
Supplementary Materials of

DPA-2: Towards a universal large atomic model for molecular and materials simulation

Duo Zhang^{*1,2,3}, Xinzijian Liu^{*1,2}, Xiangyu Zhang^{4,5}, Chengqian Zhang^{2,6}, Chun Cai^{1,2}, Hangrui Bi^{1,2}, Yiming Du^{4,5}, Xuejian Qin^{7,8}, Jiameng Huang^{5,9}, Bowen Li¹⁰, Yifan Shan^{7,8}, Jinzhe Zeng¹¹, Yuzhi Zhang², Siyuan Liu², Yifan Li¹², Junhan Chang^{2,13}, Xinyan Wang², Shuo Zhou^{2,14}, Jianchuan Liu¹⁵, Xiaoshan Luo^{16,17}, Zhenyu Wang^{17,18}, Wanrun Jiang¹, Jing Wu¹⁹, Yudi Yang¹⁹, Jiyuan Yang¹⁹, Manyi Yang²⁰, Fu-Qiang Gong²¹, Linshuang Zhang², Mengchao Shi², Fu-Zhi Dai¹, Darrin M. York¹¹, Shi Liu^{19,22}, Tong Zhu^{10,23,24}, Zhicheng Zhong^{7,8}, Jian Lv¹⁷, Jun Cheng^{21,25,26}, Weile Jia⁴, Mohan Chen^{1,6}, Guolin Ke², Weinan E^{1,27,28}, Linfeng Zhang^{1,2},[†] and Han Wang^{6,29},[‡]

¹AI for Science Institute, Beijing 100080, P. R. China

²DP Technology, Beijing 100080, P. R. China

³Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, P. R. China

⁴State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100871, P. R. China

⁵University of Chinese Academy of Sciences, Beijing 100871, P. R. China

⁶HEDPS, CAPT, College of Engineering, Peking University, Beijing 100871, P. R. China

⁷Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, P. R. China

⁸CAS Key Laboratory of Magnetic Materials and Devices and Zhejiang Province Key Laboratory of Magnetic Materials and Application Technology, Chinese Academy of Sciences, Ningbo 315201, P. R. China

⁹School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, P. R. China

¹⁰Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, P. R. China

¹¹Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA

¹²Department of Chemistry, Princeton University, Princeton, New Jersey 08540, USA

¹³College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P. R. China

¹⁴Yuanpei College, Peking University, Beijing 100871, P. R. China

¹⁵School of Electrical Engineering and Electronic Information, Xihua University, Chengdu, 610039, P. R. China

¹⁶State Key Laboratory of Superhard Materials, College of Physics, Jilin University, Changchun 130012, P. R. China

¹⁷Key Laboratory of Material Simulation Methods & Software of Ministry of Education, College of Physics, Jilin University, Changchun, 130012, P. R. China

¹⁸International Center of Future Science, Jilin University, Changchun, 130012, P. R. China

¹⁹Key Laboratory for Quantum Materials of Zhejiang Province, Department of Physics, School of Science, Westlake University, Hangzhou, Zhejiang 310030, P. R. China

²⁰Atomistic Simulations, Italian Institute of Technology, 16156 Genova, Italy

²¹State Key Laboratory of Physical Chemistry of Solid Surface, iChEM, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, 361005, P. R. China

²²Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310030, P. R. China

²³NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, P. R. China

²⁴Institute for Advanced algorithms research, Shanghai, 201306, P. R. China

²⁵Laboratory of AI for Electrochemistry (AI4EC), IKKEM, Xiamen, 361005, P. R. China

²⁶Institute of Artificial Intelligence, Xiamen University, Xiamen, 361005, P. R. China

²⁷Center for Machine Learning Research, Peking University, Beijing 100871, P. R. China

²⁸School of Mathematical Sciences, Peking University, Beijing, 100871, P. R. China

²⁹Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics, Fenghao East Road 2, Beijing 100094, P. R. China

*These authors contributed equally to this work.

[†]linfeng.zhang.zlf@gmail.com

[‡]wang_han@iapcm.ac.cn

S1 Datasets

The datasets denoted by an asterisk (*) are being published for the first time in this paper. All of 25 datasets, along with the training and test splits for both pre-training and downstream tasks, can be accessed through the AIS-Square platform at <https://www.aissquare.com>.

- Alloy^{*url}** This dataset is generated using the DP-GEN [1] scheme and comprises structure-energy-force-virial data for 53 typical metallic elements (Li, Be, Na, Mg, Al, Si, K, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, Sr, Y, Zr, Nb, Mo, Ru, Rh, Pd, Ag, Cd, In, Sn, La, Hf, Ta, W, Re, Os, Ir, Pt, Au, Pb, Ce, Pr, Nd, Sm, Gd, Tb, Dy, Ho, Er, Tm, Lu) The dataset encompasses a diverse array of crystal configurations, featuring FCC, BCC, and HCP structures, as well as intermetallic compounds and amorphous structures with stochastic vacancies. The dataset contains three categories, random substitutional solid solutions, elementary substances, and intermetallic compounds. All density functional theory (DFT) calculations were conducted using the ABACUS package [2, 3]. The exchange-correlation functional was described by the generalized gradient approximation (GGA) in the Perdew-Burke-Ernzerhof (PBE) form. Norm-conserving pseudopotentials were adopted. The cutoff energy of the plane wave basis was set to be 100 Rydberg, and the Monkhorst-Pack [4] k -point mesh was chosen with a reciprocal space resolution of 0.05 Bohr^{-1} . The self-consistent field iteration stops when the difference in total electron density of consecutive iterations is less than 10^{-6} .
- Cathode^{*url}** This dataset explores O3-type layered oxide cathodes employed in lithium-ion and sodium-ion batteries. It has been generated utilizing the DP-GEN scheme. Specifically, the systems analyzed include Li_xTMO_2 and Na_xTMO_2 , where TM represents transition metal elements including Ni, Mn, Fe, Co, and Cr. The configuration space is explored by NPT MD simulations in a wide range of temperatures and pressures, varying from 50.0 K to 1250.0 K and 0 bar to 3000 bar, respectively. The DFT calculations for this dataset were conducted using the VASP [5, 6] software, incorporating the PBE-GGA functional. The dataset comprises supercells containing twelve formula units for various systems, including LiTMO_2 , NaTMO_2 , $\text{Li}_{0.5}\text{TMO}_2$, $\text{Na}_{0.5}\text{TMO}_2$, and TMO_2 . The dataset is split into two non-overlapping subset, **Cathode-P** and **Cathode-D**, denoting the pre-training and downstream dataset, respectively. The Cathode-P dataset is composed by Li_xTMO_2 with $\text{TM} \in \{\text{Mn, Fe, Co, Cr}\}$, Na_xMnO_2 , and all TMO_2 with $\text{TM} \in \{\text{Ni, Mn, Fe, Co, Cr}\}$, while the Cathode-D dataset includes the Li_xNiO_2 and Na_xTMO_2 with $\text{TM} \in \{\text{Ni, Fe, Co, Cr}\}$.
- Cluster^{*url}** This dataset is composed of metal nano-clusters. The dataset is decomposed in a non-overlapping way into two subsets, **Cluster-P** and **Cluster-D**, which are used for pre-training and downstream tasks, respectively. The Cluster-P dataset encompasses 9 types of clusters that are composed of one element, namely Au, Ag, Al, Cu, Ni, Pt, Pd, Si, and Ru, and 15 types of clusters composed of a combination of 2 elements, namely AgCu, AgNi, AgPd, AgPt, AuAg, AuCu, AuNi, AuPd, AuPt, CuNi, CuPd, CuPt, NiPd, PtNi, and PtPd. The Cluster-D dataset includes 7 types of clusters composed of ternary combination of metal elements, i.e. AgCuPt, AuAgCu, AuAgPd, AuAgPt, AuCuPd, AuCuPt, and PtPdNi. The configurations of the clusters in the Cluster-P and Cluster-D datasets are explored using the DPGEN scheme. The DFT calculations are performed using CP2K [7] with PBE exchange-correlation functional and Grimme D3 dispersion correction [8].
- Drug^{*url}** This dataset is generated using the DPGEN approach and encompasses an extensive collection of over 1.4 million structures, comprising 8 elements H, C, N, O, F, Cl, S, and P, with the inclusion of up to 70 heavy atoms. The foundation for the initial training data was established by optimizing small molecules procured from the ChEMBL [9–11] database with the aid of Gaussian software [12]. To expand the dataset, high-temperature simulations were employed, and the data pool was further augmented by optimizing larger molecules from the ChEMBL database, followed by conducting supplementary simulations. In addition, unoptimized structures were randomly selected and subjected to simulations, resulting in the enlargement of the training set to encompass over 1 million conformations. To ensure comprehensive torsion coverage, structures originating from the ChEMBL torsion scans dataset were optimized and simulated, while enhanced sampling MD simulations

performed with molecules contributed additional structures to the dataset. The whole dataset is employed for **pre-training**.

FerroEle [13]^{url} This dataset comprises 26 ABO₃-type perovskite oxides, which span an extensive composition space containing elements such as Pb, Sr, Ba, Ca, Bi, K, Na, and their various combinations for the A-site, in addition to Ti, Nb, Zr, Mg, Zn, In, Hf, and their respective combinations for the B-site. The configurations of the materials were generated utilizing the DPGEN method and subsequently employed for the training of a universal interatomic potential for perovskite oxides, referred to as UniPero. All DFT calculations were executed with the aid of the ABACUS software [2, 3], utilizing the PBEsol functional within the GGA framework and ONCV multi-projector pseudopotentials. This dataset is divided into four distinct segments according to data complexity, as detailed in [13]. The dataset is divided without overlap into **FerroEle-P** and **FerroEle-D** subsets, used for pre-training and downstream tasks, respectively. The FerroEle-P dataset is composed by BaTiO₃, CaTiO₃, PbTiO₃, SrTiO₃, NaNbO₃, Pb(Zn_{1/3}Nb_{2/3})O₃, Pb(Zr_{1-x}Ti_x)O₃, Pb(Mg_{1/3}Nb_{2/3})O₃, Pb(In_{1/2}Nb_{1/2})O₃, Bi_{1/2}Na_{1/2}TiO₃, K_{1/2}Na_{1/2}NbO₃, Ba_xCa_{1-x}TiO₃, Ba_xSr_{1-x}TiO₃, Ca_xSr_{1-x}TiO₃, Ba_xPb_{1-x}TiO₃, Ca_xPb_{1-x}TiO₃, and Pb_xSr_{1-x}TiO₃. The FerroEle-D dataset contains perovskite oxide binary solid solutions, Bi_{1/2}Na_{1/2}TiO₃-BaTiO₃, NaNbO₃-BaTiO₃, Pb(Mg_{1/3}Nb_{2/3})O₃-BaTiO₃, Pb(Mg_{1/3}Nb_{2/3})O₃-PbTiO₃, Pb(Zn_{1/3}Nb_{2/3})O₃-PbTiO₃, Ba(Zr_{0.2}Ti_{0.8})O₃-Ba_{0.7}Ca_{0.3}TiO₃, Ba(Hf_{0.2}Ti_{0.8})O₃-Ba_{0.7}Ca_{0.3}TiO₃, and ternary solution Pb(In_{1/2}Nb_{1/2})O₃-Pb(Mg_{1/3}Nb_{2/3})O₃-PbTiO₃, Pb(Mg_{1/3}Nb_{2/3})O₃-Pb(Zn_{1/3}Nb_{2/3})O₃-PbTiO₃.

OC2M [14]^{url} This dataset constitutes a subset derived from the Open Catalyst Project’s comprehensive OC20 dataset, which is inclusive of approximately 2 million DFT data samples. This particular dataset comprises 56 distinct elements, with the samples depicting DFT relaxations associated with molecular adsorptions on various surfaces, spanning an extensive structure and chemical space. The principal focus of these samples is directed toward 82 adsorbates that hold significance in the context of renewable energy production and environmental applications. The entire dataset is utilized for **pre-training**.

SSE-PBE [15]^{url} This dataset comprises solid-state electrolytes generated through the DP-GEN method. It is composed of three distinct chemical formulas, namely Li₁₀GeP₂S₁₂, Li₁₀SiP₂S₁₂, and Li₁₀SnP₂S₁₂. All DFT calculations were conducted employing the VASP software [5, 6], with the application of PBE exchange-correlation functional. The Li₁₀GeP₂S₁₂, Li₁₀SiP₂S₁₂ configurations form the **SSE-PBE-P** dataset used for pre-training, while the Li₁₀SnP₂S₁₂ configurations provide the **SSE-PBE-D** dataset for downstream tasks.

SSE-PBESol [15]^{url} This dataset comprises solid-state electrolyte configurations labeled DFT calculations using the PBEsol exchange-correlation functional. Three formulas, namely Li₁₀GeP₂S₁₂, Li₁₀SiP₂S₁₂, and Li₁₀SnP₂S₁₂, are considered, and the configurations were with generated through a DP-GEN procedure that is independent of that used in SSE-PBE. The complete dataset serves for the **downstream** tasks.

SemiCond [16]^{url} This dataset encompasses 20 semiconductors spanning from group IIB to VIA, namely Si, Ge, SiC, BAs, BN, AlN, AlP, AlAs, InP, InAs, InSb, GaN, GaP, GaAs, CdTe, InTe-In₂Te₃, CdSe-CdSe₂, InSe-In₂Se₃, ZnS, CdS-CdS₂. The configurations are explored by the DPGEN scheme in a temperature range of approximately 50.0 K to ~4000 K and a pressure range of circa 1 bar to 50000 bar. DFT calculations, employed during the DP-GEN process, are computed utilizing the ABACUS software package [2, 3]. The energy cutoff of the DFT calculations was set to 100 Ry (~ 1361 eV) and the mesh grid for K-space sampling was 0.08 Bohr⁻¹ (~ 0.15 Å⁻¹). The dataset is divided into two non-overlapping subsets, **SemiCond-P** and **SemiCond-D**, for pre-training and downstream tasks. The Ge and Si are placed in the SemiCond-P dataset. The 11 of the two-element semiconductors, namely AlAs, AlP, BN, CdSe-CdSe₂, CdS-CdS₂, CdTe, GaAs, GaN, GaP, SiC, and ZnS are assigned to the SemiCond-P dataset, while the reset semiconductors, AlN, BAs, InAs, InP, InSb, InSe and InTe are assigned to the SemiCond-D dataset.

H₂O-PD [17]^{url} The water/ice dataset is used to train a DP model for the calculation of the phase diagram of water in the thermodynamic range of 0 to 2400 K and 0 to 50 GPa. The dataset was labeled by the VASP software [5, 6] with the SCAN exchange-correlation functional. The energy cutoff was set to 1500 eV and the spacing of the K-space lattice was 0.5 Å⁻¹. This dataset is used for pre-training.

- AgAu-PBE [18]^{url}** This dataset contains Ag and Au configurations that were generated by the DP-GEN scheme. DFT calculations were conducted employing VASP software [5, 6], in conjunction with PBE functional. The entire dataset is used for **pre-training**.
- AgAu-PBED3 [18]^{url}** This dataset contains Ag, Au, and AgAu alloy configurations that were generated by the DP-GEN scheme. The labels were generated by VASP [5, 6] using the PBE exchange-correlation function in conjugate with the D3 dispersion correction [8]. The entire dataset is used for **downstream** tasks.
- AlMgCu [19]^{url}** This dataset contains unitary, binary, and ternary alloys of Al, Cu, and Mg, i.e. $\text{Al}_x\text{Cu}_y\text{Mg}_z$ with a concentration range of $0 \leq x, y, z \leq 1, x + y + z = 1$. The configurations are explored by the DP-GEN scheme in a temperature range of 50.0 K to 2579.8 K and a pressure range of 1 to 50,000 bar (5 GPa). Energy, force, and virial labels are obtained by DFT calculations adopting the PBE functional using VASP [5, 6]. The energy cut-off for PAW basis sets is 650 eV. K-points are sampled by Monkhorst-Pack mesh with a grid spacing of 0.1 \AA^{-1} . Order 1 Methfessel-Paxton smearing is used with $\sigma = 0.22 \text{ eV}$. SCF convergence criterion for DFT calculation is $1 \times 10^{-6} \text{ eV}$. The dataset is divided into **AlUMgUCu** for pre-training and **AlMgCu-D** for downstream in a mutually exclusive way. The **AlUMgUCu** dataset contains all the unitary systems, i.e. pure Al, Mg, and Cu, while the **AlMgCu-D** contains binary and ternary alloy configurations.
- ANI-1x [20]^{url}** This dataset is generated through iterative active learning to efficiently sample chemical space relevant for machine-learned potentials. It contains over 5 million conformations of organic molecules with up to 13 heavy atoms. Initial data came from small GDB-11 [21, 22] molecules. More complex chemical space was then explored by conformational sampling of progressively larger molecules from databases like ChEMBL and GDB-13 [23]. Techniques included diverse normal mode sampling, trajectory sampling, and dimer sampling to capture molecular diversity. By combining automated active learning of molecules and conformations with rigorous sampling methods, ANI-1x provides comprehensive coverage of organic chemical space. The entire dataset is used for **downstream** tasks.
- Transition-1x [24]^{url}** The dataset comprises over 9.6 million conformations of organic small molecules, spanning more than 10,000 distinct organic chemical reactions. Each reaction involves up to seven heavy atoms, including carbon, nitrogen, and oxygen. Originating from the GDB-7 [25] database, the dataset selects structures that serve as reactants; potential products are then generated via the Growing String Method. Reaction trajectories for these reactant-product pairs are computed employing the Nudged Elastic Band (NEB) method. By selectively sampling structures produced during the NEB procedure and discarding non-physical conformations, the dataset effectively encapsulates the chemical space pertinent to reaction pathways. The entire dataset is used for **downstream** tasks.
- Other datasets** To provide a clearer understanding, detailed descriptions of the datasets can be found in the respective links attached to them. Additional datasets Cu [1]^{url}, Sn [26]^{url}, Ti [27]^{url}, V [28]^{url}, W [29]^{url}, $\text{C}_{12}\text{H}_{26}$ [30]^{url} and HfO_2 [31]^{url} are used for **pre-training**, while In_2Se_3 [32]^{url}, $\text{H}_2\text{O-DPLR}$ [33]^{url}, $\text{H}_2\text{O-SCAN0}$ [34]^{url}, $\text{H}_2\text{O-PBE0TS}$ [35]^{url} and $\text{H}_2\text{O-PBE0TS-MD}$ [36]^{url}, are utilized for **downstream** tasks. Notably, four H_2O downstream datasets ($\text{H}_2\text{O-DPLR}$, $\text{H}_2\text{O-SCAN0}$, $\text{H}_2\text{O-PBE0TS}$ and $\text{H}_2\text{O-PBE0TS-MD}$) significantly differ from the pre-training dataset $\text{H}_2\text{O-PD}$.

S2 Single-task benchmark of the generalizability

The DPA-2 model is trained in a single-task manner on the ANI-1x dataset [20] with a batch size of 20 for ~ 20.2 epochs. The dataset is comprised of $\sim 5.0\text{M}$ training and $\sim 0.5\text{M}$ test data. We have tested this DPA-2 model on the 6 test datasets provided in the Ref. [20], i.e. ANI-MD, DrugBank, GDB07-09, GDB10-13, S66x8, and Tripeptide. The RMSEs of energy and force predictions are reported in Tab. S1 and are compared to the ANI-1x model. The DPA-2 model shows superior accuracy over the ANI-1x model.

We train the GemNet-OC [41], Equiformer V2 [38], Nequip [39], Allegro [40] and DPA-2 models on the pre-training datasets using the single-task training approach, and report the test RMSEs of energy and force in Tab.S2. All the models are trained with a batch size of 1 for 1 million steps. Note that the 1 million training steps are far from enough for a large dataset like OC2M, but it is reasonable

Table S1: The test accuracy of the single-task DPA-2 model compared with ANI-1x [20]. The energy and force test RMSEs in kcal/mol and kcal/mol/Å are respectively presented. The DPA-2 model is trained with a batch size of 20 for ~ 20.2 epochs. The uncertainties of the ANI-1x were estimated by an ensemble of the ANI-1x model, while we have only trained on the DPA-2 model, thus no error uncertainty is provided.

Test set	Energy RMSE [kcal/mol]		Force RMSE [kcal/mol/Å]	
	ANI-1x	DPA-2	ANI-1x	DPA-2
ANI-MD	5.94 ± 1.48	3.31	4.24 ± 0.63	1.42
DrugBank	6.01 ± 3.01	4.22	5.35 ± 1.82	1.56
GDB07-09	1.50 ± 0.05	1.06	3.93 ± 0.08	1.10
GDB10-13	3.21 ± 0.14	1.81	6.01 ± 0.17	1.75
S66x8	3.01 ± 0.18	1.74	2.76 ± 0.30	0.98
Tripeptide	3.77 ± 0.47	2.62	4.79 ± 0.70	1.38

Table S2: The test accuracy of single-task trained models on the pre-training datasets. The energy and force test RMSEs are reported. Four models are compared, the GNO (GemNet-OC [37]), EFV2 (EquiformerV2 [38]), Nequip [39], Allegro [40] and DPA-2. All the models are trained for 1 million steps with a batch size of 1 on a GPU with 32G memory. “OOM” stands for the CUDA out-of-memory error when running the training program. The last row shows the weighted average RMSE (WARMSE) over all the datasets. The weights for the datasets are defined in Tab. 1 of the main text.

Dataset	Energy RMSE [meV/atom]					Force RMSE [meV/Å]				
	GNO	EFV2	Nequip	Allegro	DPA2	GNO	EFV2	Nequip	Allegro	DPA2
Alloy	14.3	8.5	44.0	21.4	16.8	85.1	62.7	175.6	119.4	125.7
Cathode-P	1.5	1.1	14.3	1.0	0.9	17.9	14.9	14.3	24.2	24.5
Cluster-P	47.7	34.6	75.1	54.8	31.5	69.6	104.4	216.6	174.1	126.0
Drug	40.5	29.8	21.6	13.1	12.7	93.6	807.4	187.2	100.8	125.5
FerroEle-P	1.5	1.1	1.1	0.7	0.6	17.9	13.0	23.0	28.6	28.7
OC2M	25.0	6.7	97.4	61.3	36.2	129.1	45.2	226.1	166.8	154.0
SSE-PBE-P	2.7	OOM	1.6	1.0	1.4	8.2	OOM	41.1	47.8	50.3
SemiCond	8.0	3.9	20.5	6.8	5.5	94.4	40.8	180.7	146.8	123.6
H2O-PD	OOM	OOM	0.9	OOM	0.5	OOM	OOM	27.1	OOM	24.7
AgAu-PBE	106.0	23.4	42.3	39.2	2.4	8.0	4.4	43.8	58.9	17.8
AlMgCu	5.9	1.9	38.0	18.3	2.1	9.4	5.7	48.3	40.6	19.1
Cu	6.1	1.7	6.2	1.3	1.2	5.8	3.8	16.7	8.9	8.9
Sn	8.4	5.2	18.2	5.6	4.1	33.7	19.6	62.2	40.2	54.4
Ti	44.5	19.1	27.6	6.9	5.0	87.9	48.6	137.4	85.6	113.1
V	17.9	5.6	8.8	4.2	4.1	79.3	47.4	91.6	82.1	90.8
W	79.1	46.8	20.8	4.0	5.6	81.2	51.3	160.4	101.6	108.1
C12H26	135.8	123.1	121.4	140.4	55.3	518.7	907.4	715.6	648.1	692.5
HfO2	1.2	1.0	1.5	1.4	1.0	16.1	9.1	58.8	64.0	54.2
WARMSE	22.4	14.0	36.3	23.4	13.6	74.6	188.9	142.5	108.4	99.2

Table S3: The test accuracy of the single-task (ST) and multi-task (MT) DPA-2 models on the pre-training datasets. The energy and force test RMSEs are reported. The DPA-2 MT model is trained by 8 GPUs with a total batch size of 8 for 1,000K training steps. Each datum in the mini-batch is randomly sampled from the pre-training dataset with a probability proportional to the weights defined by Tab. 1 in the main text. The DPA-2 ST models are trained with a batch size of 1 by the effective number of training steps calculated as $\text{weight}/13.2 \times 8 \times 1,000\text{K}$.

Dataset	Eff. tr. steps	Energy RMSE [meV/atom]		Force RMSE [meV/Å]	
		DPA-2 ST	DPA-2 MT	DPA-2 ST	DPA-2 MT
Alloy	1212K	16.4	36.5	123.9	169.5
Cathode-P	606K	1.2	3.3	36.6	39.8
Cluster-P	606K	33.2	34.4	169.9	162.5
Drug	1212K	11.9	20.6	117.6	128.9
FerroEle-P	606K	0.9	4.4	34.9	44.2
OC2M	1212K	36.0	29.3	175.0	157.6
SSE-PBE-P	606K	1.5	2.1	54.8	64.0
SemiCond	606K	6.6	6.5	132.8	131.9
H2O-PD	606K	0.6	3.2	31.0	39.7
AgUAu-PBE	121K	16.7	9.4	32.5	28.2
AlUMgUCu	182K	11.4	4.9	33.9	23.4
Cu	61K	3.6	3.6	22.7	18.2
Sn	61K	12.5	24.8	75.6	69.7
Ti	61K	26.2	16.3	158.5	112.4
V	61K	18.9	13.9	133.2	110.2
W	61K	20.6	24.6	176.1	157.9
C12H26	61K	92.5	62.5	894.4	710.6
HfO2	61K	2.4	3.9	107.5	102.8
WARMSE	—	14.9	18.6	111.1	116.3

to fairly compare the performance of the models under a limited time and computational resources budget. The learning rate, energy/force loss prefactors, and all other training hyper-parameters follow the default values provided by corresponding implementation packages. More details on the training settings are found in Secs. S9 and S10. It should be noted that some packages do not handle energy bias based on element type, but rely on an overall mean energy for normalization. Training in this manner can lead to a complete lack of energy convergence in many systems. To aid these models in achieving better convergence, we subtract the element-specific energy bias obtained through least squares fitting during data preprocessing, which is automatically handled internally within the DPA-2 framework. The RMSEs are weighted averaged to obtain the WARMSE score for an easy comparison of the overall accuracy on all the pre-training datasets. Here we used the same weights as those used in the multi-task pre-training, see Tab. 1 in the main text. From Tab. S2, it is observed that the DPA-2 model presents the smallest energy WARMSE, and the second smallest force WARMSE among all the models. It is noted that among the models, DPA-2, Nequip and Allegro are conservative, but GemNet-OC and Equiformer V2 are not.

S3 Multi-task training of the DPA-2 on the pre-training dataset

We train the multi-task DPA-2 model on the pre-training dataset with a batch size of 8 for 1,000K training steps and report the test energy and force RMSEs in Tab. S3. The learning rate undergoes an exponential decay from an initial value of $2e-4$ to a final value of $3e-8$, consistent with the single-task training protocol. Additionally, the gradients across different heads on each GPU are aggregated and averaged. Different pre-training datasets are trained with different weights (different probability to be sampled during training), thus for a fair comparison, the single-task models are trained by an effective number of training steps calculated by $\text{weight}/13.2 \times 8 \times 1,000\text{K}$, where 13.2 is the summation of all the dataset weights. All the single-task models share the same structure and the same learning rate as the multi-task model. The test RMSEs in Tab. S3 are weighted averaged (WARMSE) for a clear comparison over all the pre-training datasets. In terms of training the multi-task is more difficult than the single-task scheme, because the multi-task model should simultaneously fit all the datasets

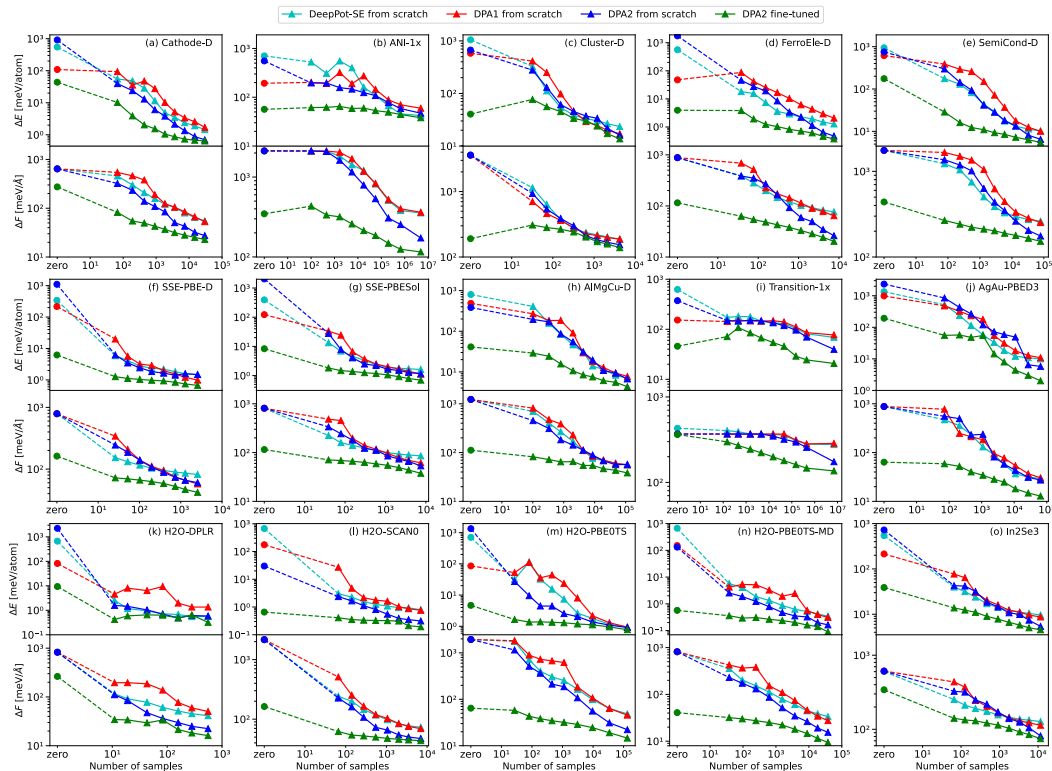


Figure S1: Comparative analysis of sample efficiency on all downstream tasks.

with the same model capacity and number of training steps as the single-task model. It is shown that the force accuracy of the multi-task DPA-2 model is almost the same as the single-task DPA-2 models, while the energy RMSE is roughly 40% higher. We are satisfied with this accuracy on the pre-training dataset. The main advantage of the multi-task model is the stronger generalizability to the datasets that are not explicitly included in the pre-training datasets.

S4 Downstream learning curves

As discussed in Sec. 2.4 of the main text, we conducted a comparative analysis of sample efficiency across all 15 downstream tasks, and illustrate the results in Fig. S1.

S5 Choices of fine-tuning heads

When fine-tuning the DPA-2 for various downstream tasks, we have the option to either select the most task-relevant head from the model’s array of pre-trained heads or randomly choose a head, or even randomly initialize a new head from scratch. In the main text, we opted for the head that closely aligns with the characteristics of each downstream task. To evaluate the robustness of the choice, we investigated the impact of different head selections on sample efficiency using the ANI-1x dataset, as depicted in Fig. S2. The results indicate that, due to the majority of parameters being pre-trained within the descriptor, the choice of fitting head had a minimal influence on the learning curve. The learning curves of Drug and FerroEle-P heads become identical when the sample size is larger than $\sim 10^3$. The learning curve of randomly initialized head converges to the other two heads after $\sim 10^4$. Compared to the total sample size of 4.8×10^6 , the influence of different fitting heads is not important. This observation confirms the resilience of the multi-task pre-training approach.

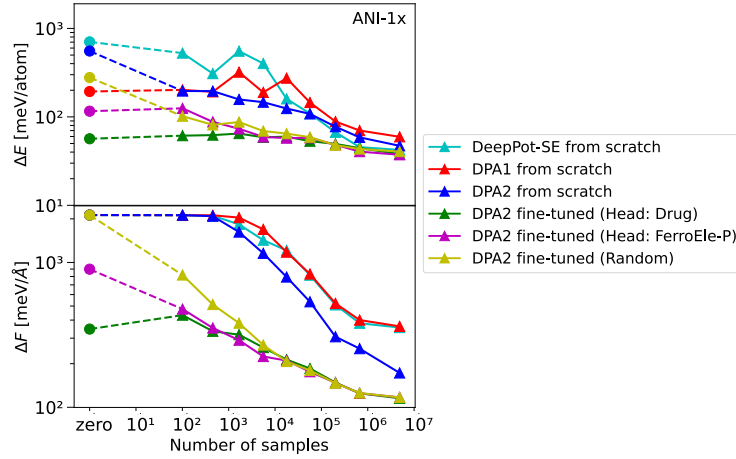


Figure S2: Evaluation of various fitting head selections on the ANI-1x dataset. The inclusion of a dataset name in parentheses indicates the initialization by a specific fitting head that was pre-trained on the corresponding dataset during multi-task pre-training. "Random" denotes the random initialization of the fitting network from scratch.

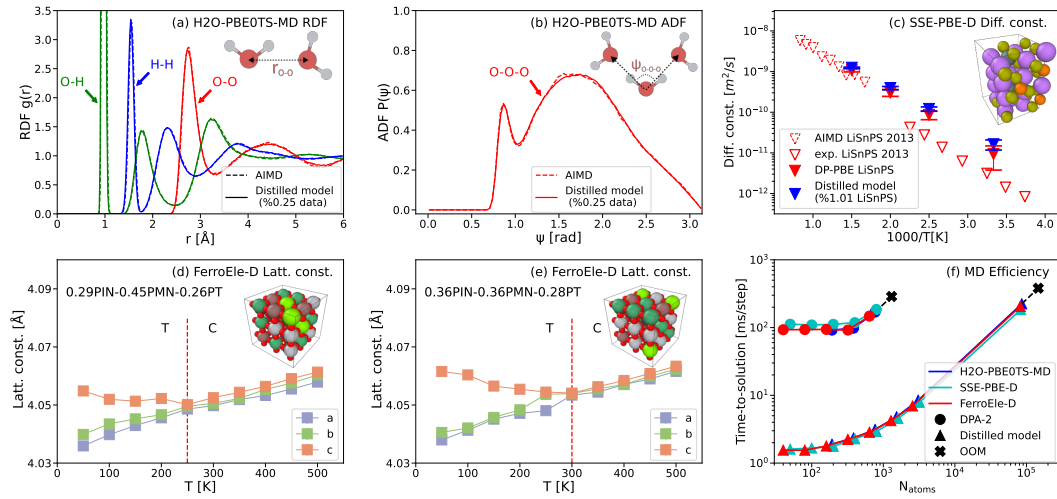


Figure S3: Evaluation of the distilled model across various downstream applications. (a-b) Comparison of the radial distribution function (RDF) and angular distribution function (ADF) for the H2O-PBE0TS-MD dataset between the reference AIMD results [36] and the distilled model. The model is distilled from a DPA-2 model fine-tuned from merely 0.25% of DFT-labeled data. (c) A comparison of diffusion constants for the solid-state electrolyte $\text{Li}_{10}\text{SnP}_2\text{S}_{12}$. The constants were determined using various methods: the distilled model, DPMD as reported in Huang et al. (2021) [15], AIMD simulations from the studies by Mo et al. (2012) and Marcolongo et al. (2017) [42, 43], and experimental findings from solid-state nuclear magnetic resonance (NMR) as documented by Kuhn et al. (2013) [44]. The distilled model is trained from a DPA-2 model fine-tuned by 1.01% of the SSE-PBE-D data. (d-e) The temperature-dependent lattice constants for the ternary solid solution ferroelectric perovskite oxides $\text{Pb}(\text{In}_{1/2}\text{Nb}_{1/2})\text{O}_3\text{-Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3\text{-PbTiO}_3$ (PIN-PMN-PT). The NPT MD simulations using the distilled model are conducted for two concentrations, 0.29PIN-0.45PMN-0.26PT and 0.36PIN-0.36PMN-0.28PT [13]. The model is distilled from a DPA-2 model fine-tuned with the complete FerroEle-P dataset and 7.86% of the FerroEle-D data. (f) Computational efficiency assessment for the aforementioned three systems, showcasing the time-to-solution as a function of the system size in the number of atoms (N_{atoms}).

Table S4: Comparison of the accuracy of few-shot fine-tuned teacher models and student models in the model distillation process across three downstream systems, along with the accuracy of DPA-1 models trained on full downstream data as a benchmark. The units for energy and force RMSEs are expressed in meV/atom and meV/Å, correspondingly.

Dataset	Teacher Model			Student Model		Full-data Model	
	DPA-2 few-shot fine-tuned			DPA-1 (wo attn)		DPA-1 (wo attn)	
	data used (%)	ΔE	ΔF	ΔE	ΔF	ΔE	ΔF
H2O-PBE0TS-MD	0.25	0.3	30.0	0.4	41.7	0.4	37.7
SSE-PBE-D	1.01	1.3	72.0	3.3	101.0	1.9	96.7
FerroEle-D	7.86	1.6	44.7	2.5	99.6	1.9	105.7

Table S5: Ablation study on the components within the repformer layer. This study investigates the impact of sequentially removing the respective components from left to right within the single- and pair-atom channels independently. The table presents the average variations of the test RMSEs across all datasets during the multi-task pre-training process. The units for energy and force RMSEs are meV/atom and meV/Å, respectively. A larger change indicates a more critical contribution to the accuracy of the model.

Structures	Increased RMSEs when removed sequentially →							
single-atom channel	conv		sym_f		sym_g		local_attn	
	ΔE	ΔF	ΔE	ΔF	ΔE	ΔF	ΔE	ΔF
	+5.1	+41.2	+0.7	+10.4	+17.1	+51.3	+14.6	+58.4
pair channel	prod_f		gate		attn			
	ΔE	ΔF	ΔE	ΔF	ΔE	ΔF		
	+4.6	+21.7	+4.7	+34.6	+0.6	+14.1		

S6 Distillation and validation by applications

As outlined in Sec. 2.5 of the main text, we executed model distillation on three representative downstream tasks: H2O-PBE0TS-MD, SSE-PBE-D, and FerroEle-D. This process was conducted in accordance with the methodologies detailed in Sec. 4.6. The main results are reported in Fig. S3. The RMSEs of both the fine-tuned DPA-2 models (teacher models) and the distilled DPA-1 models (student models) on test data are presented in Tab. S4. The DPA-1 models with the same architecture are trained on the full downstream datasets for comparison. As is explained in the main text, in the downstream task of FerroEle-D, the pre-training dataset FerroEle-P is append to the 7.86% subset of FerroEle-D for fine-tuning.

S7 Ablation study

In this section, we carry out an ablation study to investigate the importance of different structures in the repformer layer. To begin with, we introduce some notations to denote specific structures. For the update of the single-atom representation $f_i^{2,l}$, we use **conv** to signify $\frac{1}{N_{rc}^m} \sum_{j \in N_{rc}(i)} w_{ij} g_{ij}^{2,l} f_j^{2,l}$, which is the convolution part in Eq. (25) of the main text. Furthermore, **sym_f** and **sym_g** represent the substructures $\text{symm}(f_j^{2,l}, h_{ij}^{2,l})$ and $\text{symm}(g_{ij}^{2,l}, h_{ij}^{2,l})$ in Eq. (25), respectively. Meanwhile, **local_attn** denotes the local attention operator in Eq. (24). With regard to the update of pair-atom representation $g_{ij}^{2,l}$, **prod_f** stands for the element-wise product $w_{ij} \text{linear}(f_i^{2,l} \odot f_j^{2,l})$ in Eq. (29), and we divide the gated_attn($g_{ij}^{2,l}, h_{ij}$) in Eq. (29) into **gate** and **attn** to represent the $\sum_{\delta} h_{ij,\delta} h_{ik,\delta}$ in Eq. (32) and other parts in Eq. (30)–(32), respectively.

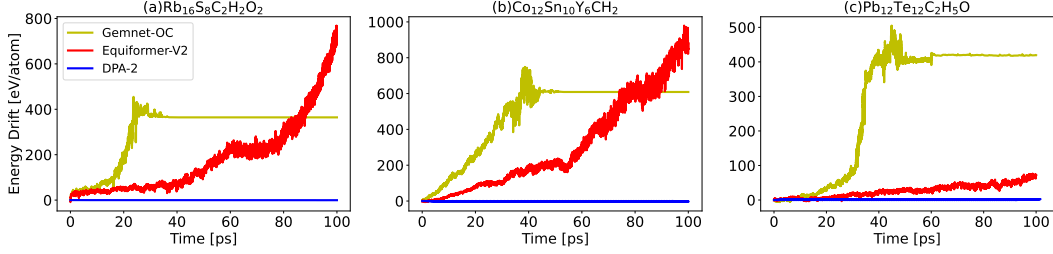


Figure S4: Drift of total energy in 100 ps-long NVE MD simulations for Gemnet-OC, Equiformer-V2, and DPA-2 models, where (a-c) are three randomly selected structures from the OC2M dataset, and the simulation temperature is initialized to 330 K.

We conducted separate ablation studies on the single-atom representation $f_i^{2,l}$ and pair-atom representation $g_{ij}^{2,l}$ within the repformer layer. Tab. S5 presents the results, where we sequentially remove the corresponding structures from left to right and observe the changes in the average RMSE across all datasets during the multi-task pre-training process. It is evident that the removal of certain structures leads to varying degrees of increase in RMSE. For the single-atom representation, all structures except **sym_f** exhibit substantial importance. Conversely, for the pair-atom representation, the **gate** component appears to be more significant than the attention operator **attn** itself.

S8 Energy conservation

To illustrate the energy conservation of the DPA-2 model and the energy drift caused by non-conservative models, we conducted NVE MD simulations on three randomly chosen OC2M structures, as depicted in Fig. S4. We downloaded the GemNet-OC and Equiformer-V2 model checkpoints from their official websites, with reported energy MAEs of 0.286 eV and 0.236 eV, and force MAEs of 0.026 eV/Å and 0.016 eV/Å, respectively. The training parameters for GemNet-OC included a batch size of 16 over 10 million steps, while Equiformer-V2 was trained with a batch size of 4 for 15 million steps. The DPA-2 model is derived from the OC2M single-task training, with the same settings in Sec. S2. It was trained using a batch size of 1 for 1 million steps, which is distant from fair convergence, as indicated by an energy MAE of 0.912 eV and a force MAE of 0.116 eV/Å. This model serves merely as an illustrative case. Despite the high energy and force accuracy of the non-conservative models (Gemnet-OC and Equiformer-V2), their lack of energy conservation results in a significant drift in the total energy during the simulation. In contrast, the DPA-2 simulations do not exhibit any energy drift.

S9 Hyper-parameters of the DPA-2 model

In our study, we employed consistent descriptor configurations for both single-task and multi-task training using the DPA-2 framework. Specifically, we configured the repinit layer with a cutoff radius (r_c^0) of 9.0 Å and a maximum neighbor count in $N_{r_c^0}^m$ of 120. The dimension of the input single-atom representation is 8. The MLP used for embedding $\text{concat}(f_i^0, f_j^0, g_{ij}^0)$ (Eq.(16) in the main text) consists of three layers, each of which has 25, 50, and 100 neurons respectively. The dimension of the pair representations $g_{ij}^{r,t}$ is 100. The first 12 dimensions is preserved in the split operation (Eq. (20)) in the symmetrization operator, so the output dimension of single-atom representation is 1200. For the repformer layers, we chose to use 12 layers, with a cutoff radius (r_c^1) of 4.0 Å and a $N_{r_c^1}^m$ of 40. We designated the dimensions of the single-atom and pair-atom feature representations to be 128 and 32, respectively. Furthermore, within each local multi-head self-attention layer, we set the hidden dimension to 128. For the gated multi-head self-attention layers, we opted for a dimension of 32 and configured both types of layers with four attention heads. The fitting network is comprised of a three-layer MLP with each layer containing 240 neurons.

Regarding the distillation model, we are at liberty to employ parameters specific to downstream tasks. Take the distillation process on the H2O-PBE0TS-MD dataset as an example, we implemented the DPA-1 model architecture excluding the attention layers. We established a cutoff radius (r_c^0) of

6.0 Å and capped the maximum number of neighbors at 120. Additionally, the MLP configuration for the embedding and fitting networks mirrors the settings used in the repinit layer of the DPA-2 framework.

For the single-task model, the total parameter count is approximately 5.15 million, with the descriptor accounting for nearly 5 million parameters and the fitting network utilizing around 0.15 million. In comparison, the multi-task model, trained on 18 pre-training datasets, boasts a total parameter count of about 7.68 million. This total includes 5 million descriptor parameters, augmented by an extra 2.68 million parameters distributed across the 18 fitting networks. The distilled model exhibits a total parameter count of around 0.53 million.

As for our training strategy, we initiated the learning rate at $2e-4$, applying an exponential decay after every 5,000 steps, and ultimately reducing it to $3e-8$ upon reaching 1 million training steps. The prefactors for both energy and force were adjusted in conjunction with the learning rate. Specifically, the energy prefactor was scaled from 0.02 to 1, while the force prefactor underwent a change from 1,000 to 1.

S10 Training details of GemNet-OC/EquiformerV2/Allegro/Nequip

For GemNet-OC and EquiformerV2, the training code is from the main branch of the GitHub repository <https://github.com/Open-Catalyst-Project/ocp> (commit hash: 9bc9373), and the training parameters are from the file `s2ef/all/gemnet/gemnet-dT.yml` and `configs/s2ef/2M/equiformer_v2/equiformer_v2_N@12_L@6_M@2.yml` in the repository except for the batch size set to 1 for a fair comparison. For Allegro, the training code is from the main branch of the GitHub repository <https://github.com/mir-group/allegro> (commit hash: 22f673c), and the training parameters are from the file `configs/example.yaml` in the repository. The batch size is also set to 1, and the learning rate is set to 0.005 which is tested to be optimal when 1 is used for batch size. For Nequip, the training code is from the main branch of the GitHub repository <https://github.com/mir-group/nequip> (commit hash: dceaf49, tag: v0.5.6), and the training parameters are from the file `configs/example.yaml` in the repository. The batch size is also set to 1, and the learning rate is set to 0.002 which is tested to be optimal when 1 is used for batch size.

For the datasets we use, the contributions to the total energy of different types of atoms may differ significantly, which leads to a large variance in energy. Training on the dataset directly with GemNet-OC, EquiformerV2, Allegro or Nequip will cause the energy loss to remain high. We preprocess the dataset by subtracting the energy bias for each atom from the total energy, where the energy bias is determined by a least-square fitting of the energies in the training data. For GemNet-OC, systems with fewer than 3 atoms are excluded otherwise errors will occur during the training process. For Allegro, systems with fewer than 2 atoms are excluded otherwise Allegro will complain about the lack of neighbors. For Nequip, as 4.0 is used as the cutoff radius, systems with a distance between the two closest atoms larger than 4.0 are excluded to avoid errors. None of GemNet-OC, EquiformerV2, Allegro, or Nequip natively support non-periodic boundary condition systems as training data. For such systems, we apply a cubic box with a side length greater than the distance between the two furthest atoms in the system plus the cutoff radius. The maximum memory required by the Allegro and Nequip programs during the data processing and training process increases as the dataset size increases. For the OC2M dataset, Allegro and Nequip are unable to complete the data processing on a machine with 155GB of memory. We divide the OC2M dataset randomly into 4 subsets and train Allegro or Nequip on them sequentially.

References

- [1] Yuzhi Zhang, Haidi Wang, Weijie Chen, Jinzhe Zeng, Linfeng Zhang, Han Wang, and Weinan E. Dp-gen: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Computer Physics Communications*, 253:107206, 2020.
- [2] Mohan Chen, GC Guo, and Lixin He. Systematically improvable optimized atomic basis sets for ab initio calculations. *Journal of Physics: Condensed Matter*, 22(44):445501, 2010.
- [3] Pengfei Li, Xiaohui Liu, Mohan Chen, Peize Lin, Xinguo Ren, Lin Lin, Chao Yang, and Lixin He. Large-scale ab initio simulations based on systematically improvable atomic basis. *Computational Materials Science*, 112:503–517, 2016.

- [4] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.
- [5] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science*, 6(1):15–50, 1996.
- [6] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.
- [7] Thomas D Kühne, Marcella Iannuzzi, Mauro Del Ben, Vladimir V Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Z Khaliullin, Ole Schütt, Florian Schiffmann, et al. Cp2k: An electronic structure and molecular dynamics software package-quickstep: Efficient and accurate electronic structure calculations. *The Journal of Chemical Physics*, 152(19):194103, 2020.
- [8] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of chemical physics*, 132(15), 2010.
- [9] Simon Jupp, James Malone, Jerven Bolleman, Marco Brandizi, Mark Davies, Leyla Garcia, Anna Gaulton, Sebastien Gehant, Camille Laibe, Nicole Redaschi, et al. The ebi rdf platform: linked open data for the life sciences. *Bioinformatics*, 30(9):1338–1339, 2014.
- [10] A Patrícia Bento, Anna Gaulton, Anne Hersey, Louisa J Bellis, Jon Chambers, Mark Davies, Felix A Krüger, Yvonne Light, Lora Mak, Shaun McGlinchey, et al. The chembl bioactivity database: an update. *Nucleic acids research*, 42(D1):D1083–D1090, 2014.
- [11] Mark Davies, Michał Nowotka, George Papadatos, Francis Atkinson, Gerard JP Van Westen, Nathan Dedman, Rodrigo Ochoa, and John P Overington. Mychembl: a virtual platform for distributing cheminformatics tools and open data. *Challenges*, 5(2):334–337, 2014.
- [12] MJ Frisch, GW Trucks, H Bernhard Schlegel, GE Scuseria, MA Robb, JR Cheeseman, G Scalmani, VPGA Barone, GA Petersson, HJRA Nakatsuji, et al. Gaussian 16, 2016.
- [13] Jing Wu, Jiyuan Yang, Yuan-Jinsheng Liu, Duo Zhang, Yudi Yang, Yuzhi Zhang, Linfeng Zhang, Shi Liu, et al. Universal interatomic potential for perovskite oxides. *Physical Review B*, 108(18):L180104, 2023.
- [14] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *ACS Catalysis*, 11(10):6059–6072, 2021.
- [15] Jianxing Huang, Linfeng Zhang, Han Wang, Jinbao Zhao, Jun Cheng, and Weinan E. Deep potential generation scheme and simulation protocol for the li10gep2s12-type superionic conductors. *The Journal of Chemical Physics*, 154(9):094703, 2021.
- [16] Jianchuan Liu, Xingchen Zhang, Yuzhi Zhang, Duo Zhang, Linfeng Zhang, and Mohan Chen. Machine-learning-based interatomic potentials for group iib to via semiconductors: A comparative study of universal and independent models. *arXiv preprint arXiv:2311.11305*, 2023.
- [17] Linfeng Zhang, Han Wang, Roberto Car, and Weinan E. Phase diagram of a deep potential water model. *Physical review letters*, 126(23):236001, 2021.
- [18] YiNan Wang, LinFeng Zhang, Ben Xu, XiaoYang Wang, and Han Wang. A generalizable machine learning potential of ag–au nanoalloys and its application to surface reconstruction, segregation and diffusion. *Modelling and Simulation in Materials Science and Engineering*, 30(2):025003, 2021.
- [19] Wanrun Jiang, Yuzhi Zhang, Linfeng Zhang, and Han Wang. Accurate deep potential model for the al–cu–mg alloy in the full concentration space. *Chinese Physics B*, 30(5):050706, 2021.
- [20] Justin S Smith, Ben Nebgen, Nicholas Lubbers, Olexandr Isayev, and Adrian E Roitberg. Less is more: Sampling chemical space with active learning. *The Journal of chemical physics*, 148(24):241733, 2018.

- [21] Tobias Fink, Heinz Bruggesser, and Jean-Louis Reymond. Virtual exploration of the small-molecule chemical universe below 160 daltons. *Angewandte Chemie International Edition*, 44(10):1504–1508, 2005.
- [22] Tobias Fink and Jean-Louis Reymond. Virtual exploration of the chemical universe up to 11 atoms of c, n, o, f: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. *Journal of chemical information and modeling*, 47(2):342–353, 2007.
- [23] Lorenz C Blum and Jean-Louis Reymond. 970 million druglike small molecules for virtual screening in the chemical universe database gdb-13. *Journal of the American Chemical Society*, 131(25):8732–8733, 2009.
- [24] Mathias Schreiner, Arghya Bhowmik, Tejs Vegge, Jonas Busk, and Ole Winther. Transition1x-a dataset for building generalizable reactive machine learning potentials. *Scientific Data*, 9(1):779, 2022.
- [25] Lars Ruddigkeit, Ruud Van Deursen, Lorenz C Blum, and Jean-Louis Reymond. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.
- [26] Tao Chen, Fengbo Yuan, Jianchuan Liu, Huayun Geng, Linfeng Zhang, Han Wang, and Mohan Chen. Modeling the high-pressure solid and liquid phases of tin from deep potentials with ab initio accuracy. *Physical Review Materials*, 7(5):053603, 2023.
- [27] Tongqi Wen, Rui Wang, Lingyu Zhu, Linfeng Zhang, Han Wang, David J Srolovitz, and Zhaoxuan Wu. Specialising neural network potentials for accurate properties and application to the mechanical response of titanium. *npj Computational Materials*, 7(1):206, 2021.
- [28] Rui Wang, Xiaoxiao Ma, Linfeng Zhang, Han Wang, David J Srolovitz, Tongqi Wen, and Zhaoxuan Wu. Classical and machine learning interatomic potentials for bcc vanadium. *Physical Review Materials*, 6(11):113603, 2022.
- [29] Xiaoyang Wang, Yinan Wang, Linfeng Zhang, Fuzhi Dai, and Han Wang. A tungsten deep neural-network potential for simulating mechanical property degradation under fusion service environment. *Nuclear Fusion*, 2022.
- [30] Jinzhe Zeng, Linfeng Zhang, Han Wang, and Tong Zhu. Exploring the chemical space of linear alkane pyrolysis via deep potential generator. *Energy & fuels*, 35(1):762–769, 2020.
- [31] Jing Wu, Yuzhi Zhang, Linfeng Zhang, and Shi Liu. Deep learning of accurate force field of ferroelectric hfo 2. *Physical Review B*, 103(2):024108, 2021.
- [32] Jing Wu, Liyi Bai, Jiawei Huang, Liyang Ma, Jian Liu, and Shi Liu. Accurate force field of two-dimensional ferroelectrics from deep learning. *Physical Review B*, 104(17):174107, 2021.
- [33] Linfeng Zhang, Han Wang, Maria Carolina Muniz, Athanassios Z Panagiotopoulos, Roberto Car, et al. A deep potential model with long-range electrostatic interactions. *The Journal of Chemical Physics*, 156(12), 2022.
- [34] Chunyi Zhang, Fujie Tang, Mohan Chen, Jianhang Xu, Linfeng Zhang, Diana Y Qiu, John P Perdew, Michael L Klein, and Xifan Wu. Modeling liquid water by climbing up jacob’s ladder in density functional theory facilitated by using deep neural network potentials. *The Journal of Physical Chemistry B*, 125(41):11444–11456, 2021.
- [35] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and Weinan E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical review letters*, 120(14):143001, 2018.
- [36] Robert A DiStasio, Biswajit Santra, Zhaofeng Li, Xifan Wu, and Roberto Car. The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *The Journal of chemical physics*, 141(8), 2014.

- [37] Johannes Gasteiger, Muhammed Shuaibi, Anuroop Sriram, Stephan Günnemann, Zachary Ulissi, C Lawrence Zitnick, and Abhishek Das. How do graph networks generalize to large and diverse molecular systems? *arXiv preprint arXiv:2204.02782*, 2022.
- [38] Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.
- [39] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):1–11, 2022.
- [40] Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *arXiv preprint arXiv:2204.05249*, 2022.
- [41] Johannes Gasteiger, Muhammed Shuaibi, Anuroop Sriram, Stephan Günnemann, Zachary Ulissi, C Lawrence Zitnick, and Abhishek Das. Gemnet-oc: developing graph neural networks for large and diverse molecular simulation datasets. *arXiv preprint arXiv:2204.02782*, 2022.
- [42] Yifei Mo, Shyue Ping Ong, and Gerbrand Ceder. First principles study of the $\text{Li}_{10}\text{GeP}_2\text{S}_{12}$ lithium super ionic conductor material. *Chemistry of Materials*, 24(1):15–17, 2012.
- [43] Aris Marcolongo and Nicola Marzari. Ionic correlations and failure of nernst-einstein relation in solid-state electrolytes. *Physical Review Materials*, 1(2):025402, 2017.
- [44] Alexander Kuhn, Jürgen Köhler, and Bettina V Lotsch. Single-crystal x-ray structure analysis of the superionic conductor $\text{Li}_{10}\text{GeP}_2\text{S}_{12}$. *Physical Chemistry Chemical Physics*, 15(28):11620–11622, 2013.