Robust and Interpretable Adaptation of Equivariant Materials Foundation Models via Sparsity-promoting Fine-tuning

Youngwoo Cho*, Seunghoon Yi*, Wooil Yang,

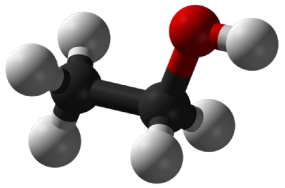
Sungmo Kang, Young-Woo Son, Jaegul Choo,

Joonseok Lee, Soo Kyung Kim, Hongkee Yoon

ICLR 2026

Motivation: Density Functional Theory [1/3]

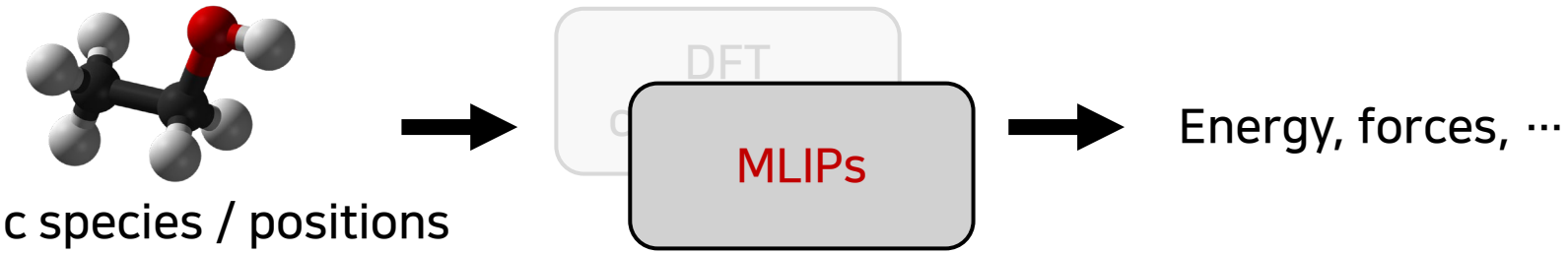
- Density functional theory (DFT)
  - The standard approach for computing material properties
  - Based on quantum mechanical principles
  - Given a material structure, DFT provides various physical quantities.
    - Input: Atomic species & coordinates
    - Output: Energy, forces, electron density, vibrational properties
  - However, its $O(N^3)$ computational cost severely limits scalability.
    - N: the number of atoms in the system.



Atomic species / positions → DFT calculation → Energy, forces, ⋯

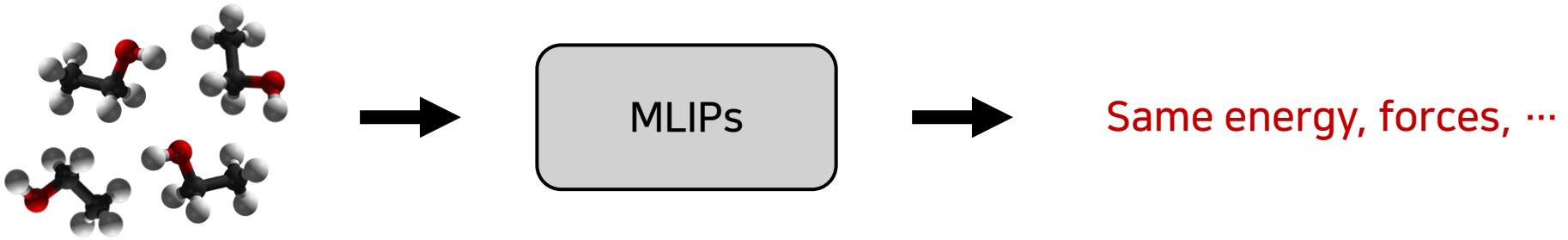
Motivation: Replacing DFT with MLIPs [2/3]

- Machine learning interatomic potential (MLIP)
  - MLIPs can serve as an efficient alternative to DFT calculations.
    - They learn to approximate the potential energy surface (PES) from DFT data.
    - This enables near-DFT accuracy with significantly reduced computational cost.
  - Today, foundation MLIPs pre-trained on large-scale datasets have emerged
    - MACE (Batatia et al., 2022), NequIP (Batzner et al., 2022), eSEN (Fu et al., 2025)
    - However, they require domain-specific fine-tuning due to:
      - Modeling defective or distorted structures, ⋯
      - Matching the specific exchange-correlation functional



Motivation: Fine-tuning MLIPs [3/3]

- A key constraint in fine-tuning MLIPs: Equivariance
  - Fine-tuning method must preserve translational and rotational symmetries.
  - Prior work on equivariance-preserving fine-tuning:
    - ELoRA (Wang et al., 2025)
    - GeoAda (Zhao et al., 2025)
- Is there a more effective fine-tuning method for equivariant MLIPs?
  - Proposed solution: Sparsity-promoting approach



Proposed method [1/4]

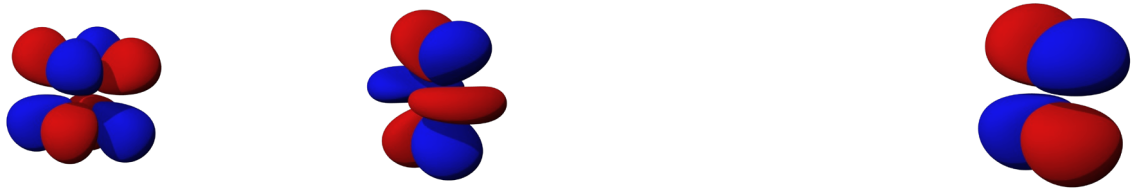
- Background: How equivariant MLIPs work?
  - Fundamental operations in equivariant MLIPs are tensor products:

$$\phi_{k_3,\ell_3,m_3}^{(\ell_1,\ell_2\to\ell_3)}(\boldsymbol{u}\otimes\boldsymbol{v}) = \sum_{k_1,k_2}\sum_{m_1,m_2} W_{\ell_3\ell_2\ell_1}^{k_3k_2k_1} C_{\ell_1m_1,\ell_2m_2}^{\ell_3m_3} u_{k_1\ell_1m_1} v_{k_2\ell_2m_2}$$

    - $\boldsymbol{u}$ & $\boldsymbol{v}$: atomic-level features, analogous to atomic orbitals being coupled
    - Pre-defined weights: Clebsch-Gordan coefficients $C_{\ell_1m_1,\ell_2m_2}^{\ell_3m_3}$
    - Learnable components: Linear combination coefficients $W_{\ell_3\ell_2\ell_1}^{k_3k_2k_1}$
    - Check mathematical details in Geiger and Smidts (2022).

Proposed method [2/4]

- Background: How equivariant MLIPs work?
  - Intuitively, weights that satisfy equivariance are pre-defined,
    and the model operates through linear combinations of these weights.



$$\boldsymbol{\theta} = \alpha_1\boldsymbol{\theta}_1 + \alpha_2\boldsymbol{\theta}_2 + \ldots + \alpha_n\boldsymbol{\theta}_n$$

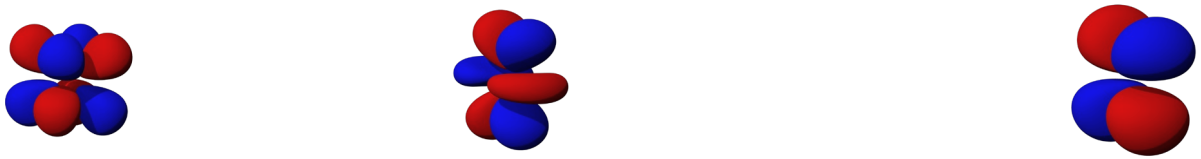
α: Learnable
$\boldsymbol{\theta}$: Frozen

  - Here, each weight θ corresponds to the concept of atomic orbitals.
  - Only the coefficients (alphas) are updated during model training.
    - In the previous slide, α corresponds to W

Proposed method [3/4]

- Fine-tuning as weight update: $W' = W + \Delta W$
  - W: pre-trained weights, ΔW: domain-specific adaptation
- What if we make $\Delta W \rightarrow \Delta W^{k_3 k_2 k_1}_{\ell_3 \ell_2 \ell_1}$ sparse?
  - Only a minimal number of parameters are updated.
  - Equivariance is naturally preserved (ΔW operates on W, not on C).

$$\phi'^{(\ell_1,\ell_2\rightarrow\ell_3)}_{k_3,\ell_3,m_3}(\boldsymbol{u}\otimes\boldsymbol{v}) = \sum_{k_1,k_2}\sum_{m_1,m_2}(W^{k_3k_2k_1}_{\ell_3\ell_2\ell_1} + \Delta W^{k_3k_2k_1}_{\ell_3\ell_2\ell_1})C^{\ell_3 m_3}_{\ell_1 m_1,\ell_2 m_2} u_{k_1\ell_1 m_1} v_{k_2\ell_2 m_2}$$



$$\boldsymbol{\theta} = (\alpha_1 + \cancel{\Delta\alpha_1})\boldsymbol{\theta}_1 + (\alpha_2 + \Delta\alpha_2)\boldsymbol{\theta}_2 + \ldots + (\alpha_n + \cancel{\Delta\alpha_n})\boldsymbol{\theta}_n$$

Proposed method [4/4]

- How to promote sparsity in ΔW?
  - The proposed method starts from soft threshold reparameterization (STR).
    - STR is originally introduced in the computer vision domain.
  - Sparsity is governed by a learnable, per-parameter scalar threshold δ.
    - Parameters with magnitudes below δ in ΔW are pruned before the forward pass:

$$\Delta W_t \leftarrow S(\Delta W_t, \delta_t) \coloneqq \mathrm{sign}(\Delta W_t) \odot \mathrm{ReLU}(|\Delta W_t| - \delta_t)$$

    - During training, δ is governed by its own learning rate and a separate weight decay.
      - Key modification in fine-tuning equivariant MLIPs
      - Initial value of $\delta = 0.001$

Kusupati et al., ICML (2020). Link: https://proceedings.mlr.press/v119/kusupati20a.html

Experimental results: rMD17 dataset [1/6]

- Evaluation on molecular datasets with MACE-OFF23
  - Our method outperforms baselines while updating only a few parameters.
    - See our paper for the full results.

Subset	Metric	Method					
		Zero shot	Scratch	Full	ELoRA	Ours (L)	Ours (H)
Aspirin	Energy	4.72	0.60±0.07	0.19±0.01	0.21±0.01	0.17±0.01	0.20±0.02
	Force	354.42	25.55±0.70	8.09±0.09	8.52±0.08	7.56±0.08	8.22±0.09
	Sparsity	-	-	-	-	83.19 (27.50)	96.84 (86.38)
Benzene	Energy	4.50	0.16±0.07	0.02±0.01	0.02±0.00	0.01±0.00	0.01±0.00
	Force	233.45	7.63±0.14	1.67±0.08	0.94±0.02	0.90±0.04	0.96±0.03
	Sparsity	-	-	-	-	83.50 (28.83)	96.85 (86.40)
Malonaldehyde	Energy	7.47	0.98±0.12	0.20±0.01	0.24±0.01	0.21±0.04	0.22±0.01
	Force	405.37	22.66±0.35	7.91±0.09	8.61±0.06	7.58±0.13	8.03±0.13
	Sparsity	-	-	-	-	83.54 (29.00)	97.39 (88.73)
Paracetamol	Energy	5.33	0.44±0.04	0.12±0.01	0.13±0.01	0.11±0.01	0.13±0.01
	Force	232.60	21.39±0.16	6.74±0.01	6.65±0.03	5.92±0.05	6.53±0.04
	Sparsity	-	-	-	-	83.29 (27.91)	96.92 (86.72)

Experimental results: LAM dataset [2/6]

- Evaluation on inorganic crystals dataset with MACE-MP-0b3
  - Consistent trends observed across inorganic crystal systems.
    - See our paper for the full results.

Subset	Metric	Method					
		Zero shot	Scratch	Full	ELoRA	Ours (L)	Ours (H)
Al∪Mg∪Cu	Energy	238.42	27.90±4.22	1.33±0.14	1.63±0.12	1.83±0.22	2.23±0.21
	Force	43.17	8.23±0.71	5.59±0.03	6.82±0.08	6.10±0.07	6.63±0.11
	Sparsity	-	-	-	-	97.61 (36.59)	99.62 (89.98)
Cu	Energy	359.05	3.31±1.20	0.66±0.01	0.79±0.01	0.65±0.02	0.71±0.04
	Force	50.94	7.08±0.72	3.96±0.00	5.29±0.02	4.32±0.02	4.80±0.02
	Sparsity	-	-	-	-	97.57 (35.38)	99.58 (88.82)
Sn	Energy	31905.00	8.60±1.53	32.74±1.36	9.33±0.30	2.18±0.01	2.49±0.05
	Force	114.22	29.58±0.72	25.82±0.53	32.46±0.64	23.92±0.40	25.39±0.47
	Sparsity	-	-	-	-	97.46 (32.59)	99.42 (84.72)
SSE-PBE	Energy	20.33	0.53±0.03	0.29±0.02	0.41±0.01	0.33±0.03	0.32±0.01
	Force	82.38	12.74±0.04	8.13±0.18	10.85±0.17	8.92±0.28	9.78±0.17
	Sparsity	-	-	-	-	96.77 (14.16)	99.33 (82.12)

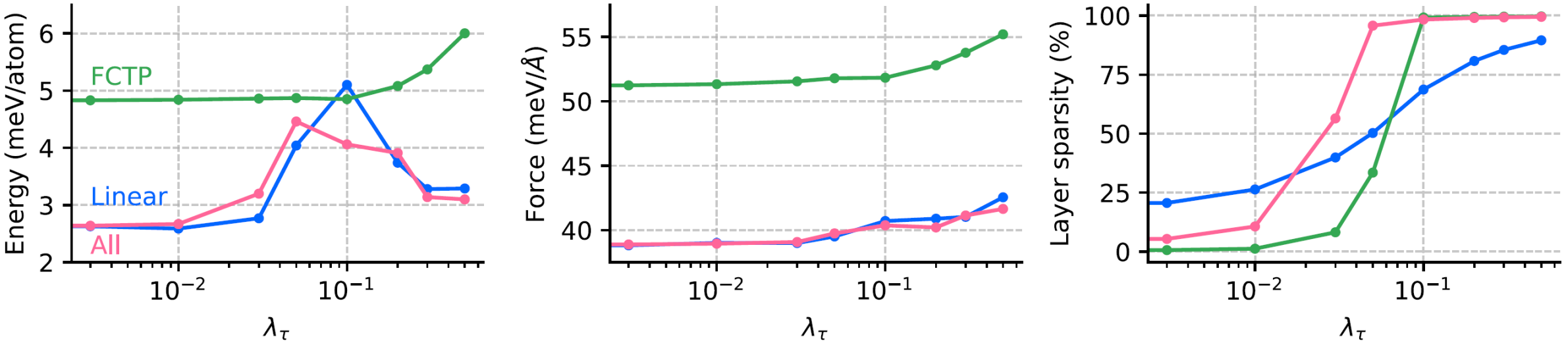
Experimental results: Another backbone [3/6]

- Evaluation with NequIP-OAM-L as the foundation model
  - Our method generalizes across different equivariant architectures.
    - See our paper for the full results.

rMD17		Method Full	Method ELoRA	Method Ours	LAM		Method Full	Method ELoRA	Method Ours
Aspirin	E	0.37±0.01	0.38±0.05	0.33±0.04	Al∪Mg∪Cu	E	1.33±0.16	1.17±0.17	1.10±0.15
	F	16.54±0.88	20.50±0.21	16.54±0.27		F	5.19±0.37	5.22±0.34	5.29±0.35
	Sp.	-	-	98.22 (71.87)		Sp.	-	-	93.90 (3.73)
Benzene	E	0.16±0.06	0.05±0.02	0.05±0.00	Cu	E	0.67±0.06	0.35±0.01	0.49±0.05
	F	9.43±0.28	4.59±0.58	4.26±0.34		F	2.81±0.05	2.76±0.02	3.11±0.06
	Sp.	-	-	98.80 (81.04)		Sp.	-	-	94.23 (8.87)
Malonaldehyde	E	0.35±0.01	0.42±0.10	0.35±0.01	Sn	E	1.55±0.03	1.60±0.06	1.56±0.04
	F	13.52±0.82	14.64±0.42	13.60±0.30		F	20.83±0.52	20.94±0.45	20.08±0.46
	Sp.	-	-	98.40 (74.75)		Sp.	-	-	93.86 (3.14)
Paracetamol	E	0.29±0.00	0.40±0.06	0.25±0.04	SSE-PBE	E	0.21±0.01	0.22±0.02	0.24±0.02
	F	15.13±0.38	21.59±0.87	15.82±0.91		F	6.40±0.21	6.91±0.26	7.58±0.20
	Sp.	-	-	98.26 (72.47)		Sp.	-	-	94.41 (11.81)

Experimental results: Ablation study [4/6]

- Ablation study on the effect of the threshold weight decay (Ti dataset)
  - Updating only the linear layers shows a similar trend to updating all layers.
  - Indicates that most adaptation occurs in linear layers, enabling extremely high sparsity with minimal parameter updates.



Experimental results: Magnetism [5/6]

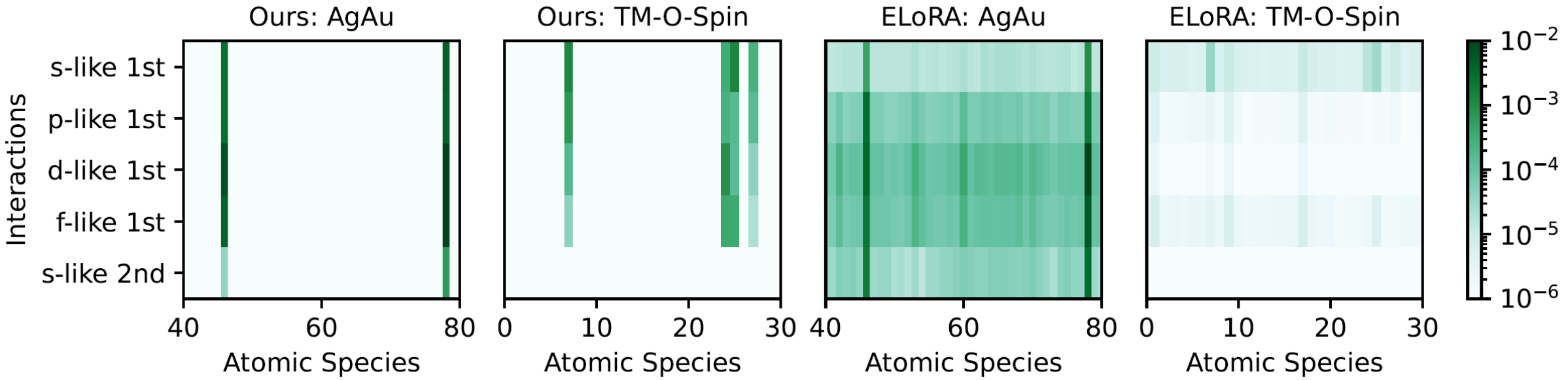
- Extension towards magnetic moment prediction
  - Evaluation on TM-O-Spin dataset (top, custom) and MP-mag dataset (bottom)
  - The proposed method successfully extends non-magnetic MLIPs to magnetic properties.
    - It demonstrates generality for diverse physical property predictions.

Metric	Method				
	Scratch	Full	ELoRA	Ours (L)	Ours (H)
Energy	27.28±6.73	11.58±2.87	12.44±1.88	9.50±0.11	10.57±1.18
Force	174.70±17.68	96.89±47.09	116.31±2.70	48.75±7.68	74.15±33.76
Magnetic moment	0.035±0.013	0.029±0.008	0.038±0.014	0.028±0.007	0.030±0.006
Sparsity (total)	-	-	-	90.21±0.01	91.83±0.01
Sparsity (layer)	-	-	-	43.17±0.40	89.70±0.29

Method	Metric				
	Energy	Force	Magnetic moment	Sparsity (total)	Sparsity (layer)
Full	18.75±0.80	37.56±0.16	0.015±0.000	-	-
ELoRA	17.37±0.81	40.34±0.21	0.013±0.000	-	-
Ours					
└ Linear, $\lambda_\tau = 0.01$	17.56±0.89	42.11±0.35	0.012±0.000	90.15±0.04	41.27±1.08
└ Linear, $\lambda_\tau = 0.3$	18.40±0.86	39.60±0.37	0.014±0.000	91.83±0.01	89.79±0.36
└ All, $\lambda_\tau = 0.05$	16.56±0.71	40.75±0.71	0.011±0.000	30.78±0.12	20.90±0.15
└ All, $\lambda_\tau = 0.3$	17.20±0.75	39.59±0.45	0.013±0.000	87.53±0.04	94.00±0.05

Experimental results: Model interpretation [6/6]

- Model interpretation
  - The proposed method updates only physically meaningful parameters.
  - Sparse fine-tuning for equivariant MLIPs is effective and interpretable for domain-specific calibration.



Conclusion

- Take home message
  - Sparsity-promoting fine-tuning achieves competitive accuracy with minimal parameter updates.
  - The method preserves equivariance naturally and extends to diverse tasks (energy, forces, magnetism).
  - Sparsity patterns reveal physically interpretable signatures (e.g., d-orbital contributions in transition metals).
- Outlook
  - Structured sparse pre-training for efficiency and interpretability

Thank you

- Affiliations
  - KAIST: Youngwoo Cho, Jaegul Choo
  - Seoul National University: Seunghoon Yi, Joonseok Lee
  - KIAS: Wooil Yang, Sungmo Kang, Young-Woo Son
  - Ewha Womans University: Soo Kyung Kim
  - Kangwon National University: Hongkee Yoon

- Code will be available at:
  - https://github.com/Lactobacillus/equivariant-sparse-finetuning