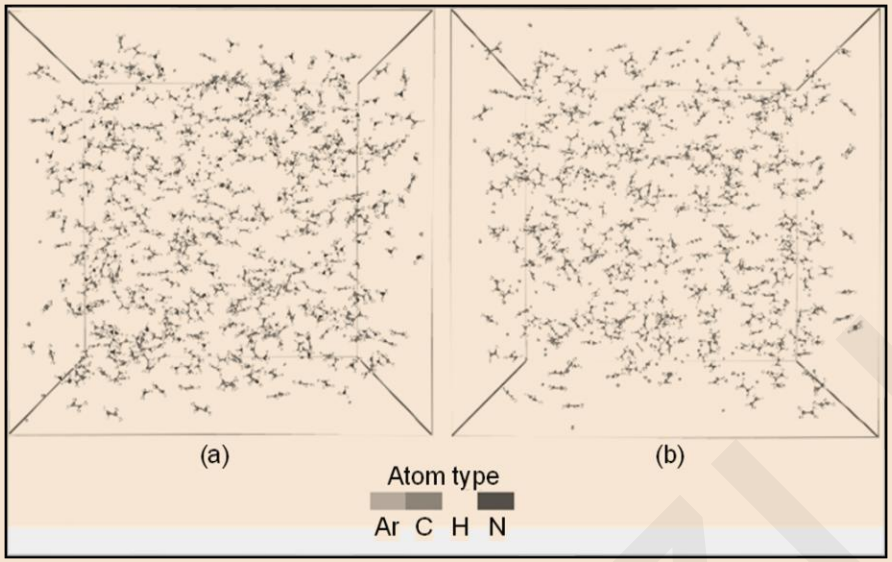
Journal of Zhejiang University-SCIENCE A

Data-driven neural surrogates for ReaxFF molecular dynamics simulations in engine-relevant combustion chemistry

Yuchao YAN, Qiao HUANG, Tianfang XIE, Jinlong LIU

Cite this as: Yuchao YAN, Qiao HUANG, Tianfang XIE, Jinlong LIU, 2026. Data-driven neural surrogates for ReaxFF molecular dynamics simulations in engine-relevant combustion chemistry. *Journal of Zhejiang University-SCIENCE A*, 27(8):825-836. https://doi.org/10.1631/jzus.A2500457

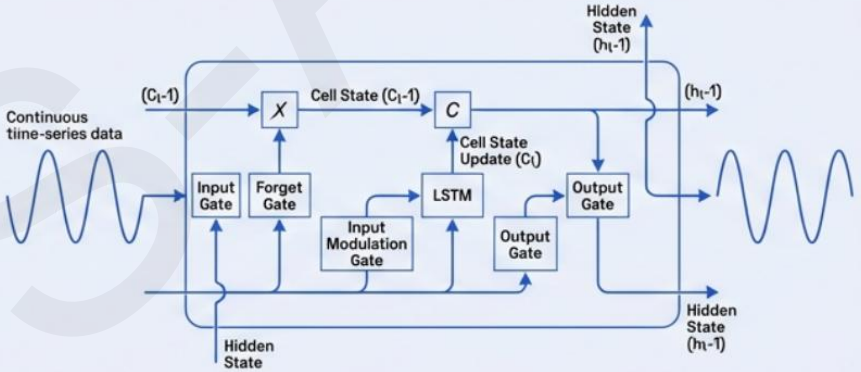
The Bottleneck: ReaxFF MD Simulations



ReaxFF MD

- Computationally Expensive
- A Workflow Bottleneck
- Essential but Time-Consuming

The Innovation: Data-Driven Surrogates

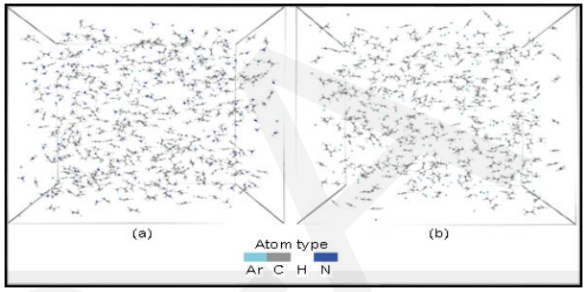


Machine Learning

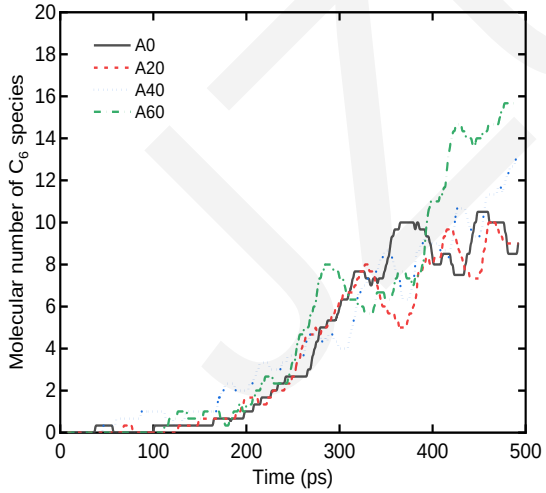
- Reduces Computational Cost
- Broad Applicability Across Engine Systems

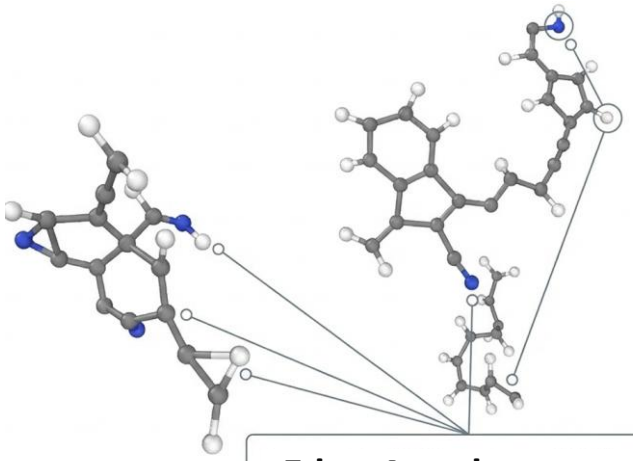
The Testing Ground: C_2H_4/NH_3 pyrolysis

Prediction Target
Temporal Evolution of C_6 species



DATASET GENERATED BY ReaxFF MD			
[A0]	[A20]	[A40]	[A60]
0% NH_3	20% NH_3	40% NH_3	60% NH_3
Training	Training	Test	Training



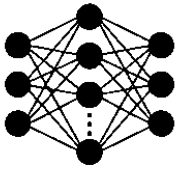
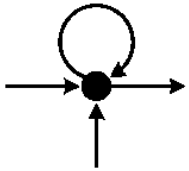
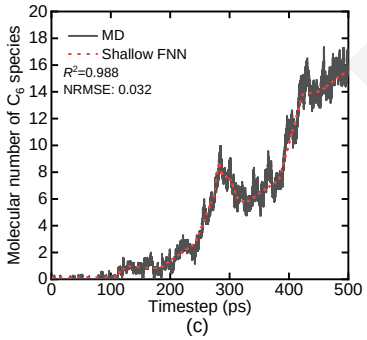
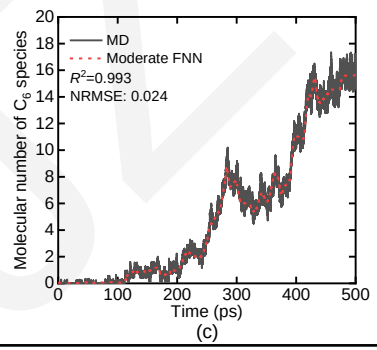
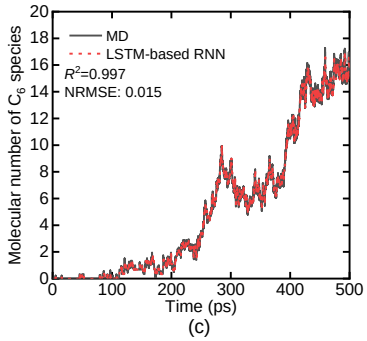


Edge-Attachment Effects of N-PAHs

***C_6* Evolution: Nonlinear Kinetics as the ML Target**

Machine Learning Models

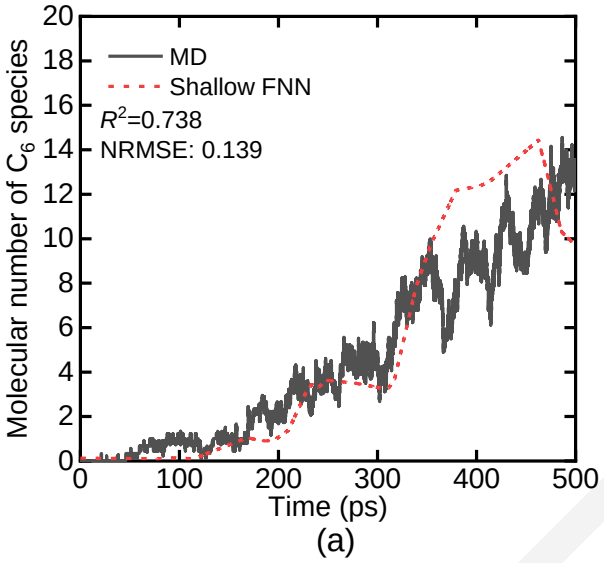
Task: Predict the exact number of C_6 molecules over a 500-ps trajectory.

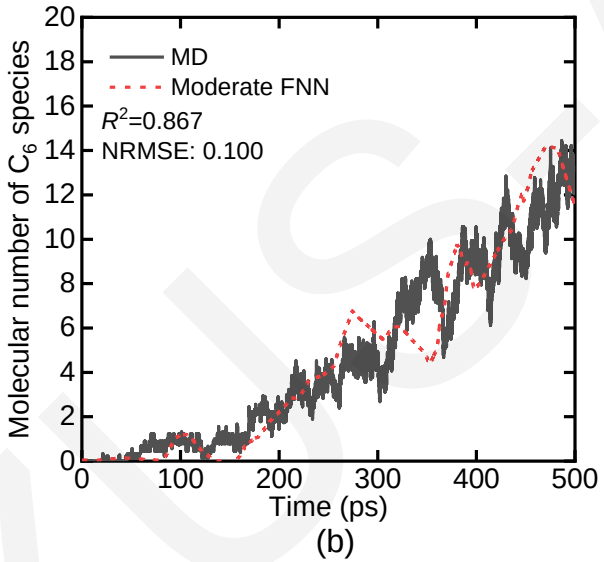
Shallow FNN	Moderate FNN	LSTM-based RNN
1 Hidden Layer 64 Neurons per Layer	3 Hidden Layers 64 Neurons per Layer	1 LSTM Layer 50 Neurons + Gates
Inputs Time and Ammonia Ratio		**Inputs** Time-Series Data
Memory State None		**Memory State** Gated Internal Memory
 **Processes each input [t, r] independently**		 **Utilizes input, forget, and output gates to regulate information flow**
 		

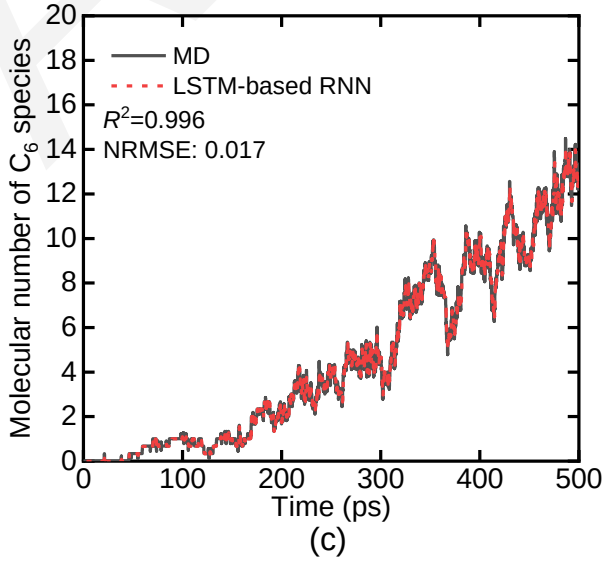
- **All models perform well on the training set, but generalization differs across architectures**

Machine Learning Models

Generalization Performance on the Test Dataset







Shallow FNN
R^2=0.738
Suboptimal generalization

Moderate FNN
R^2=0.867
Depth alone cannot address the lack of memory

LSTM
R^2=0.996
Highly accurate with strong generalization

Conclusions

• **Validated MD Data:** The MD simulations accurately capture the inhibitory effect of NH_3 on PAH formation, providing a robust modeling foundation.

• **FNN Shortfalls:** A memoryless architecture leads to poor generalization in history-dependent dynamics.

• **LSTM Superiority:** Gated memory captures the reaction history, ensuring high accuracy on unseen data.

• **Broad Impact:** Sequence-aware surrogates enable faster workflows in engine combustion research.

