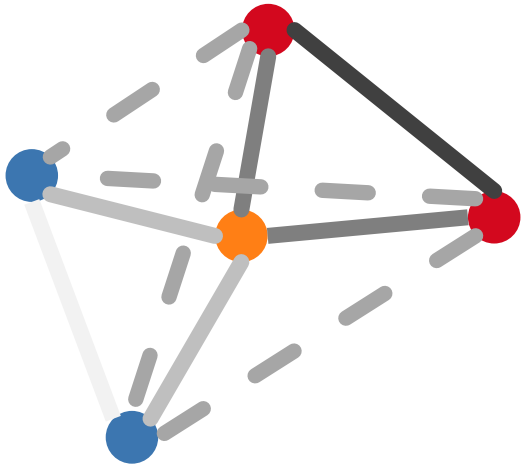
Material concept



Molecular synthesis

To find the molecule with desired properties.
We need explore the molecule database (e.g. gdb-17),
and **predict molecular properties.**

Our Task



Input
(molecule)

Properties:

U0 (Atomization energy at 0K)

U (Atomization energy at room temperature)

H (Enthalpy at room temperature)

G (Free energy of atomization)

.
.
.

Output
(properties)

J. Chem. Inf. Model. 2012. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. Ruddigkeit Lars, van Deursen Ruud, Blum L. C.; Reymond J.-L.

Challenge:

- Molecular quantum interactions are **highly complex** and hard to model.
- The amount of labeled molecule data is significantly limited, which requires a **generalizable** approach for the prediction.
- The molecule data is unbalanced: most of the molecules are small and few of them are large, thus the model should be **transferable**.

02

Related Work

DFT (Density Functional Theory)

- Classic physical methods which could date back to 1960s.
- States that the quantum interactions between particles (e.g., atoms) create the correlation and entanglement of molecules which are closely related to their inherent properties
- Pros:
  - Accurate
  - Widely used
- Cons:
  - Extremely time consuming

- Journal of Physics. 2014. Behler, Jörg. "Representing potential energy surfaces by high-dimensional neural network potentials."
- Journal of Chemical Physics. 2017. Cubuk, Ekin D., et al. "Representations in neural network based empirical potentials."

Traditional ML models

Representations:

- BOB (bag of bonds)
- Coulomb matrix
- HDAD (histogram of distance, angle and dihedral angle)
- .
- .
- .

Models:

- Kernel ridge regression
- Random forest
- Elastic Net
- .
- .
- .

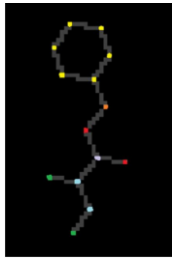
Cons:

- Hand crafted features need much domain expertise
- Be restricted in practice

Journal of chemical theory and computation. 2013. Faber, F. A.; Hutchison, L.; Huang, B.; Gilmer, J.; Schoenholz, S. S.; Dahl, G. E.; Vinyals, O.; Kearnes, S.; Riley, P. F.; and von Lilienfeld, O. A. 2017. Prediction errors of molecular machine learning models lower than hybrid dft error.

Deep Neural Networks I

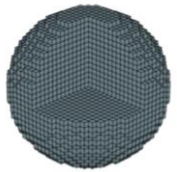
Use grid-like data as input



1. Images

C\C=C(/C)C
(=O)OCC1=
CC=CC=C1

2. Text



3. Sphere

- Could utilize the models in CV/NLP
- Initiative grid-like transformation usually caused information loss

1. KDD'18. ChemNet: A Transferable and Generalizable Deep Neural Network for Small-Molecule Property Prediction
2. ACS'18. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules
3. NIPS'17. Spherical convolutions and their application in molecular modelling

Deep Neural Networks II

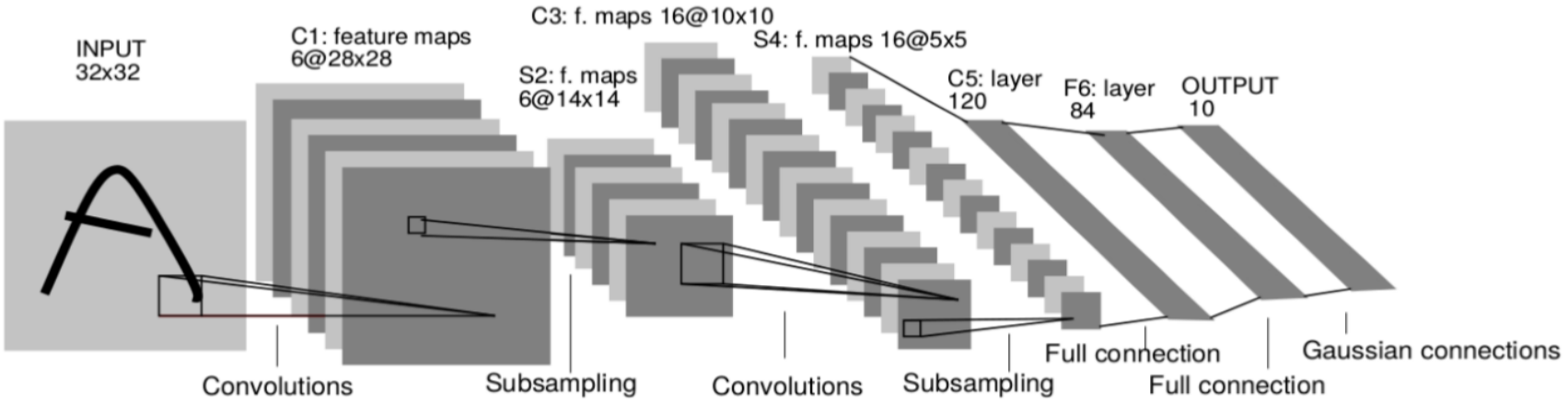
Use graph-like data as input

- Deep Tensor Neural Network
- Sch Net
- Message Passing Neural Network

- Implement the conv-operator in graph
- Achieve some superior experimental results
- Have not utilize the multilevel property
- Bad generalizability and transferability

- Nature Comm'17. Quantum-chemical insights from deep tensor neural networks
- NIPS'17 SchNet: A continuous-filter convolutional neural network for modeling quantum interactions
- ICML'17 Neural Message Passing for Quantum Chemistry

02 Key properties of CNNs



- ☺ Convolutional **(Translation invariance)**
- ☺ Scale Separation **(Compositionality)**
- ☺ Filters localized in space **(Deformation Stability)**
- ☺ $\mathcal{O}(1)$ parameters per filter (independent of input image size n)
- ☺ $\mathcal{O}(n)$ complexity per layer (filtering done in the spatial domain)
- ☺ $\mathcal{O}(\log n)$ layers in classification tasks

LeCun et al. 1989

Multilevel Graph Convolutional Network (MGCN)

03

Potential Energy Surfaces

$$E = \sum_i E_i = \sum_i \sum_{j \neq i} E_{ij}$$

- Behler, Jörg. "Representing potential energy surfaces by high-dimensional neural network potentials." *Journal of Physics: Condensed Matter* 26.18 (2014): 183001.
- Cubuk, Ekin D., et al. "Representations in neural network based empirical potentials." *The Journal of Chemical Physics* 147.2 (2017): 024104.

Atom-centered symmetry functions

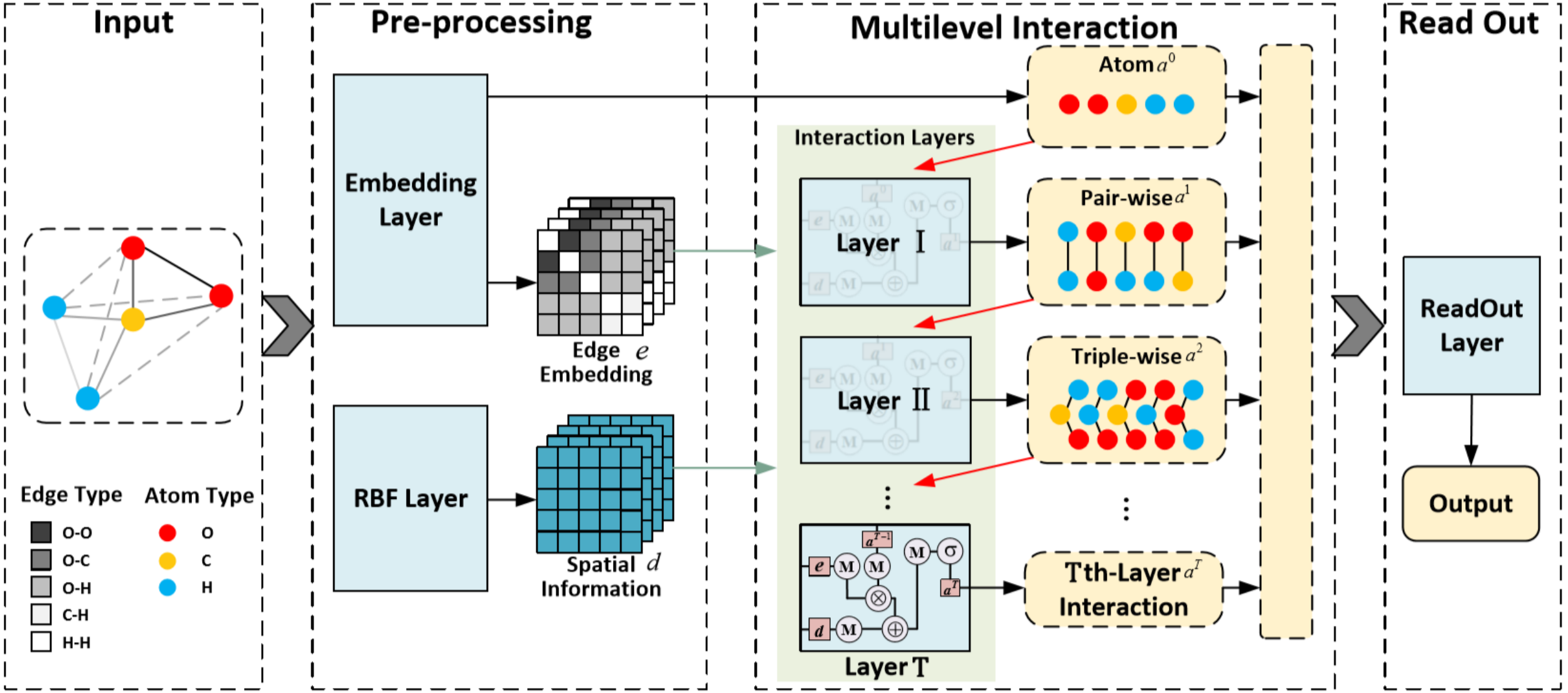
$$G_i^1 = \sum_{j=1}^{N_{\text{atom}}} f_c\left(R_{ij}\right)$$

$$G_i^2 = \sum_{j=1}^{N_{\text{atom}}} \mathrm{e}^{-\eta\left(R_{ij}-R_{\mathrm{s}}\right)^2} \cdot f_{\mathrm{c}}\left(R_{ij}\right)$$

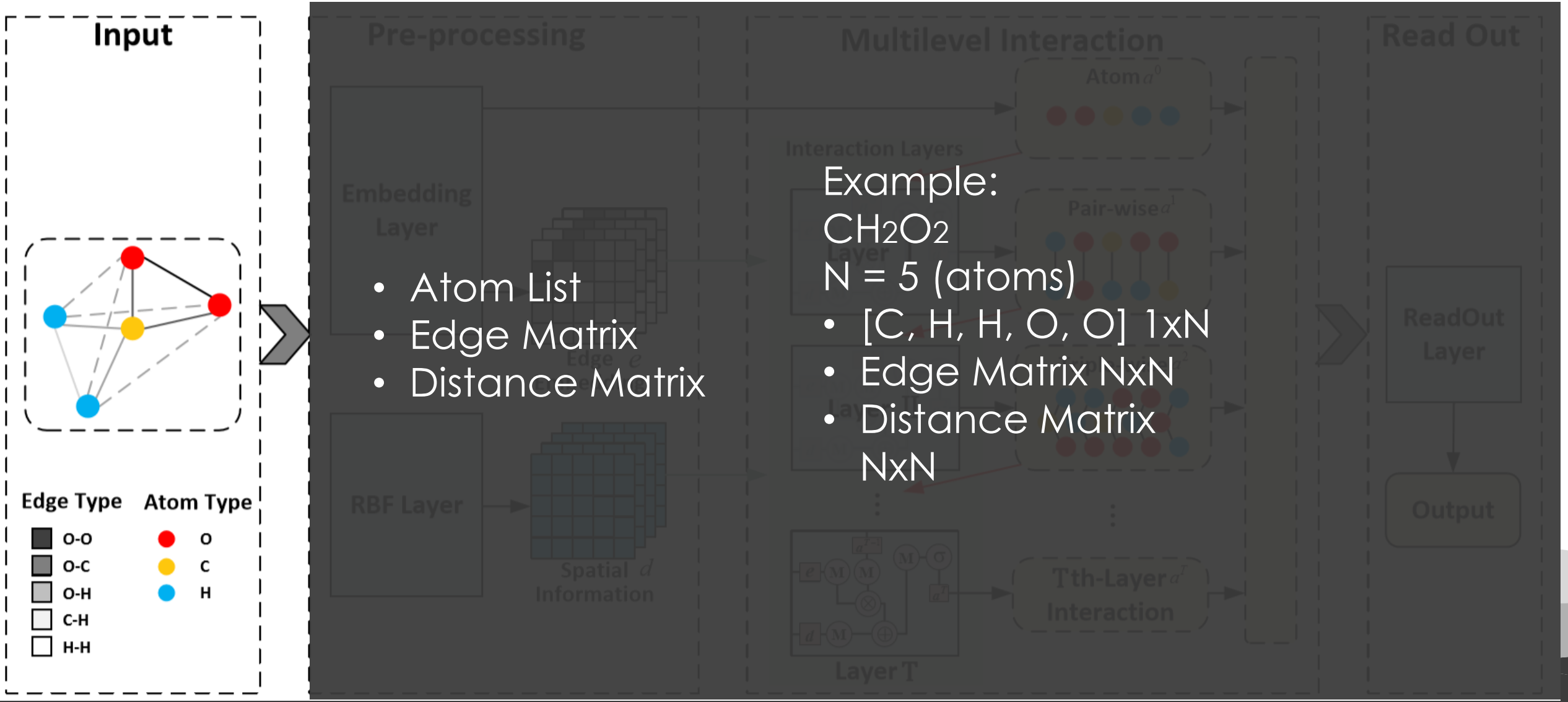
$$G_i^4 = 2^{1-\zeta} \sum_{j\neq i} \sum_{k\neq i,j} \left[\left(1+\lambda\cdot\cos\theta_{ijk}\right)^{\zeta} \times \mathrm{e}^{-\eta\left(R_{ij}^2+R_{ik}^2+R_{jk}^2\right)} \cdot f_{\mathrm{c}}\left(R_{ij}\right) \cdot f_{\mathrm{c}}\left(R_{ik}\right) \cdot f_{\mathrm{c}}\left(R_{jk}\right) \right]$$

$$G_i^5 = 2^{1-\zeta} \sum_{j\neq i} \sum_{k\neq i,j} \left[\left(1+\lambda\cdot\cos\theta_{ijk}\right)^{\zeta} \cdot \mathrm{e}^{-\eta\left(R_{ij}^2+R_{ik}^2\right)} \cdot f_{\mathrm{c}}\left(R_{ij}\right) \cdot f_{\mathrm{c}}\left(R_{ik}\right) \right]$$

Overview


Input
Pre-processing
Multilevel Interaction
Read Out
Embedding Layer
RBF Layer
Edge e Embedding
Spatial d Information
Interaction Layers
Layer I
Layer II
Layer T
Atom a^0
Pair-wise a^1
Triple-wise a^2
Tth-Layer a^T Interaction
ReadOut Layer
Output
Edge Type
O-O
O-C
O-H
C-H
H-H
Atom Type
O
C
H

Input

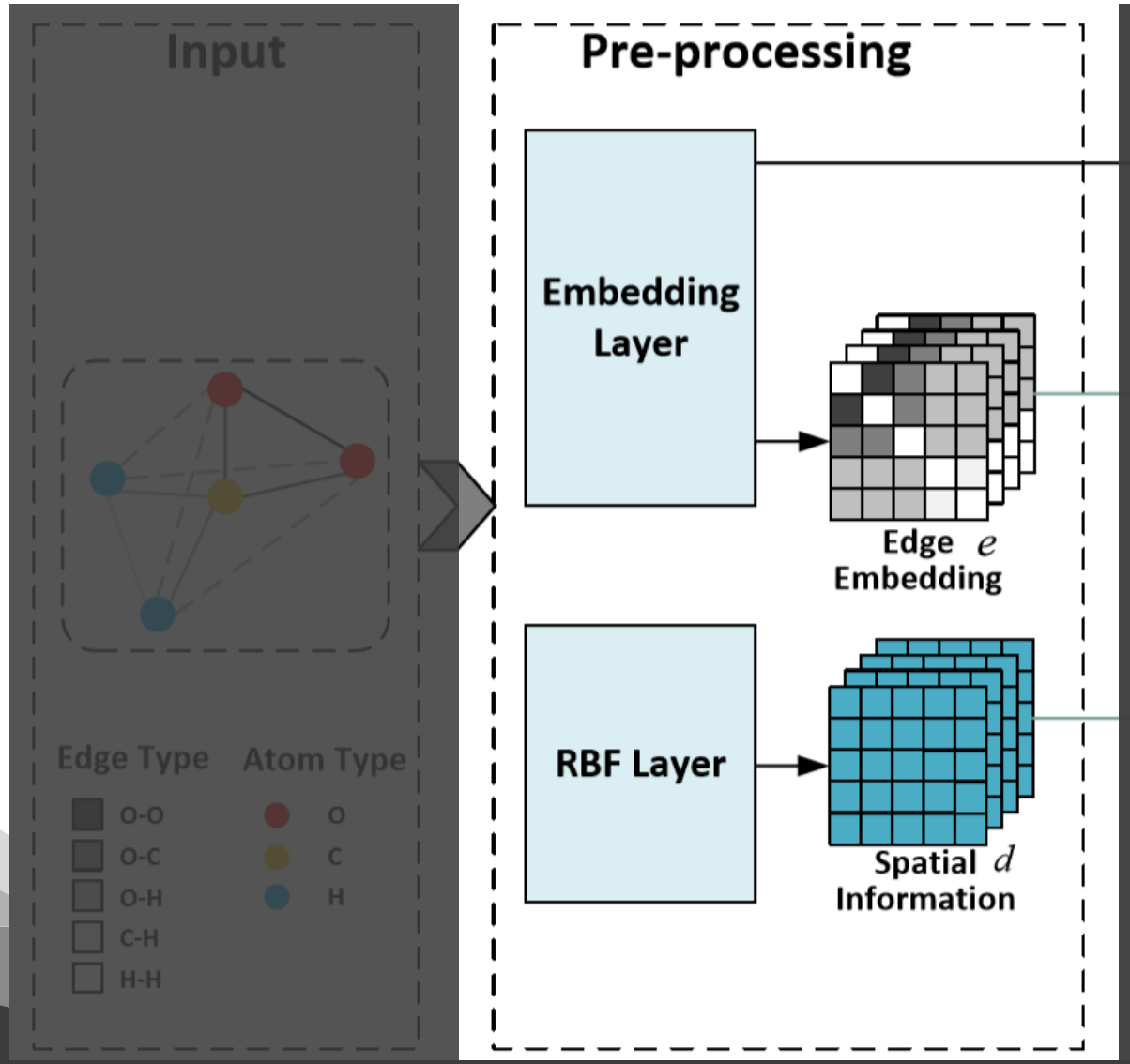


- Atom List
- Edge Matrix
- Distance Matrix

Example:
CH_2O_2
N = 5 (atoms)

- [C, H, H, O, O] 1xN
- Edge Matrix NxN
- Distance Matrix NxN

Pre-processing



Embedding Layer: generate initial representation of edges and atom.

- Atom embedding: A^0 $N \times K$
- Edge embedding: E $N \times N \times K$

Radial Basis Function Layer: convert distance matrix to robust distance tensors

$$RBF(x) = \overset{K}{\underset{i=1}{\frown}} \mathrm{h}(\|x - \mu_i\|)$$

- h - RBF function
- D $N \times N \times K$

Interaction Layers

In each interaction layer: model will generate the atomic representations at **higher level**

$$\boldsymbol{a}_i^{l+1} = \sum_{j=1, j\neq i}^{N} \mathrm{h}_v(\boldsymbol{a}_j^l, \boldsymbol{e}_{ij}^l, \boldsymbol{d}_{ij})$$

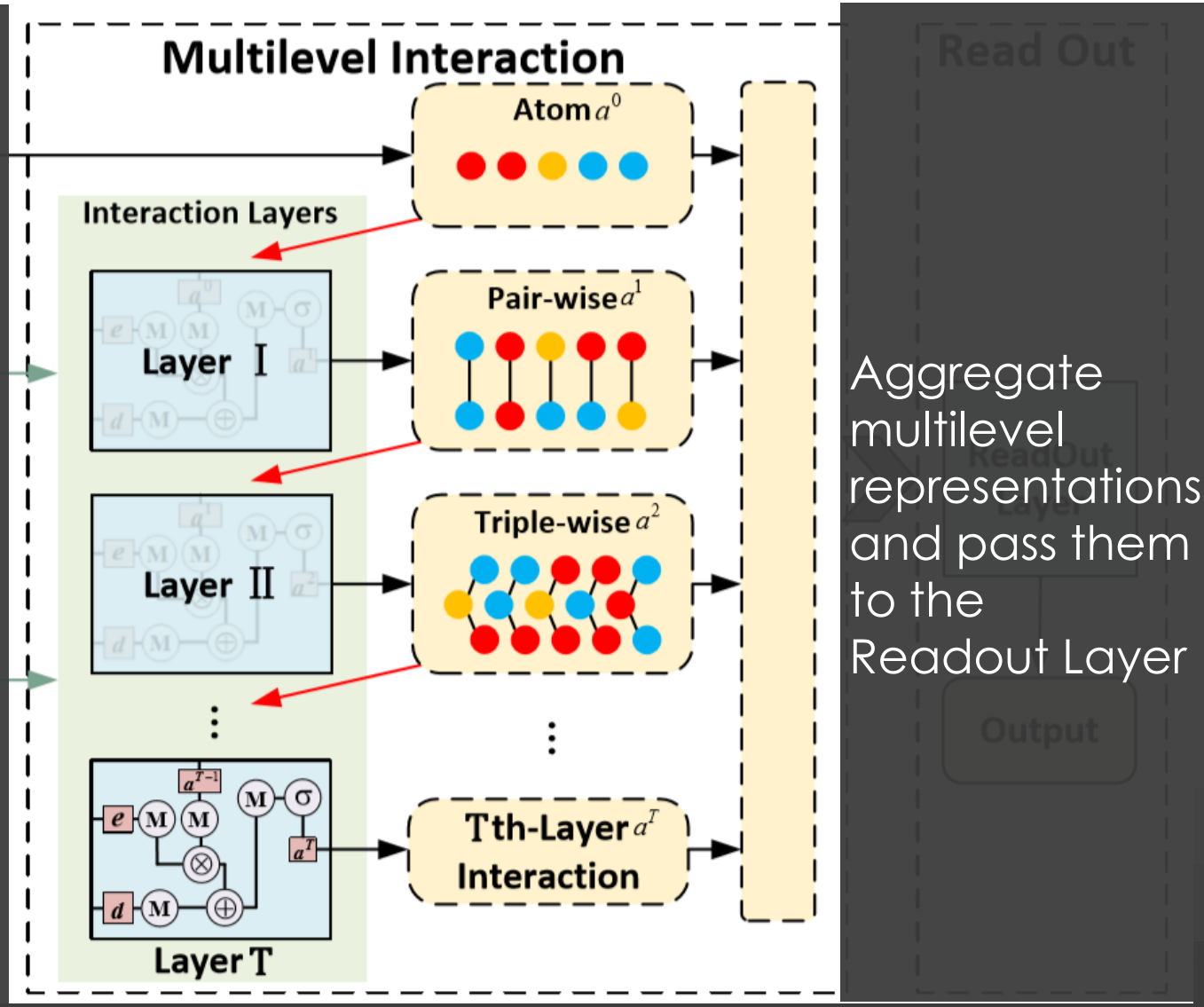
and update the edge representation:

$$\boldsymbol{e}_{ij}^{l+1} = \mathrm{h}_e(\boldsymbol{a}_i^l, \boldsymbol{a}_j^l, \boldsymbol{e}_{ij}^l)$$

In detail:

$$\mathrm{h}_v = \sigma(W^{uv}(\mathrm{M}^{fa}(\boldsymbol{a}_j^l) \odot \mathrm{M}^{fd}(\boldsymbol{d}_{ij}) \oplus \mathrm{M}^{fe}(\boldsymbol{e}_{ij}))$$

$$\mathrm{h}_e = \eta \boldsymbol{e}_{ij}^l \oplus (1-\eta) W^{ue} \boldsymbol{a}_i^l \odot \boldsymbol{a}_j^l$$

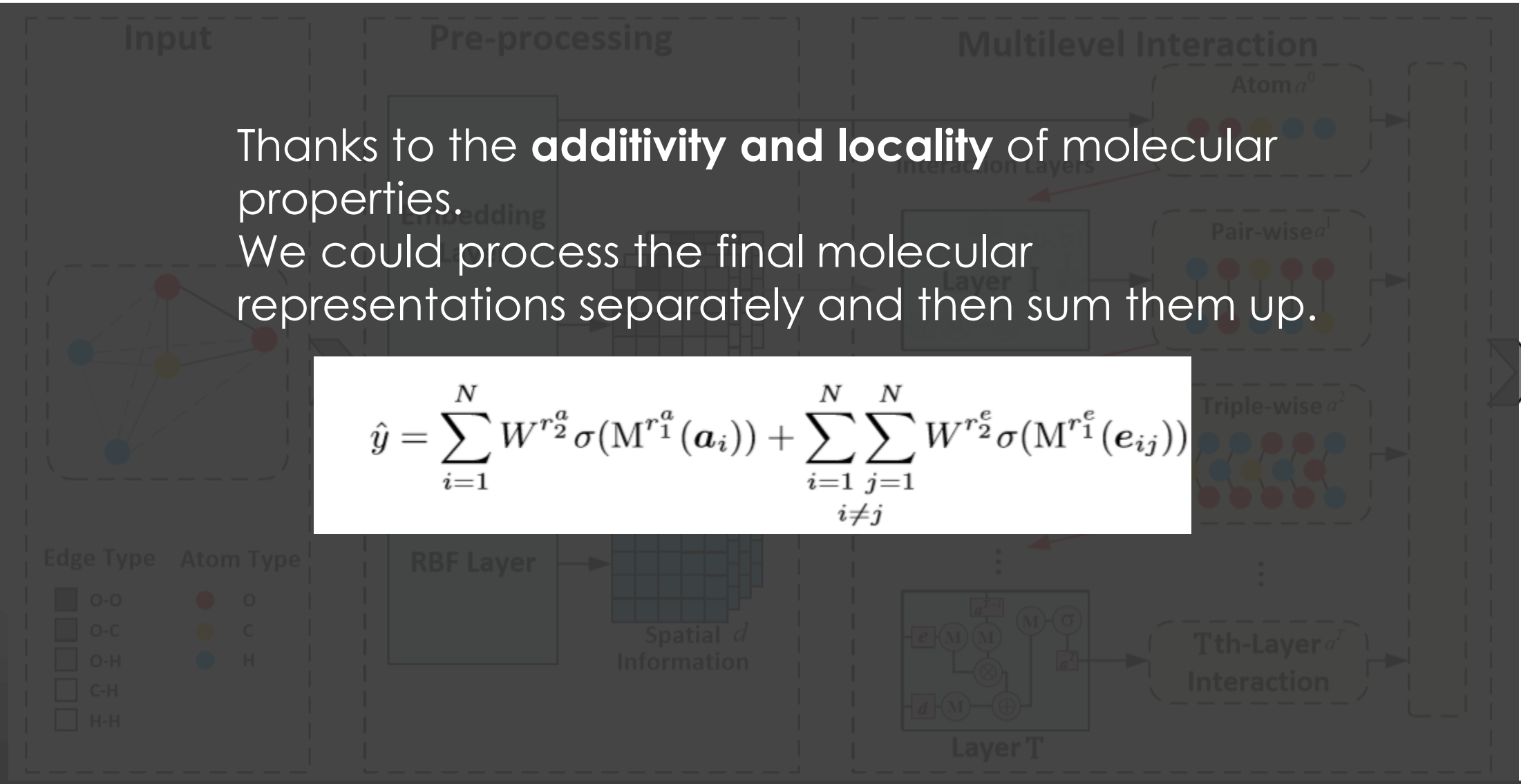


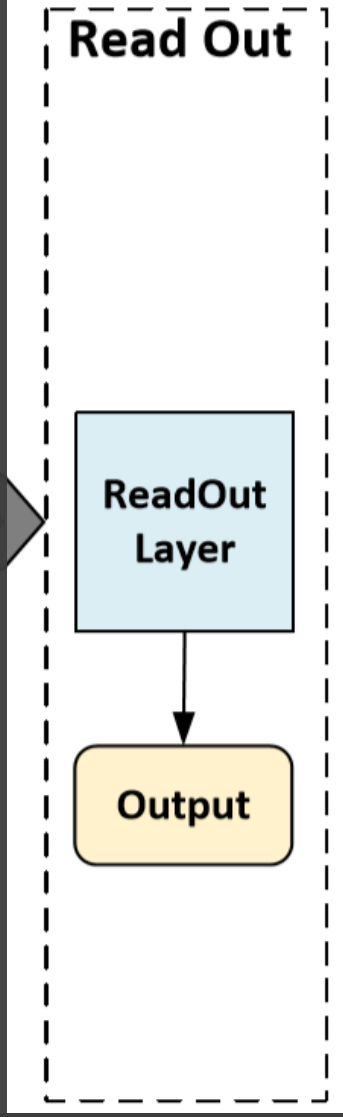
Aggregate multilevel representations and pass them to the Readout Layer

Read Out Layer

Thanks to the **additivity and locality** of molecular properties.
We could process the final molecular representations separately and then sum them up.

$$\hat{y} = \sum_{i=1}^{N} W^{r_2^a} \sigma(\mathrm{M}^{r_1^a}(\boldsymbol{a}_i)) + \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} W^{r_2^e} \sigma(\mathrm{M}^{r_1^e}(\boldsymbol{e}_{ij}))$$





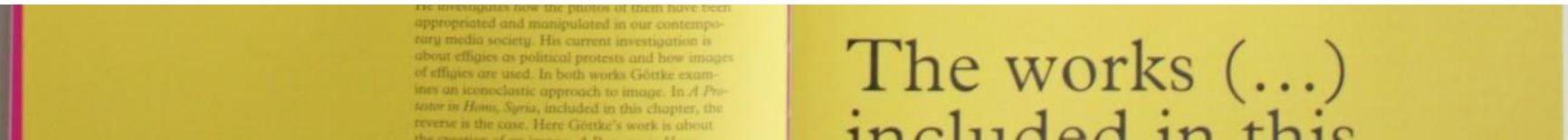
Discussion

Generalizability:

- Coordinates -> Distance tensor: translation rotation invariance.
- Element-wise operations: index invariance.
- Drop-out.

Transferability:

- First-level knowledge are structure/spatial-irrelevanted.
- Pre-trained embedding.



04

Experiment

Data sets

QM9

- Most well-known data set
- Contains 134k stable molecules
- 13 different properties

ANI-1

- Contains **20 million** unstable molecules
- Only one property

Table 1: Predictive accuracy of different models in QM9

Properties	U_0	U	G	H	C_v	$\varepsilon_{\text{HOMO}}$	$\varepsilon_{\text{LUMO}}$	$\Delta\varepsilon$	ω_1	ZPVE	$\langle R^2 \rangle$	μ	α
Unit	eV	eV	eV	eV	cal/molK	eV	eV	eV	cm^{-1}	eV	Bohr^2	Debye	Bohr^3
DFT Error	0.1	0.1	0.1	0.1	0.34	-	-	-	28	0.0097	-	0.1	0.4
Chemical Acc.	0.043	0.043	0.043	0.043	0.05	0.043	0.043	0.043	10	0.00122	1.2	0.1	0.1
RF+BAML	0.2000	-	-	-	0.451	0.1070	0.1180	0.1410	2.71	0.01320	51.10	0.434	0.638
KRR+BOB	0.0667	-	-	-	0.092	0.0948	0.1220	0.1480	13.20	0.00364	0.98	0.423	0.298
KRR+HDAD	0.0251	-	-	-	0.044	0.0662	0.0842	0.1070	23.10	0.00191	1.62	0.334	0.175
GG	0.0421	-	-	-	0.084	0.0567	0.0628	0.0877	6.22	0.00431	6.30	0.247	0.161
enn-s2s	0.0194	0.0194	0.0168	0.0189	0.040	0.0426	0.0374	0.0688	1.90	0.00152	0.18	**0.030**	0.092
DTNN	0.0364	0.0377	0.0385	0.0357	0.089	0.0982	0.1053	0.1502	4.23	0.00312	0.30	0.257	0.131
SchNet	0.0134	0.0189	0.0196	0.0182	0.067	0.0507	**0.0372**	0.0795	3.83	0.00172	0.27	0.071	0.073
MGCN	**0.0129**	**0.0144**	**0.0146**	**0.0162**	**0.038**	**0.0421**	0.0574	**0.0642**	**1.67**	**0.00112**	**0.11**	0.056	**0.030**

Table 2: Predictive accuracy of different models in ANI-1

Methods	DTNN	SchNet	MGCN
MAE	0.113	0.108	**0.078**

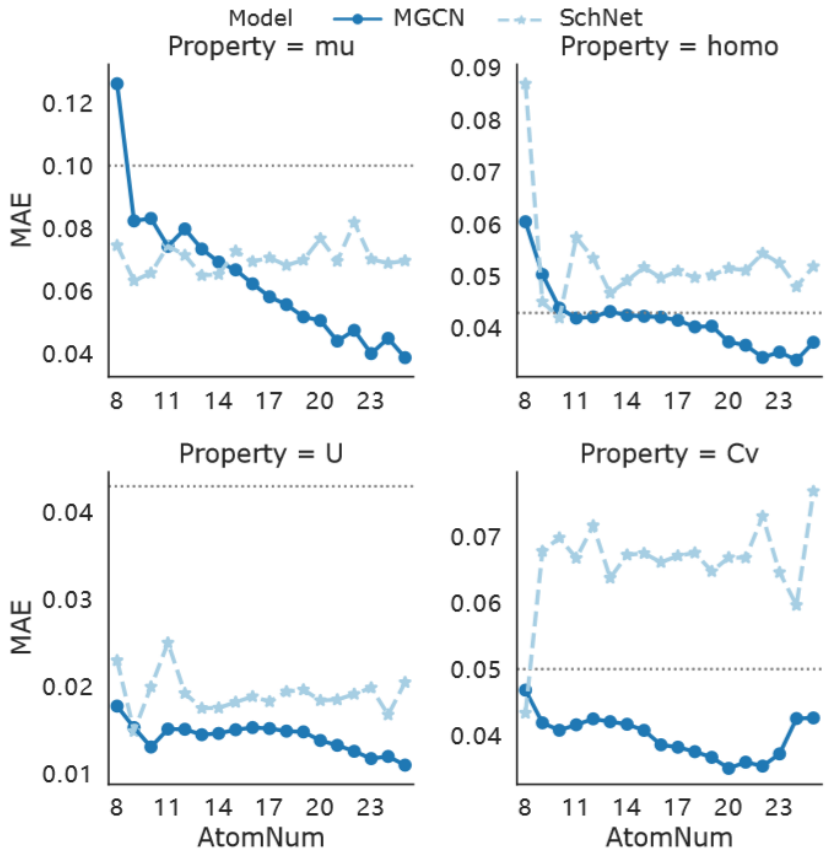


Figure 3: MAE of prediction in different size molecules.

Table 3: Performance comparison in varied size training set

N	SchNet	DTNN	enn-s2s	MGCN
50,000	0.0256	0.0408	0.0249	**0.0229**
100,000	0.0147	0.0364	-	**0.0142**
110,462	0.0134	-	0.0194	**0.0129**

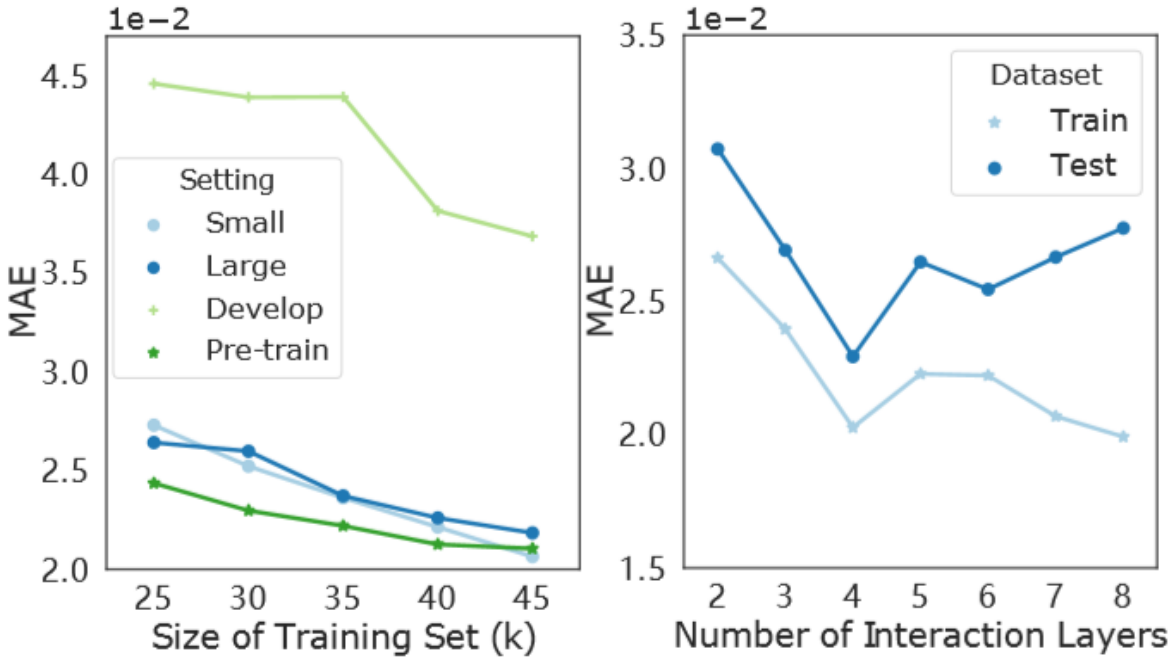
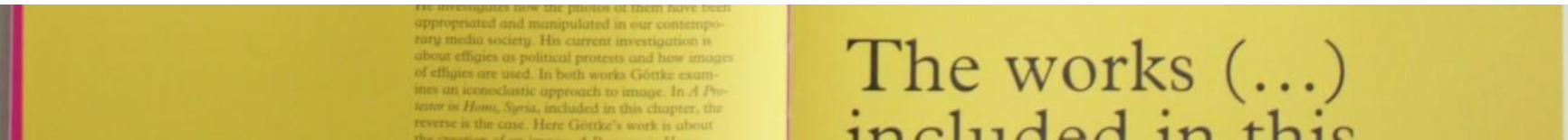


Figure 4: *a*(left). Performance comparison in the training set with different size. *b*(right). Predictive performance of models with different number of interaction layers.

Conclusion

- Propose a well designed Multilevel Convolutional Neural Network (MGCN) for predicting molecular properties.
- Model the quantum Interaction from a multilevel view using molecular graph as input.
- MGCN model is transferable and generalizable.



Thanks for listening.