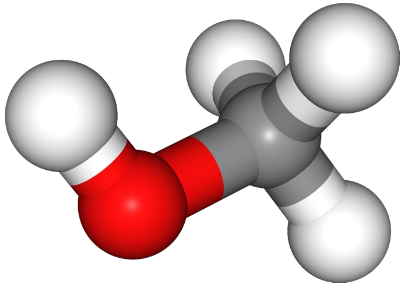
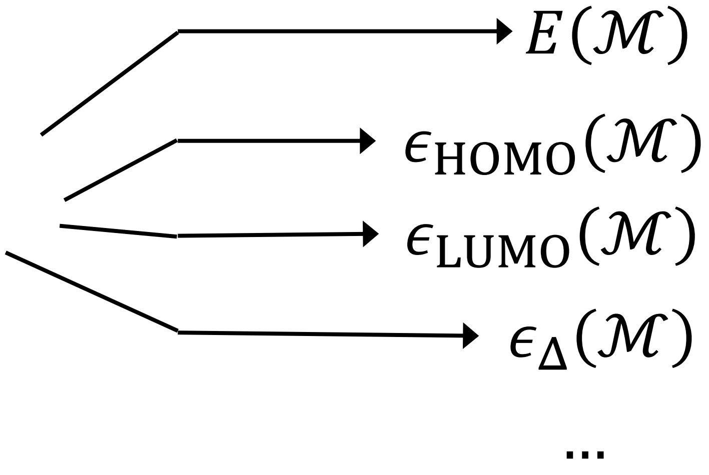
Self-Consistency Training for Density-Functional-Theory Hamiltonian Prediction

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 AI for Science [#]Corresponding authors

Hamiltonian Prediction



$\mathcal{M} \coloneqq \{\mathcal{Z}, \mathcal{R}\}$



$E(\mathcal{M})$

$\epsilon_{\mathrm{HOMO}}(\mathcal{M})$

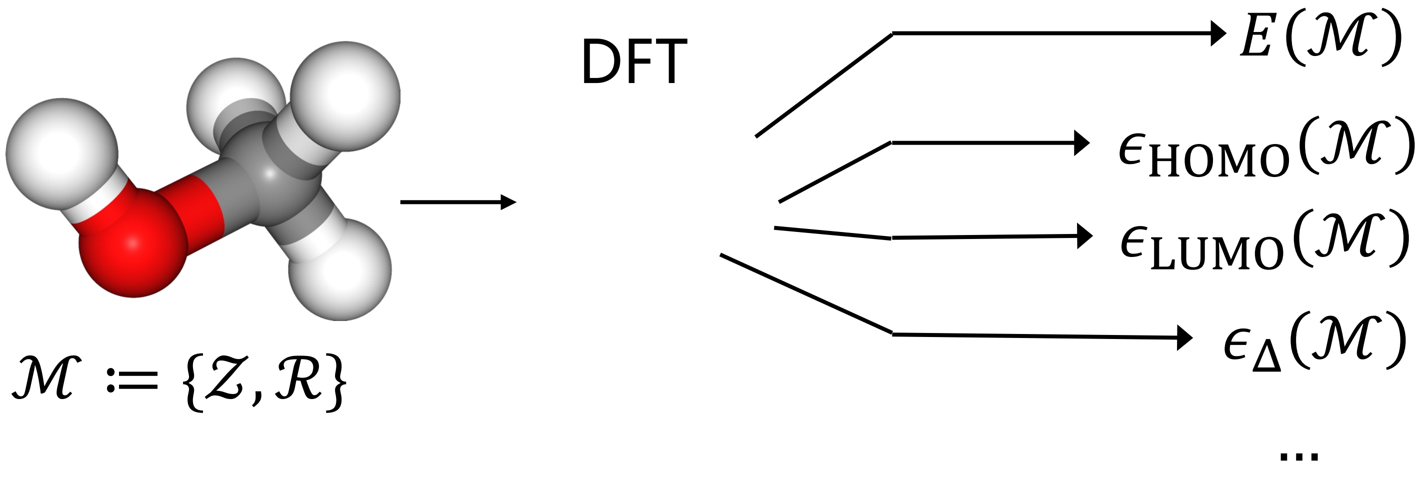
$\epsilon_{\mathrm{LUMO}}(\mathcal{M})$

$\epsilon_{\Delta}(\mathcal{M})$

…

- Molecular properties: interaction among electrons and atomic nuclei

Hamiltonian Prediction



$\mathcal{M} := \{\mathcal{Z}, \mathcal{R}\}$

DFT

$E(\mathcal{M})$

$\epsilon_{\mathrm{HOMO}}(\mathcal{M})$

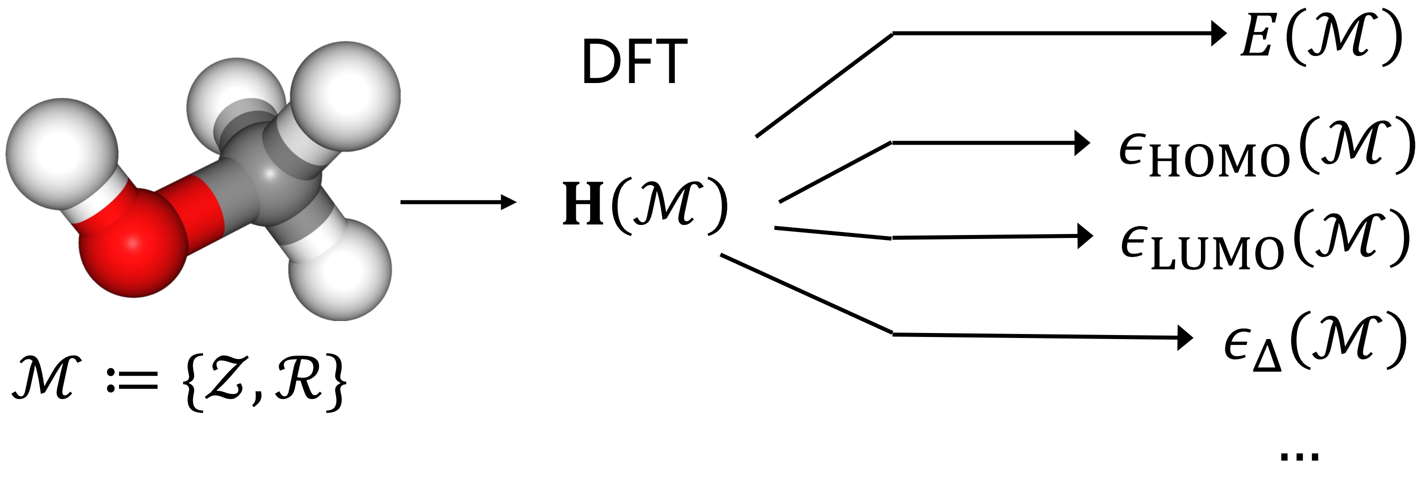
$\epsilon_{\mathrm{LUMO}}(\mathcal{M})$

$\epsilon_{\Delta}(\mathcal{M})$

…

- Molecular properties: interaction among electrons and atomic nuclei
- DFT: solve electronic structure hence properties

Hamiltonian Prediction



DFT

$\mathbf{H}(\mathcal{M})$

$E(\mathcal{M})$

$\epsilon_{\mathrm{HOMO}}(\mathcal{M})$

$\epsilon_{\mathrm{LUMO}}(\mathcal{M})$

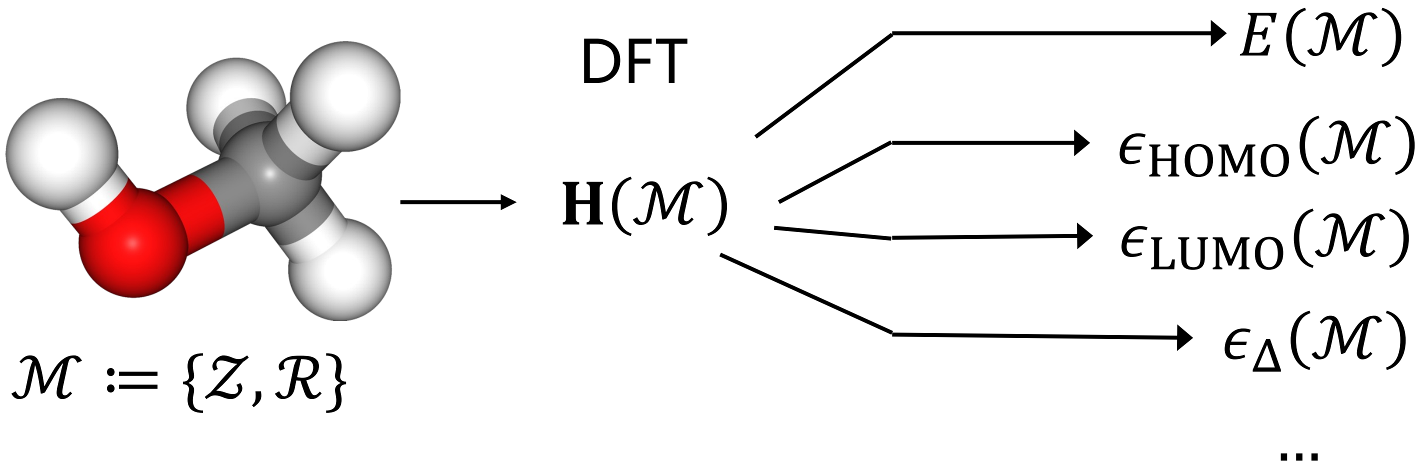
$\epsilon_{\Delta}(\mathcal{M})$

...

$\mathcal{M} \coloneqq \{\mathcal{Z}, \mathcal{R}\}$

- Molecular properties: interaction among electrons and atomic nuclei
- DFT: solve electronic structure hence properties
- Hamiltonian: raw DFT solution, derive all properties

Hamiltonian Prediction



DFT

$\mathcal{M} \coloneqq \{\mathcal{Z}, \mathcal{R}\}$

$\mathbf{H}(\mathcal{M})$

$E(\mathcal{M})$

$\epsilon_{\mathrm{HOMO}}(\mathcal{M})$

$\epsilon_{\mathrm{LUMO}}(\mathcal{M})$

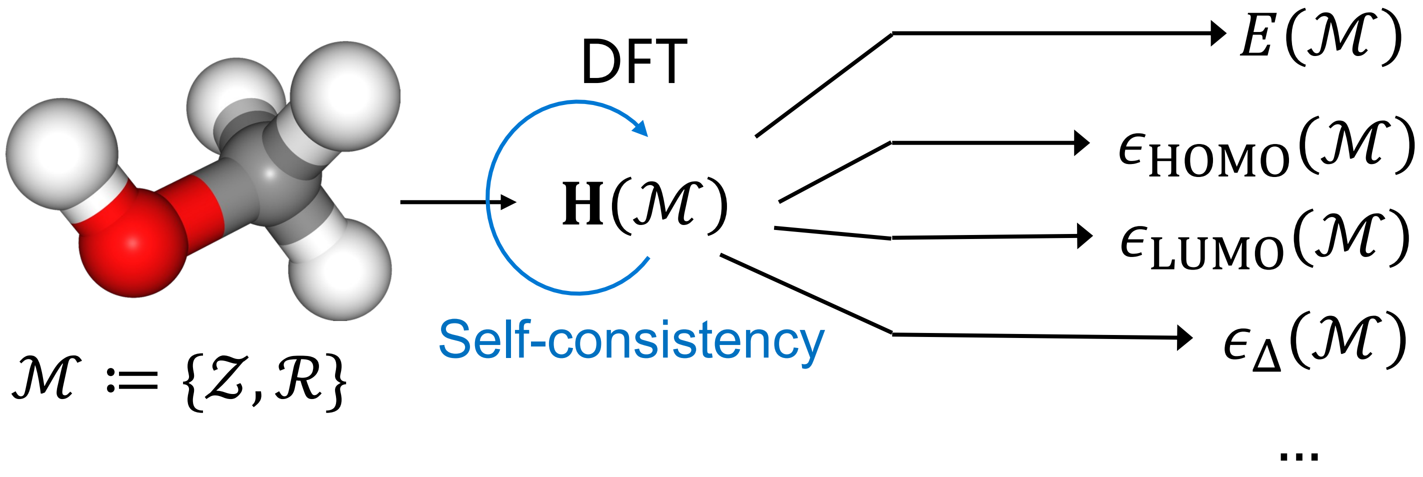
$\epsilon_{\Delta}(\mathcal{M})$

...

- Molecular properties: interaction among electrons and atomic nuclei
- DFT: solve electronic structure hence properties
- Hamiltonian: raw DFT solution, derive all properties
- Hamiltonian prediction:

  "unified" predictor, provide all properties that DFT can

Hamiltonian Prediction



$\mathcal{M} \coloneqq \{\mathcal{Z}, \mathcal{R}\}$

DFT

$\mathbf{H}(\mathcal{M})$

Self-consistency

$E(\mathcal{M})$

$\epsilon_{\mathrm{HOMO}}(\mathcal{M})$

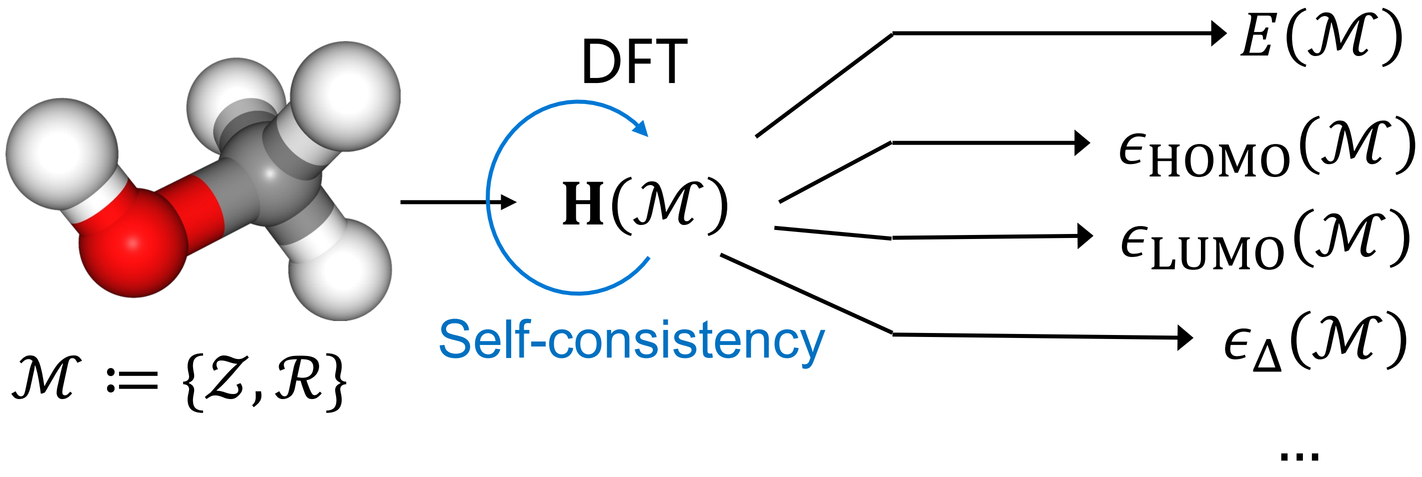
$\epsilon_{\mathrm{LUMO}}(\mathcal{M})$

$\epsilon_{\Delta}(\mathcal{M})$

…

- Hamiltonian prediction has a self-consistency principle: **Training without label!**
  - Distinction from common property prediction: data-free training / self-improvement
  - Compensating data scarcity with scientific laws

Hamiltonian Prediction


$\mathcal{M} \coloneqq \{\mathcal{Z}, \mathcal{R}\}$
DFT
$\mathbf{H}(\mathcal{M})$
Self-consistency
$E(\mathcal{M})$
$\epsilon_{\text{HOMO}}(\mathcal{M})$
$\epsilon_{\text{LUMO}}(\mathcal{M})$
$\epsilon_{\Delta}(\mathcal{M})$
...

- Hamiltonian prediction has a self-consistency principle: **Training without label!**
  - Distinction from common property prediction: data-free training / self-improvement
  - Compensating data scarcity with scientific laws
- Unique benefits:
  - Exact generalization to arbitrary workload beyond labeled data
  - Amortization of DFT calculation: more efficient than running DFT to generate labels

Background: DFT Formulation

- Describe N-electron state by orbitals $\{\phi_i(\mathbf{r})\}_{i=1}^{N}$ ➔ coefficients $\mathbf{C}$ under a basis set
- Solve for the electron state $\mathbf{C}$ of molecular structure $\mathcal{M}$ by minimizing:

$$E_{\mathcal{M}}(\mathbf{C}), \text{ s.t. } \mathbf{C}^{\top}\mathbf{S}_{\mathcal{M}}\,\mathbf{C} = \mathbf{I}.$$

- Solve the optimization problem:

$$\boxed{\underbrace{\mathbf{H}_{\mathcal{M}}(\mathbf{C})}_{:=\frac{1}{2}\nabla_{\mathbf{C}}E_{\mathcal{M}}(\cdot)}\,\mathbf{C} = \mathbf{S}_{\mathcal{M}}\,\mathbf{C}\,\boldsymbol{\epsilon}.}$$

Kohn-sham equation

Background: DFT Formulation

- Describe N-electron state by orbitals $\{\phi_i(\mathbf{r})\}_{i=1}^{N}$ ➔ coefficients $\mathbf{C}$ under a basis set
- Solve for the electron state $\mathbf{C}$ of molecular structure $\mathcal{M}$ by minimizing:

$$E_{\mathcal{M}}(\mathbf{C}),\ \text{s.t. } \mathbf{C}^{\top}\mathbf{S}_{\mathcal{M}}\,\mathbf{C} = \mathbf{I}.$$

- Solve the optimization problem:

$$\boxed{\underbrace{\mathbf{H}_{\mathcal{M}}(\mathbf{C})}_{:=\frac{1}{2}\nabla_{\mathbf{C}}E_{\mathcal{M}}(\cdot)}\,\mathbf{C} = \mathbf{S}_{\mathcal{M}}\,\mathbf{C}\,\boldsymbol{\epsilon}.}$$

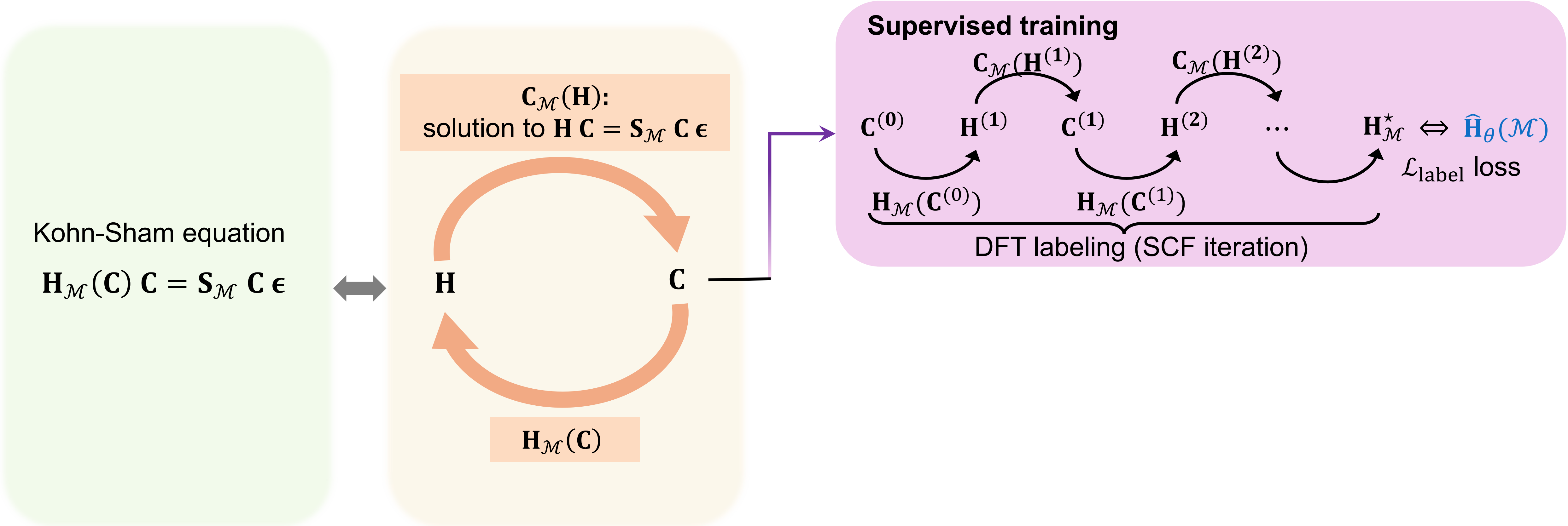
Kohn-sham equation

Self-Consistent Field (SCF) iteration:

$\mathbf{C}^{(k-1)}$ ➔ $\mathbf{H}^{(k)} = \mathbf{H}_{\mathcal{M}}\left(\mathbf{C}^{(k-1)}\right)$ ➔ $\mathbf{C}^{(k)} := \mathbf{C}_{\mathcal{M}}\left(\mathbf{H}^{(k)}\right)$ which solves $\mathbf{H}^{(k)}\mathbf{C} = \mathbf{S}_{\mathcal{M}}\mathbf{C}\boldsymbol{\epsilon}$

➔ $\mathbf{H}_{\mathcal{M}}^{\star}$ after convergence

DFT Calculation➔ Self-Consistency Training



Kohn-Sham equation

$\mathbf{H}_{\mathcal{M}}(\mathbf{C})\,\mathbf{C} = \mathbf{S}_{\mathcal{M}}\,\mathbf{C}\,\boldsymbol{\epsilon}$

$\mathbf{C}_{\mathcal{M}}(\mathbf{H})$:
solution to $\mathbf{H}\,\mathbf{C} = \mathbf{S}_{\mathcal{M}}\,\mathbf{C}\,\boldsymbol{\epsilon}$

$\mathbf{H}_{\mathcal{M}}(\mathbf{C})$

Supervised training

$$\mathbf{C}^{(0)} \xrightarrow{\mathbf{H}_{\mathcal{M}}(\mathbf{C}^{(0)})} \mathbf{H}^{(1)} \xrightarrow{\mathbf{C}_{\mathcal{M}}(\mathbf{H}^{(1)})} \mathbf{C}^{(1)} \xrightarrow{\mathbf{H}_{\mathcal{M}}(\mathbf{C}^{(1)})} \mathbf{H}^{(2)} \xrightarrow{\mathbf{C}_{\mathcal{M}}(\mathbf{H}^{(2)})} \cdots \longrightarrow \mathbf{H}^{\star}_{\mathcal{M}} \Leftrightarrow \hat{\mathbf{H}}_{\theta}(\mathcal{M})$$

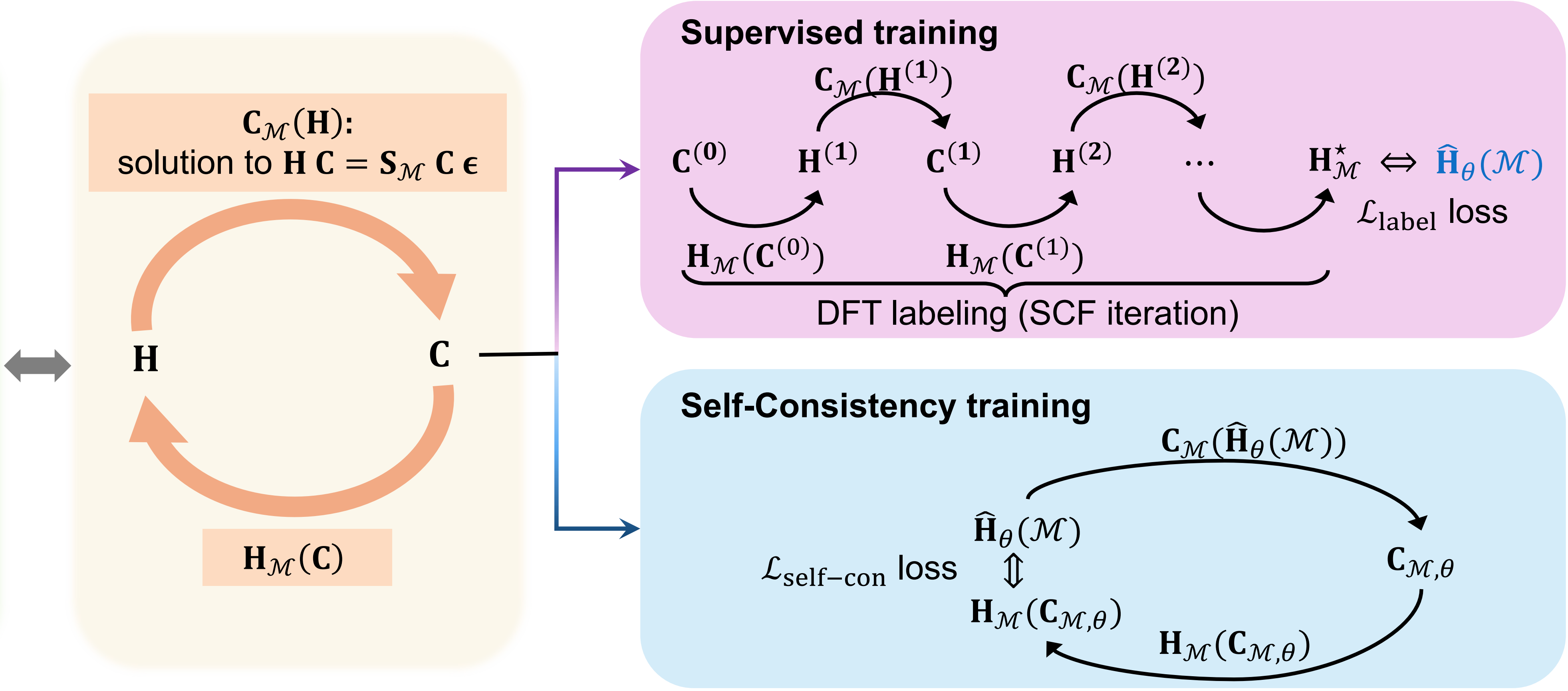
$\mathcal{L}_{\text{label}}$ loss

DFT labeling (SCF iteration)

DFT Calculation➔ Self-Consistency Training

Kohn-Sham equation

$$\mathbf{H}_{\mathcal{M}}(\mathbf{C})\,\mathbf{C} = \mathbf{S}_{\mathcal{M}}\,\mathbf{C}\,\epsilon$$



$\mathbf{C}_{\mathcal{M}}(\mathbf{H})$:
solution to $\mathbf{H}\,\mathbf{C} = \mathbf{S}_{\mathcal{M}}\,\mathbf{C}\,\epsilon$

$\mathbf{H}$ $\mathbf{C}$

$\mathbf{H}_{\mathcal{M}}(\mathbf{C})$

Supervised training

$\mathbf{C}^{(0)}$ $\mathbf{H}^{(1)}$ $\mathbf{C}^{(1)}$ $\mathbf{H}^{(2)}$ $\cdots$ $\mathbf{H}^{\star}_{\mathcal{M}} \Leftrightarrow \hat{\mathbf{H}}_{\theta}(\mathcal{M})$

$\mathbf{C}_{\mathcal{M}}(\mathbf{H}^{(1)})$ $\mathbf{C}_{\mathcal{M}}(\mathbf{H}^{(2)})$

$\mathbf{H}_{\mathcal{M}}(\mathbf{C}^{(0)})$ $\mathbf{H}_{\mathcal{M}}(\mathbf{C}^{(1)})$

$\mathcal{L}_{\text{label}}$ loss

DFT labeling (SCF iteration)

Self-Consistency training

$\mathbf{C}_{\mathcal{M}}(\hat{\mathbf{H}}_{\theta}(\mathcal{M}))$

$\hat{\mathbf{H}}_{\theta}(\mathcal{M})$ $\Updownarrow$ $\mathbf{H}_{\mathcal{M}}(\mathbf{C}_{\mathcal{M},\theta})$

$\mathcal{L}_{\text{self-con}}$ loss

$\mathbf{C}_{\mathcal{M},\theta}$

$\mathbf{H}_{\mathcal{M}}(\mathbf{C}_{\mathcal{M},\theta})$

$$L_{\text{sc}}(\theta;\mathcal{D}) = \frac{1}{|\mathcal{D}|}\sum_{\mathcal{M}\sim\mathcal{D}} \left\| \hat{\mathbf{H}}_{\theta}(\mathcal{M}) - \mathbf{H}_{\mathcal{M}}\left(\mathbf{C}_{\mathcal{M}}\left(\hat{\mathbf{H}}_{\theta}(\mathcal{M})\right)\right) \right\|_{\mathrm{F}}^{2}$$

Self-Consistency Training

- Self-consistency loss

$$L_{\mathrm{sc}}(\theta;\mathcal{D}) = \frac{1}{|\mathcal{D}|}\sum\nolimits_{\mathcal{M}\sim\mathcal{D}} \left\| \widehat{\mathbf{H}}_{\theta}(\mathcal{M}) - \mathbf{H}_{\mathcal{M}}\left(\mathbf{C}_{\mathcal{M}}\left(\widehat{\mathbf{H}}_{\theta}(\mathcal{M})\right)\right)\right\|_{\mathrm{F}}^{2}$$

  - Not just a regularization: it determines the DFT solution (label).

Self-Consistency Training

- Self-consistency loss

$$L_{\mathrm{sc}}(\theta;\mathcal{D}) = \frac{1}{|\mathcal{D}|}\sum\nolimits_{\mathcal{M}\sim\mathcal{D}} \left\| \hat{\mathbf{H}}_\theta(\mathcal{M}) - \mathbf{H}_{\mathcal{M}}\left(\mathbf{C}_{\mathcal{M}}\left(\hat{\mathbf{H}}_\theta(\mathcal{M})\right)\right) \right\|_{\mathrm{F}}^2$$

  - Not just a regularization: it determines the DFT solution (label).
  - Minimizing the gap unnecessarily drives $\hat{\mathbf{H}}_\theta(\mathcal{M})$ towards $\mathbf{H}_{\mathcal{M}}\left(\mathbf{C}_{\mathcal{M}}\left(\hat{\mathbf{H}}_\theta(\mathcal{M})\right)\right)$.
    - The latter may even be farther from the solution, in which case both are driven to the solution.
    - Should not apply stop-gradient to the latter.

Self-Consistency Training

- Hamiltonian prediction
  - Roto-translational/$\mathrm{SE}(3)$ equivariance
  - QHNet [Yu'23]: an $\mathrm{SE}(3)$-equivariant GNN balance efficiency and accuracy

[Yu'23] Yu H, Xu Z, Qian X, et al. Efficient and equivariant graph networks for predicting quantum Hamiltonian[C]//International Conference on Machine Learning. PMLR, 2023.

Self-Consistency Training

- Hamiltonian prediction
  - Roto-translational/$\mathrm{SE}(3)$ equivariance
  - QHNet [Yu'23]: an $\mathrm{SE}(3)$-equivariant GNN balance efficiency and accuracy
- Hamiltonian reconstruction

$$L_{\mathrm{sc}}(\theta;\mathcal{D}) = \frac{1}{|\mathcal{D}|}\sum\nolimits_{\mathcal{M}\sim\mathcal{D}} \left\| \hat{\mathbf{H}}_{\theta}(\mathcal{M}) - \mathbf{H}_{\mathcal{M}}\left(\mathbf{C}_{\mathcal{M}}\left(\hat{\mathbf{H}}_{\theta}(\mathcal{M})\right)\right) \right\|_{\mathrm{F}}^{2}$$

  - Numerically stable implementation of differentiation through eigensolver $\mathbf{C}_{\mathcal{M}}(\mathbf{H})$.
  - GPU implementation of Hamiltonian construction $\mathbf{H}_{\mathcal{M}}(\mathbf{C})$

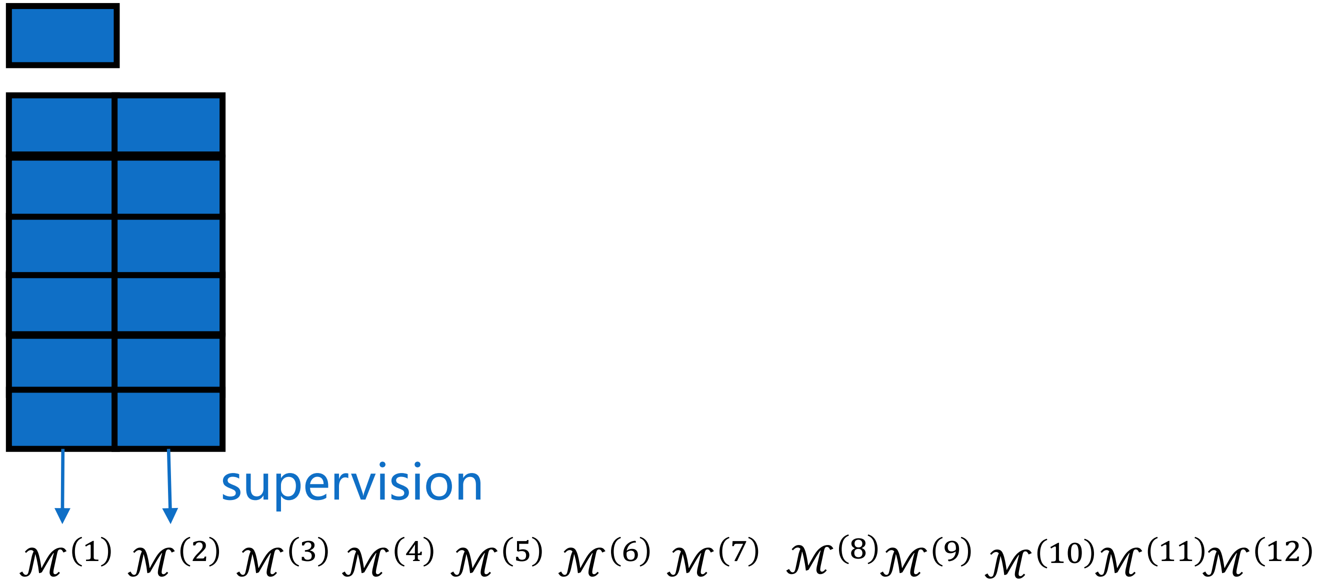
[Yu'23] Yu H, Xu Z, Qian X, et al. Efficient and equivariant graph networks for predicting quantum Hamiltonian[C]//International Conference on Machine Learning. PMLR, 2023.

Unique Benefits

- Generalization beyond labeled data: $\mathcal{L}_{\text{label}}\left(\theta;\underbrace{\overline{\mathcal{D}^{(1)}}}_{\text{limited labeled dataset}}\right) + \lambda\,\mathcal{L}_{\text{self-con}}\left(\theta;\underbrace{\mathcal{D}^{(2)}}_{\text{unlimited unlabeled dataset}}\right)$

Unique Benefits

- Generalization beyond labeled data: $\mathcal{L}_{\text{label}}\big(\theta;\underbrace{\overline{\mathcal{D}^{(1)}}}_{\text{limited labeled dataset}}\big) + \lambda\,\mathcal{L}_{\text{self-con}}\big(\theta;\underbrace{\mathcal{D}^{(2)}}_{\text{unlimited unlabeled dataset}}\big)$
- Amortization effect: efficiency over DFT labeling



Cost of one iteration:

DFT calculation:

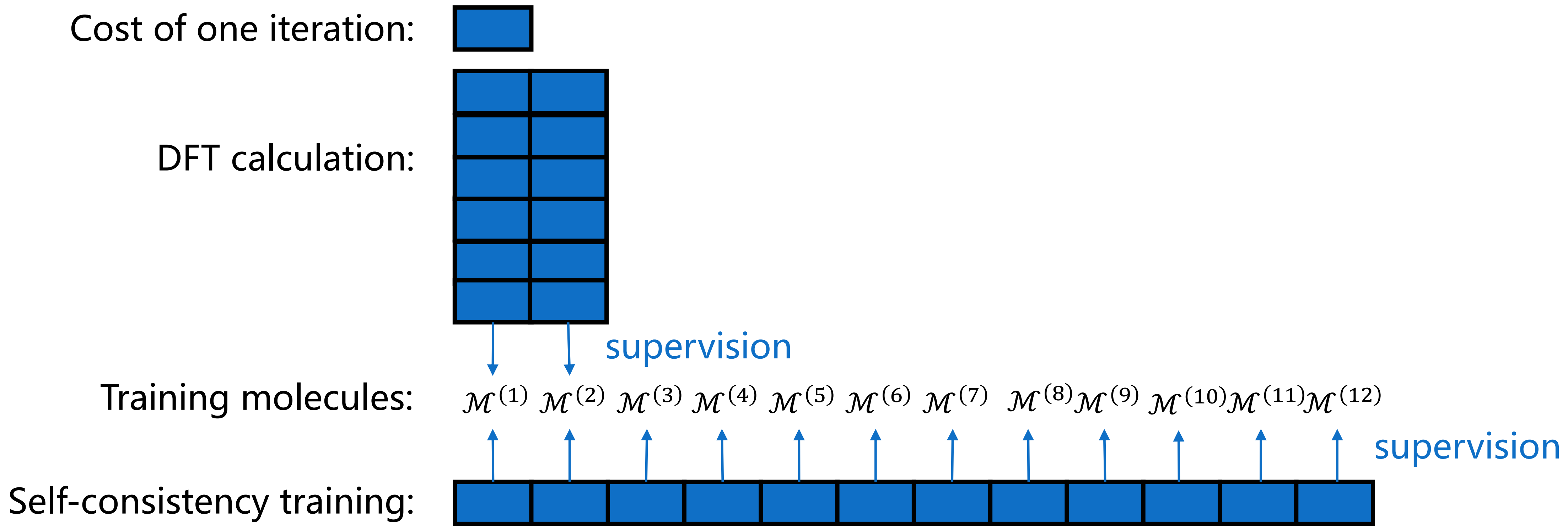
supervision

Training molecules: $\mathcal{M}^{(1)}\ \mathcal{M}^{(2)}\ \mathcal{M}^{(3)}\ \mathcal{M}^{(4)}\ \mathcal{M}^{(5)}\ \mathcal{M}^{(6)}\ \mathcal{M}^{(7)}\ \mathcal{M}^{(8)}\mathcal{M}^{(9)}\ \mathcal{M}^{(10)}\mathcal{M}^{(11)}\mathcal{M}^{(12)}$

Unique Benefits

- Generalization beyond labeled data: $\mathcal{L}_{\text{label}}\left(\theta; \underbrace{\overline{\mathcal{D}^{(1)}}}_{\text{limited labeled dataset}}\right) + \lambda\, \mathcal{L}_{\text{self-con}}\left(\theta; \underbrace{\mathcal{D}^{(2)}}_{\text{unlimited unlabeled dataset}}\right)$

- Amortization effect: efficiency over DFT labeling



Cost of one iteration:

DFT calculation:

supervision

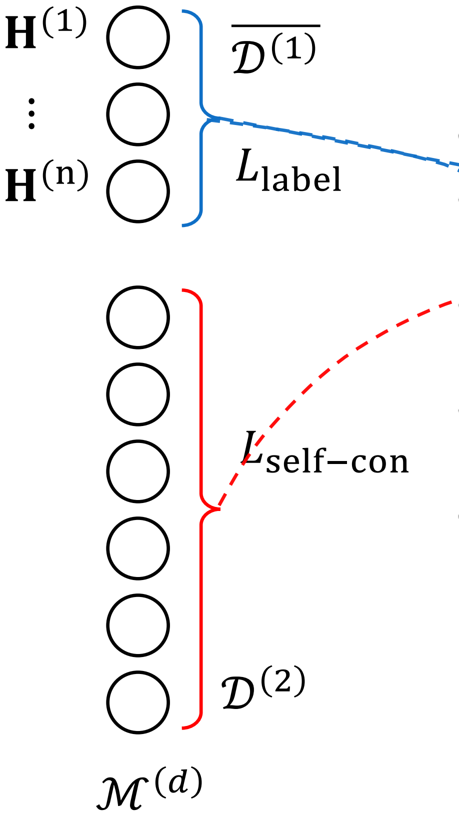
Training molecules: $\mathcal{M}^{(1)}$ $\mathcal{M}^{(2)}$ $\mathcal{M}^{(3)}$ $\mathcal{M}^{(4)}$ $\mathcal{M}^{(5)}$ $\mathcal{M}^{(6)}$ $\mathcal{M}^{(7)}$ $\mathcal{M}^{(8)}$ $\mathcal{M}^{(9)}$ $\mathcal{M}^{(10)}$ $\mathcal{M}^{(11)}$ $\mathcal{M}^{(12)}$

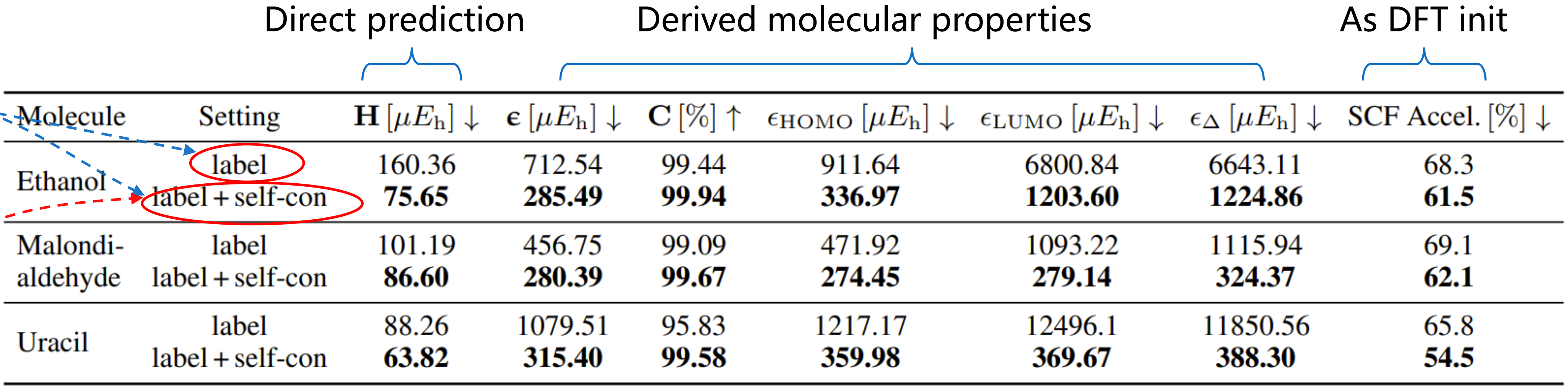
supervision

Self-consistency training:

Results: Generalization beyond Labeled Data

- Data-scarce scenario (MD17): $\left|\overline{\mathcal{D}^{(1)}}\right| = 100, \left|\mathcal{D}^{(2)}\right| = 24,900$

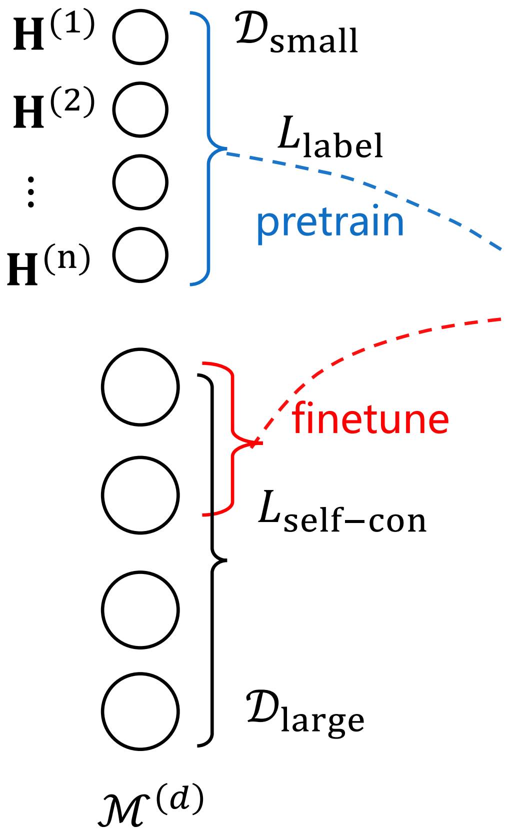


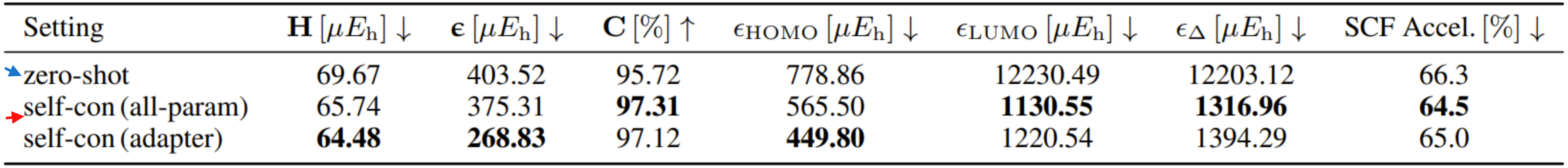


		Direct prediction	Derived molecular properties					As DFT init
Molecule	Setting	$\mathbf{H}$ $[\mu E_\text{h}]\downarrow$	$\boldsymbol{\epsilon}$ $[\mu E_\text{h}]\downarrow$	$\mathbf{C}$ $[\%]\uparrow$	ϵ_HOMO $[\mu E_\text{h}]\downarrow$	ϵ_LUMO $[\mu E_\text{h}]\downarrow$	ϵ_Δ $[\mu E_\text{h}]\downarrow$	SCF Accel. $[\%]\downarrow$
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**

Results: Generalization beyond Labeled Data

- Out-of-distribution (OOD) scenario (QH9)
  labeled small molecules + finetune on unlabeled large molecules ➔ test on large molecules





Setting	$\mathbf{H}$ $[\mu E_\mathrm{h}]\downarrow$	$\boldsymbol{\epsilon}$ $[\mu E_\mathrm{h}]\downarrow$	$\mathbf{C}$ $[\%]\uparrow$	ϵ_{HOMO} $[\mu E_\mathrm{h}]\downarrow$	ϵ_{LUMO} $[\mu E_\mathrm{h}]\downarrow$	ϵ_{Δ} $[\mu E_\mathrm{h}]\downarrow$	SCF Accel. $[\%]\downarrow$
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

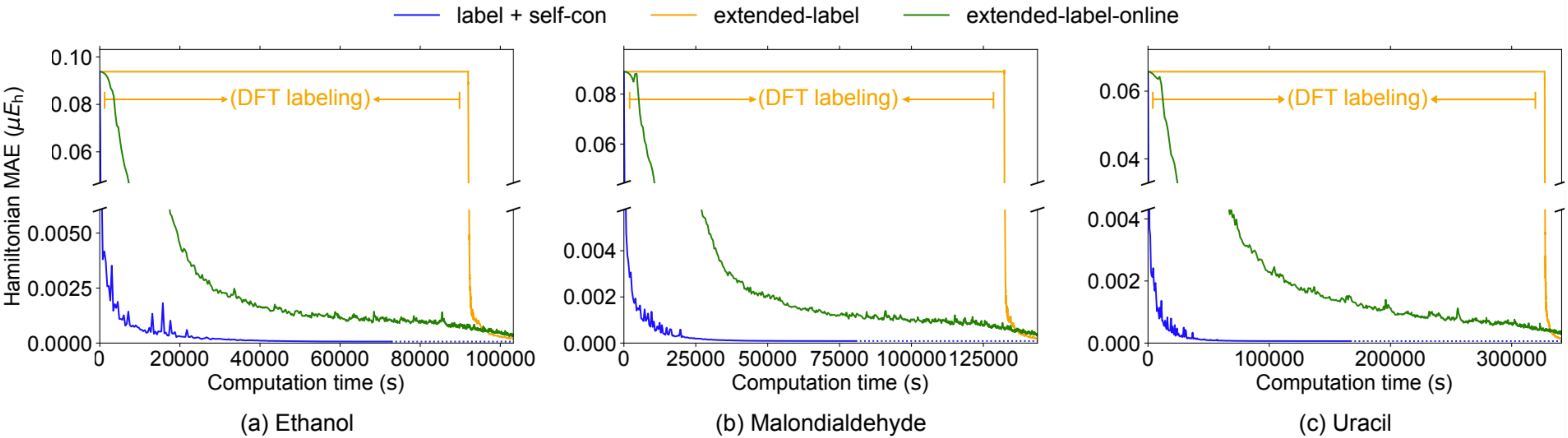
Results: Training Efficiency by Amortization

- Data-scarce scenario (MD17): $|\mathcal{D}^{(1)}| = 100, |\mathcal{D}^{(2)}| = 24900$

$\mathcal{L}_{\text{self-con}}$ on unlabeled

Label the unlabeled *then* $\mathcal{L}_{\text{label}}$

Label the unlabeled *while* $\mathcal{L}_{\text{label}}$



(a) Ethanol

(b) Malondialdehyde

(c) Uracil

accuracy-cost (*computation time*) curve

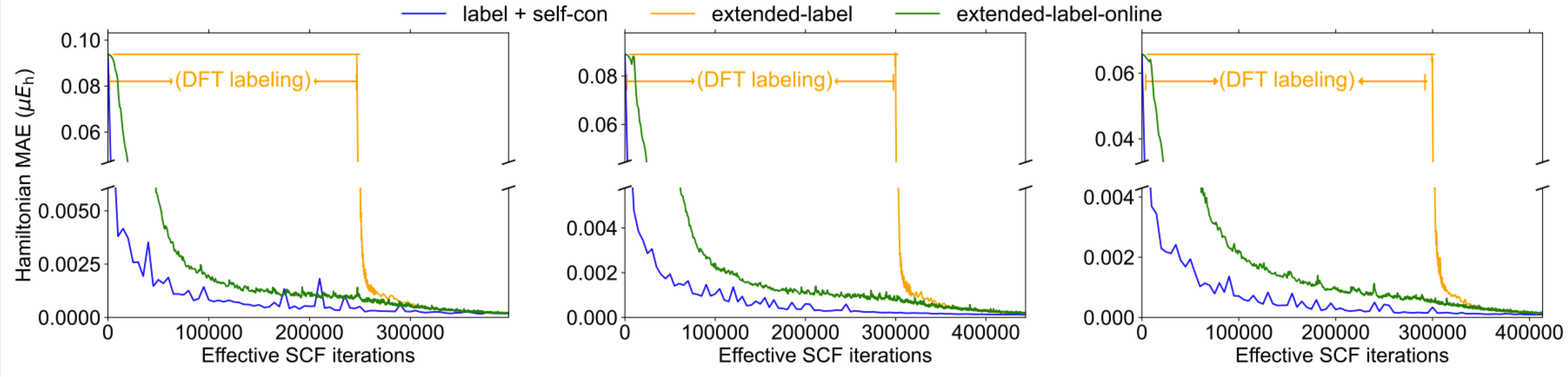
Results: Amortization of DFT

- Data-scarce scenario (MD17): $\left|\overline{\mathcal{D}^{(1)}}\right| = 100, \left|\mathcal{D}^{(2)}\right| = 24900$

$\mathcal{L}_{\text{self-con}}$ on unlabeled

Label the unlabeled *then* $\mathcal{L}_{\text{label}}$

Label the unlabeled *while* $\mathcal{L}_{\text{label}}$

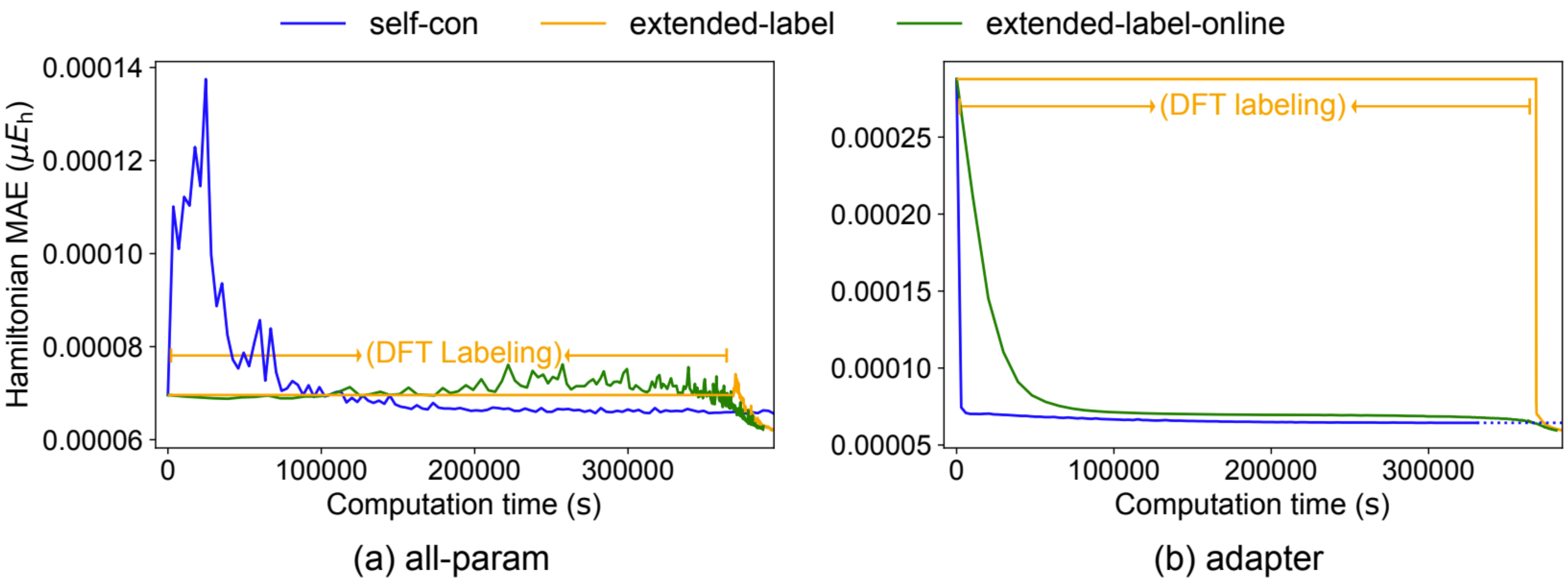


accuracy-cost (*effective SCF iterations*) curve

Results: Amortization of DFT

- OOD scenario (QH9)

  labeled small molecules + finetune on unlabeled large molecules ➔ test on large molecules

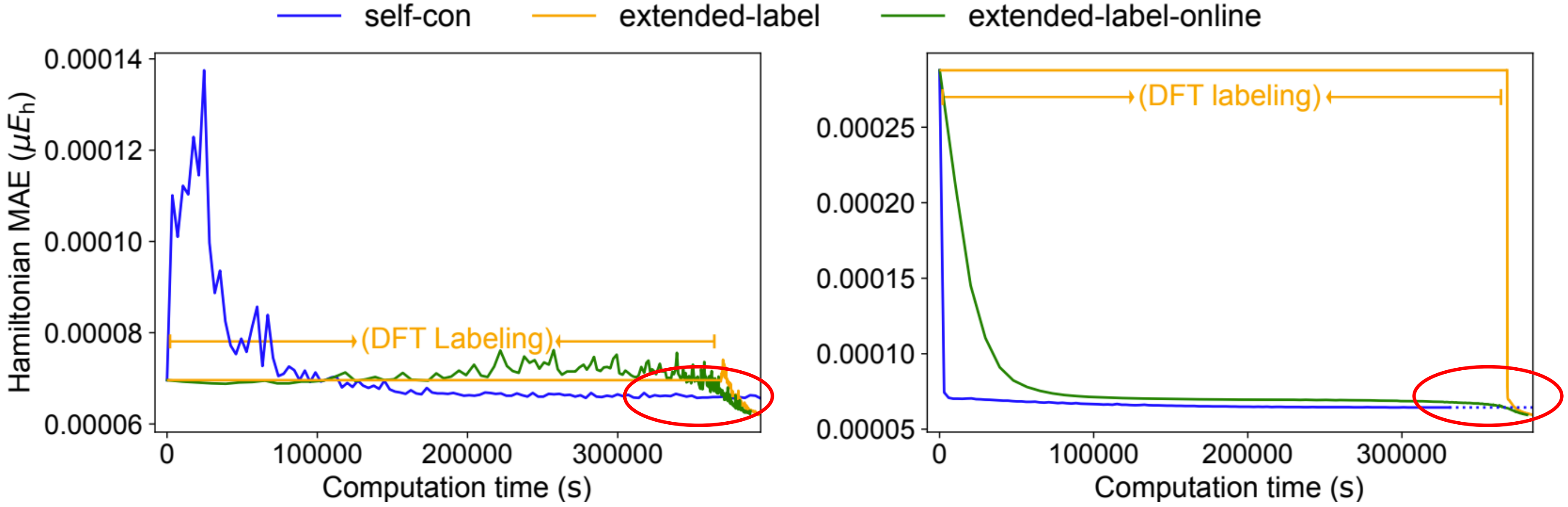


accuracy-cost (computation time) curve

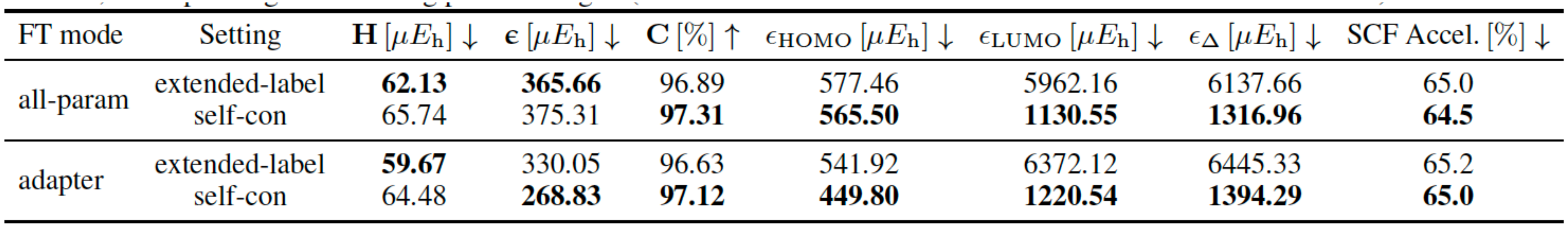
Results: Amortization of DFT

- OOD scenario (QH9)

labeled small molecules + finetune on unlabeled large molecules ➔ test on large molecules



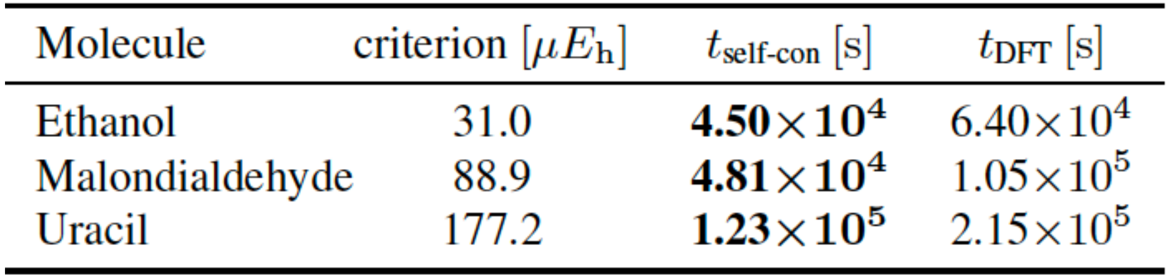
Final performance: self-consistency training gives better derived molecular properties!



FT mode	Setting	**H** [μE_h] ↓	ϵ [μE_h] ↓	**C** [%] ↑	ϵ_{HOMO} [μE_h] ↓	ϵ_{LUMO} [μE_h] ↓	ϵ_{Δ} [μE_h] ↓	SCF Accel. [%] ↓
all-param	extended-label	**62.13**	**365.66**	96.89	577.46	5962.16	6137.66	65.0
	self-con	65.74	375.31	**97.31**	**565.50**	**1130.55**	**1316.96**	**64.5**
adapter	extended-label	**59.67**	330.05	96.63	541.92	6372.12	6445.33	65.2
	self-con	64.48	**268.83**	**97.12**	**449.80**	**1220.54**	**1394.29**	**65.0**

Results: Amortization of DFT

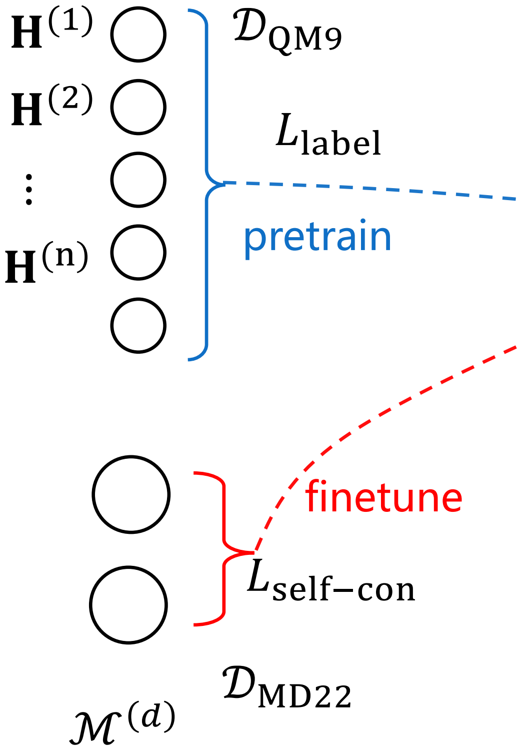
- Direct acceleration over DFT calculation
  - Self-consistency training time vs. DFT computation time to reach the same level of electronic energy accuracy



Molecule	criterion $[\mu E_\mathrm{h}]$	$t_\text{self-con}$ [s]	t_DFT [s]
Ethanol	31.0	$\mathbf{4.50 \times 10^4}$	6.40×10^4
Malondialdehyde	88.9	$\mathbf{4.81 \times 10^4}$	1.05×10^5
Uracil	177.2	$\mathbf{1.23 \times 10^5}$	2.15×10^5

Results: Extending Applicable Scale of Hamiltonian Prediction

- Labeled QM9 (≤ 31atoms)+Finetune on unlabeled larger molecules → test on MD22 (ALA3: 42 atoms, DHA: 56 atoms)
  - vs. zero-shot generalization
  - vs. SOTA end-to-end property predictors



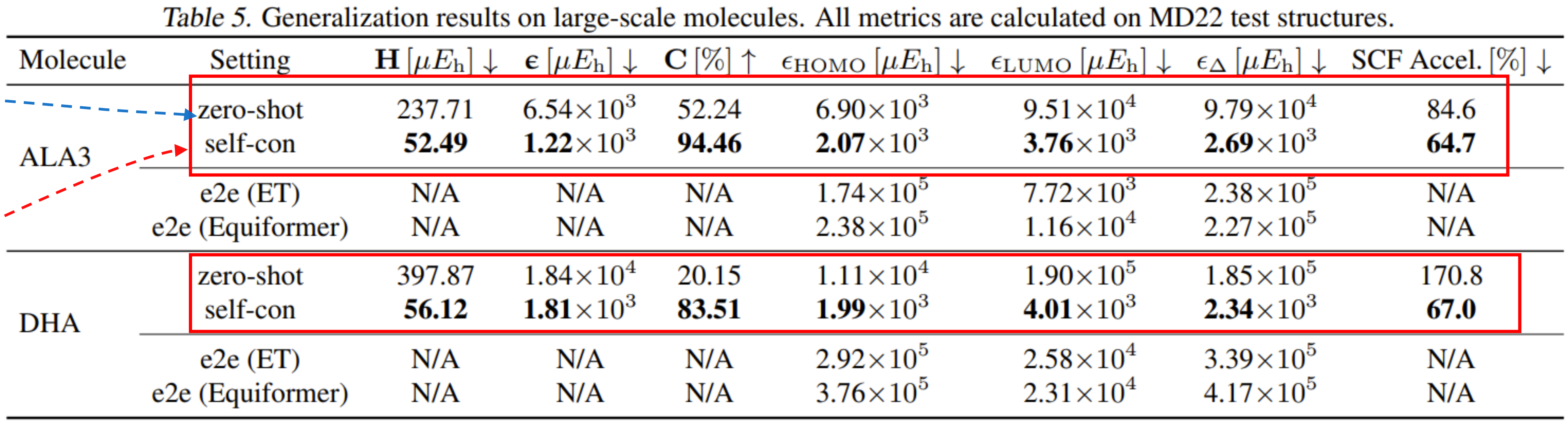
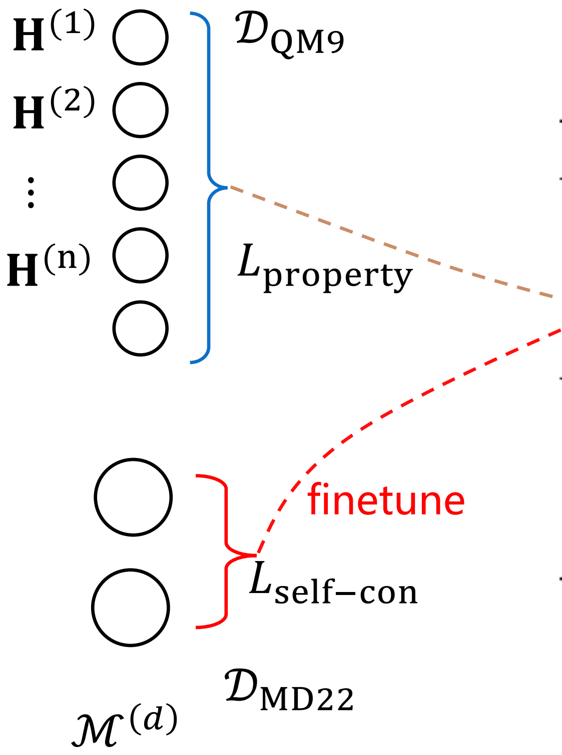


Table 5. Generalization results on large-scale molecules. All metrics are calculated on MD22 test structures.

Molecule	Setting	$\mathbf{H}$ $[\mu E_\mathrm{h}]\downarrow$	$\boldsymbol{\epsilon}$ $[\mu E_\mathrm{h}]\downarrow$	$\mathbf{C}$ $[\%]\uparrow$	ϵ_{HOMO} $[\mu E_\mathrm{h}]\downarrow$	ϵ_{LUMO} $[\mu E_\mathrm{h}]\downarrow$	ϵ_{Δ} $[\mu E_\mathrm{h}]\downarrow$	SCF Accel. $[\%]\downarrow$
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**	**1.22**$\times10^3$	**94.46**	**2.07**$\times10^3$	**3.76**$\times10^3$	**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
DHA	zero-shot	397.87	1.84×10^4	20.15	1.11×10^4	1.90×10^5	1.85×10^5	170.8
	self-con	**56.12**	**1.81**$\times10^3$	**83.51**	**1.99**$\times10^3$	**4.01**$\times10^3$	**2.34**$\times10^3$	**67.0**
	e2e (ET)	N/A	N/A	N/A	2.92×10^5	2.58×10^4	3.39×10^5	N/A
	e2e (Equiformer)	N/A	N/A	N/A	3.76×10^5	2.31×10^4	4.17×10^5	N/A

Results: Extending Applicable Scale of Hamiltonian Prediction

- QM9 (≤ 31atoms) → MD22 (ALA3: 42 atoms, DHA: 56 atoms)
  - vs. zero-shot generalization
  - vs. SOTA end-to-end property predictors



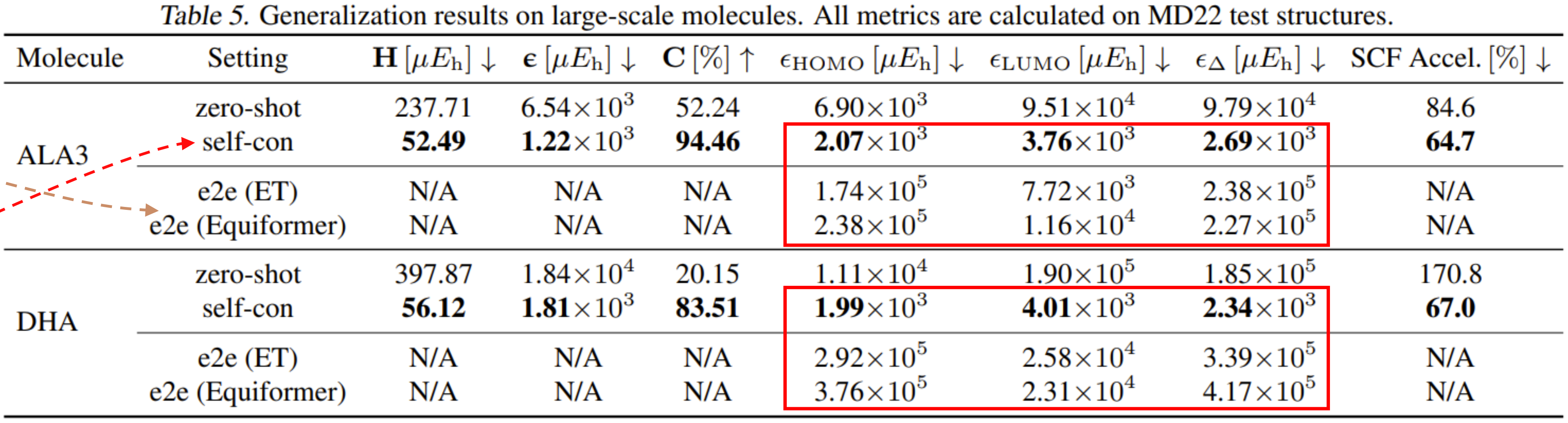


Table 5. Generalization results on large-scale molecules. All metrics are calculated on MD22 test structures.

Molecule	Setting	**H** [μE_h] ↓	**ϵ** [μE_h] ↓	**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
DHA	zero-shot	397.87	1.84×10^4	20.15	1.11×10^4	1.90×10^5	1.85×10^5	170.8
	self-con	**56.12**	$\mathbf{1.81}\times10^3$	**83.51**	$\mathbf{1.99}\times10^3$	$\mathbf{4.01}\times10^3$	$\mathbf{2.34}\times10^3$	**67.0**
	e2e (ET)	N/A	N/A	N/A	2.92×10^5	2.58×10^4	3.39×10^5	N/A
	e2e (Equiformer)	N/A	N/A	N/A	3.76×10^5	2.31×10^4	4.17×10^5	N/A

西安交通大學 XI'AN JIAOTONG UNIVERSITY | IAIR Est. 1986 Institute of Artificial Intelligence and Robotics | Microsoft

Thank you

https://arxiv.org/pdf/2403.09560

changliu@microsoft.com
mothful123@gmail.com