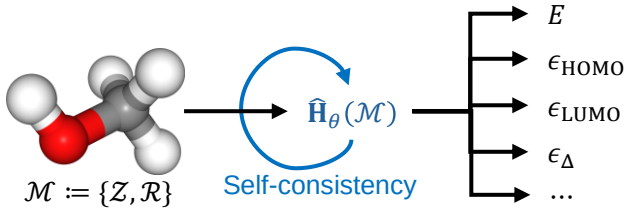
Self-Consistency Training for Density-Functional-Theory Hamiltonian Prediction

He Zhang[1,2], Chang Liu[2], Zun Wang[2], Xinran Wei[2], Siyuan Liu[2], Nanning Zheng[1], Bin Shao[2], Tie-Yan Liu[2]

[1]Xi'an Jiaotong University, [2]Microsoft Research AI for Science ✉ changliu@microsoft.com

MOTIVATION



Hamiltonian prediction can provide all that DFT can. Yet, its applicability is limited by scarce labeled data.

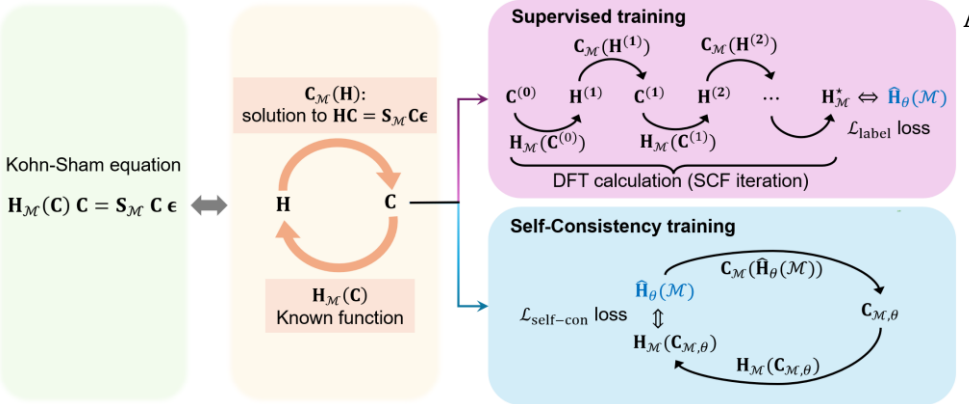
→ **We highlight it can be done variationally by self-consistency training:**

- **Label-free training**: Addressing data scarcity and enhancing generalization
- **Better efficiency** than DFT labeling for supervised training, due to amortization.

THE FRAMEWORK OF SELF-CONSISTENCY TRAINING

Density Functional Theory:

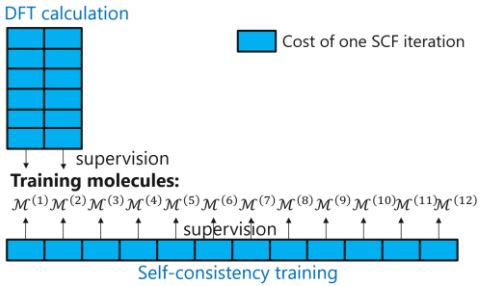
- Describe N-electron state by orbitals $\mathbf{C}$.
- Solve for $\mathbf{C}$ for a given molecular structure $\mathcal{M}$ by:



Self-Consistency Training:

$$\mathcal{L}_{\text{self-con}}(\theta) = \sum_\mathcal{M} \left\| \hat{\mathbf{H}}_\theta(\mathcal{M}) - \mathbf{H}_\mathcal{M}\left(\mathbf{C}_\mathcal{M}\left(\hat{\mathbf{H}}_\theta(\mathcal{M})\right)\right)\right\|_{\text{F}}^2$$

Amortization of SCF Iteration



Given the same computational budget, self-consistency training provides supervision on **more molecular structures**.

EXPERIMENTS

- Generalization improvement in the ***data-scare*** scenario (MD17)

Molecule	Setting	$\mathbf{H}$ [μE_h] ↓	$\boldsymbol{\epsilon}$ [μE_h] ↓	$\mathbf{C}$ [%] ↑	ϵ_{HOMO} [μE_h] ↓	ϵ_{LUMO} [μE_h] ↓	ϵ_Δ [μE_h] ↓	SCF Accel. [%] ↓
Ethanol	label	160.36	712.54	99.44	911.64	6800.84	6643.11	68.3
	label + self-con	**75.65**	**285.49**	**99.94**	**336.97**	**1203.60**	**1224.86**	**61.5**
Malondi-aldehyde	label	101.19	456.75	99.09	471.92	1093.22	1115.94	69.1
	label + self-con	**86.60**	**280.39**	**99.67**	**274.45**	**279.14**	**324.37**	**62.1**
Uracil	label	88.26	1079.51	95.83	1217.17	12496.1	11850.56	65.8
	label + self-con	**63.82**	**315.40**	**99.58**	**359.98**	**369.67**	**388.30**	**54.5**

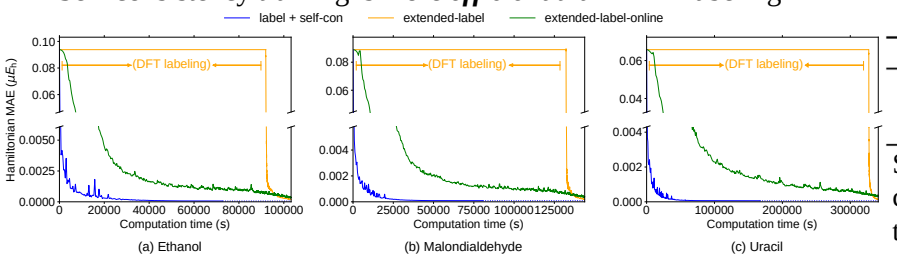
- Generalization improvement in the ***out-of-distribution*** scenario (QH9)

Setting	$\mathbf{H}$ [μE_h] ↓	$\boldsymbol{\epsilon}$ [μE_h] ↓	$\mathbf{C}$ [%] ↑	ϵ_{HOMO} [μE_h] ↓	ϵ_{LUMO} [μE_h] ↓	ϵ_Δ [μE_h] ↓	SCF Accel. [%] ↓
zero-shot	69.67	403.52	95.72	778.86	12230.49	12203.12	66.3
self-con (all-param)	65.74	375.31	**97.31**	565.50	**1130.55**	**1316.96**	**64.5**
self-con (adapter)	**64.48**	**268.83**	97.12	**449.80**	1220.54	1394.29	65.0

- Generalization to the ***larger-scale*** molecules (MD22)

Molecule	Setting	$\mathbf{H}$ [μE_h] ↓	$\boldsymbol{\epsilon}$ [μE_h] ↓	$\mathbf{C}$ [%] ↑	ϵ_{HOMO} [μE_h] ↓	ϵ_{LUMO} [μE_h] ↓	ϵ_Δ [μE_h] ↓	SCF Accel. [%] ↓
ALA3	zero-shot	237.71	6.54×10^3	52.24	6.90×10^3	9.51×10^4	9.79×10^4	84.6
	self-con	**52.49**	$\mathbf{1.22}\times10^3$	**94.46**	$\mathbf{2.07}\times10^3$	$\mathbf{3.76}\times10^3$	$\mathbf{2.69}\times10^3$	**64.7**
	e2e (ET)	N/A	N/A	N/A	1.74×10^5	7.72×10^3	2.38×10^5	N/A
	e2e (Equiformer)	N/A	N/A	N/A	2.38×10^5	1.16×10^4	2.27×10^5	N/A

- Self-consistency training is more ***efficient*** than DFT Labeling



Molecule	$t_{\text{self-con}}$ [s]	t_{DFT} [s]
Ethanol	$\mathbf{4.50}\times10^4$	6.40×10^4
Malondialdehyde	$\mathbf{4.81}\times10^4$	1.05×10^5
Uracil	$\mathbf{1.23}\times10^5$	2.15×10^5

Self-consistency training costs less than DFT to reach the same level of accuracy.