

Supplementary Information (ESI) for

Machine-learning-assisted discovery of a stable Li_3As_2 intermediate phase in the Li–As binary system and its electrochemical implications

Shuaishuai Ge, Weiduo Zhu,* and Haidi Wang*

School of Physics, Hefei University of Technology, Hefei, Anhui 230009, China

*Corresponding authors: weiduo@hfut.edu.cn; haidi@hfut.edu.cn

Table S1 Composition and sampling details of the training dataset for the machine-learning inter-atomic potential (MLIP).

Subset ID	Configuration type	Generation details (phases)	T (K)	N atoms	Raw frames	Final dataset
Perfect	Bulk crystalline snapshots (AIMD)	All parent phases	300–1200	16–144	362,265	788
Vacancy	Li-vacancy snapshots (AIMD)	Li2As_Fd-3m, Li2As_P1, Li3As_P63-mm, Li3As_Fm-3m, Li3As_I4-mmm, Li3As_Pm-3m, LiAs_P21-c, LiAs_R-3m, Li4As_Ccca, Li5As4_C2-m, Li31As20_I4-mcm	600–900	14–143	411,777	1128
Interstitial	Li-interstitial snapshots (AIMD)	Li2As_Fd-3m, Li2As_P1, Li3As_P63-mm, Li3As_Fm-3m, Li3As_I4-mmm, Li3As_Pm-3m, LiAs_P21-c, LiAs_R-3m, Li4As_Ccca, Li5As4_C2-m	600–900	17–145	108,534	155
Surface	Slab snapshots (AIMD)	Li3As_P63-mm, Li3As_Fm-3m, Li3As_I4-mmm, Li3As_Pm-3m + LiAs_P21-c, LiAs_P4-mmm, LiAs_Pmma, LiAs_C2-m, LiAs_R-3m, LiAs_Fm-3m, LiAs_F-43m, LiAs_P-6m2 (Miller indices: 100/110/111)	600	64–144	50,105	345
Melt	Liquid / melt snapshots (AIMD)	Generated from Li3As_P63-mm, LiAs_P21-c, LiAs5_P6-mmm	2000	48–96	8,000	46
Strain	Strained / displaced structures	Thirty-three known parent phases subjected to normal and shear strains of $\pm 1\%$, $\pm 3\%$, and $\pm 5\%$ (with symmetry preserved), combined with random atomic displacements of 0.01 Å	0	16–144	1000	1000

Note: For all ab initio molecular dynamics (AIMD)-derived subsets, preliminary sampling was performed every 15 steps after discarding the first 1000 molecular dynamics (MD) steps. Structural deduplication and dimensionality reduction were then carried out using smooth overlap of atomic positions (SOAP) descriptors ($r_{cut} = 6$, $n_{max} = 8$, $l_{max} = 6$) and the farthest-point sampling (FPS) algorithm, targeting approximately 2500 frames in total. These structures were combined with 1000 statically perturbed elastic configurations to construct a representative training set of approximately 3500 configurations. High-accuracy density functional theory (DFT) single-point calculations were subsequently performed for all selected structures to obtain energies, forces, and stresses, using the following key VASP settings: PREC=Accurate, ENCUT=650, EDIFF=1e-08, LREAL=False, IBRION=-1, NSW=0, and ISIF=2.

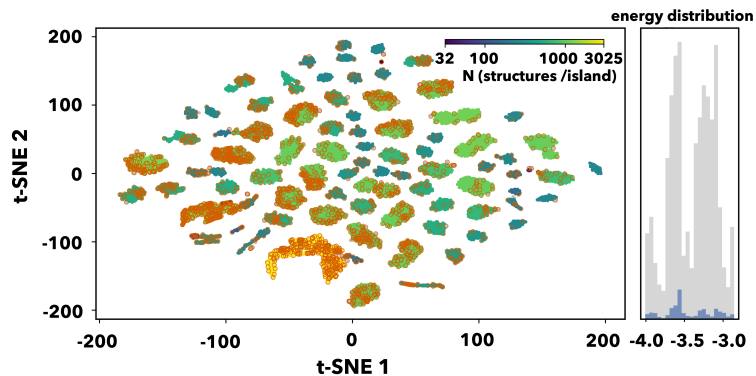


Fig. S1 Energy coverage and structural diversity of the dataset. The main panel shows the two-dimensional t-distributed stochastic neighbor embedding (t-SNE) of the structural descriptors. Data points are colored by the number of structures within each local cluster (N structures/island), and the representative configurations selected for the final training set are marked by red open circles. The right inset shows the histogram of the per-atom energy distribution for all sampled configurations (gray) and the selected training-set configurations (blue).

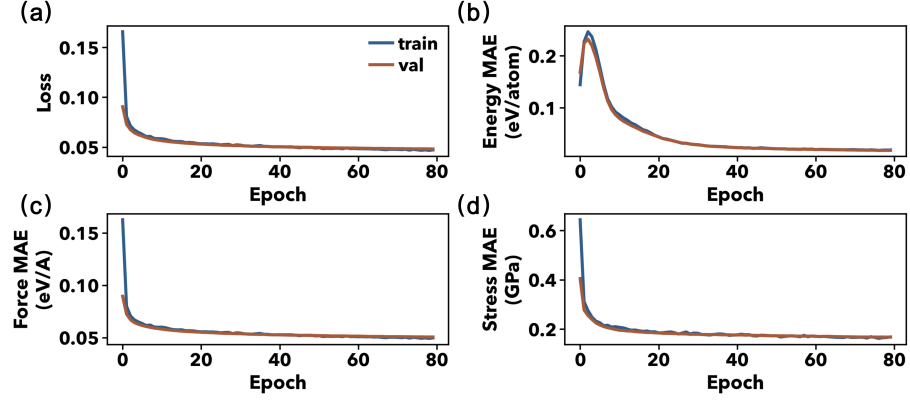


Fig. S2 Training convergence of the machine-learning interatomic potential. Evolution of the training and validation errors during the first 80 epochs of the 180-epoch fine-tuning run: (a) total loss; (b) energy mean absolute error (MAE) normalized per atom (eV atom^{-1}); (c) force MAE (eV \AA^{-1}); and (d) stress MAE (GPa).

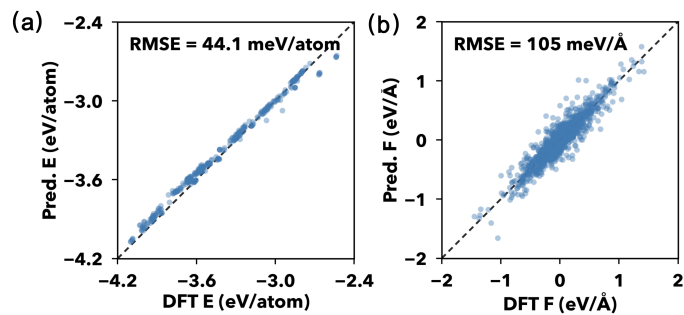


Fig. S3 Predictive performance of the base MatterSim model on the independent test set. (a,b) Diagonal parity plots comparing the energy and force predictions of the unfine-tuned base model with the DFT labels, respectively.

Table S2 Comparison of lattice parameters and elastic constants predicted by density functional theory (DFT) and MatterSim for the stable Li_3As_2 , Li_3As , and LiAs phases.

Phase	Variant	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	V (Å ³)	B_H (GPa)	G_H (GPa)	E (GPa)	ν
Li_3As_2	DFT (TOTAL)	5.08	5.08	6.77	100.55	79.45	88.97	168.41	39.18	24.40	60.61	0.24
	MatterSim	5.21	5.21	6.53	98.35	81.65	90.68	173.39	42.35	22.40	57.13	0.28
Li_3As ($P6_3/mmc$)	DFT (TOTAL)	4.36	4.36	7.78	90.00	90.00	120.00	128.15	35.36	30.47	71.01	0.17
	MatterSim	4.36	4.36	7.78	90.00	90.00	120.00	128.17	34.98	31.49	72.67	0.15
LiAs ($P2_1/c$)	DFT (TOTAL)	5.80	5.27	10.79	90.00	117.97	90.00	291.30	32.05	16.93	43.19	0.28
	MatterSim	5.93	5.29	10.38	90.00	116.58	90.00	290.72	47.57	20.03	52.69	0.32

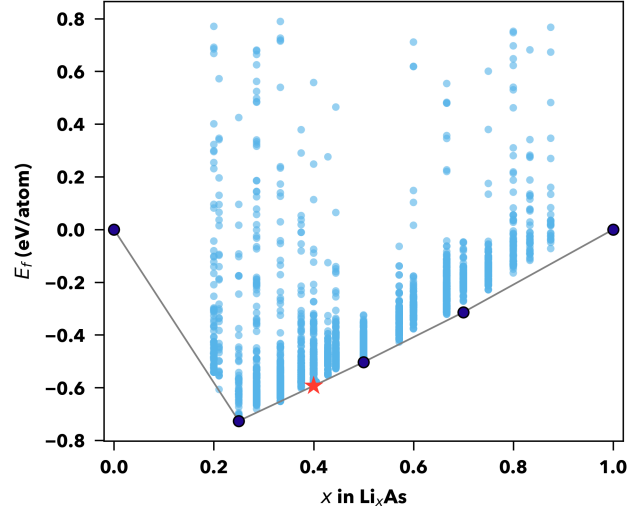


Fig. S4 Li–As composition–formation-energy distribution and 0 K convex hull of 2304 CALYPSO-generated candidate structures after DFT relaxation.

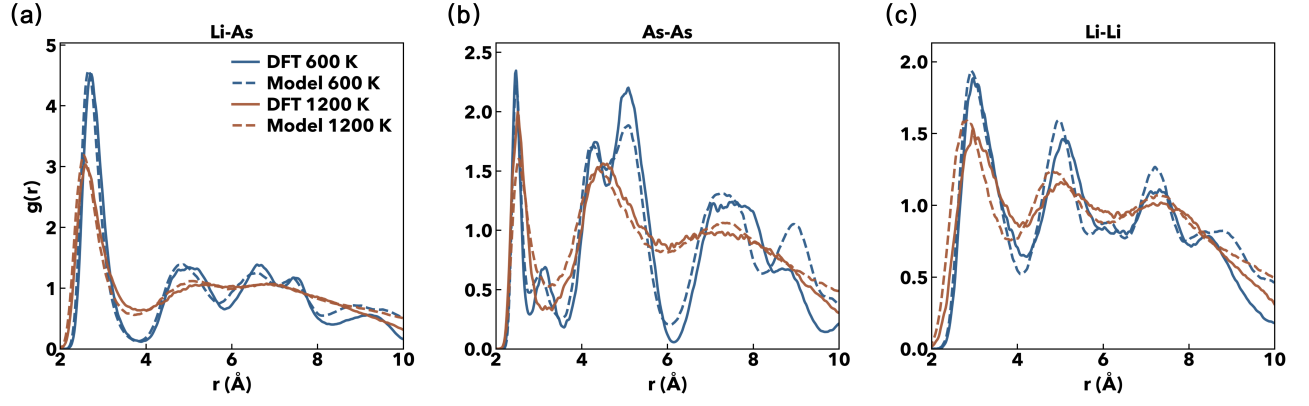


Fig. S5 Structural validation of the fine-tuned (FT) MatterSim potential. Radial distribution functions $g(r)$ of (a) Li-As, (b) As-As, and (c) Li-Li pairs obtained from AIMD and the FT MatterSim model at 600 K and 1200 K.

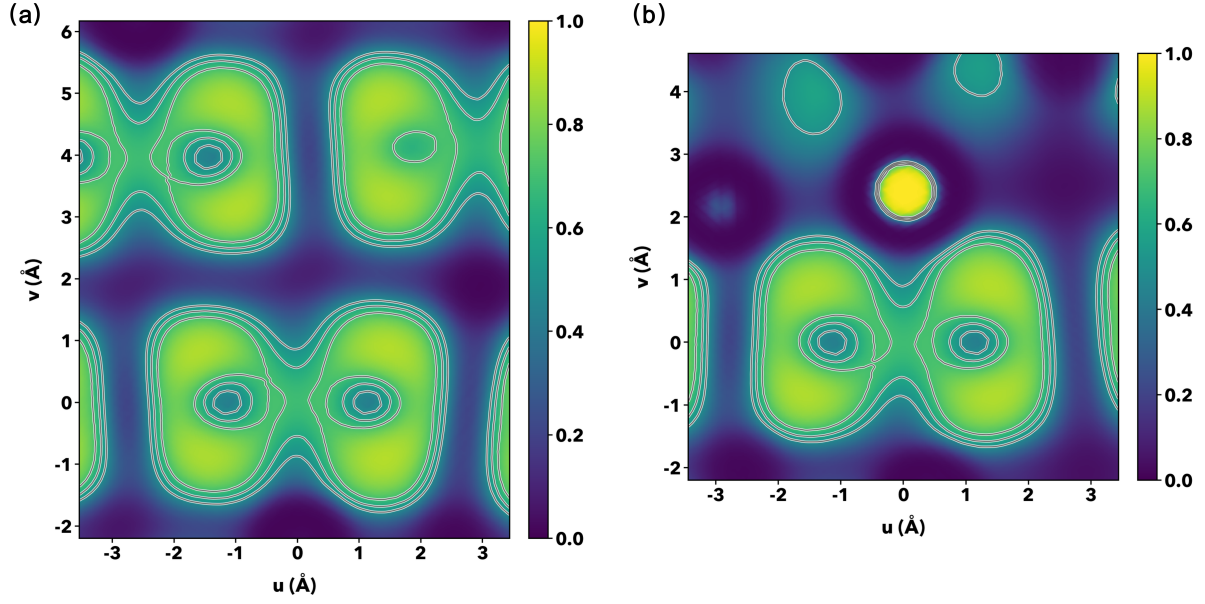


Fig. S6 Electron localization function (ELF) maps in planes defined by the shortest As–As bond and a third atom. White and black contours correspond to $\text{ELF} = 0.5, 0.6$, and 0.7 . (a) Plane defined by the shortest As–As bond and an As or Li atom near the bond midpoint. (b) Plane defined by the shortest As–As bond and the nearest Li atom.

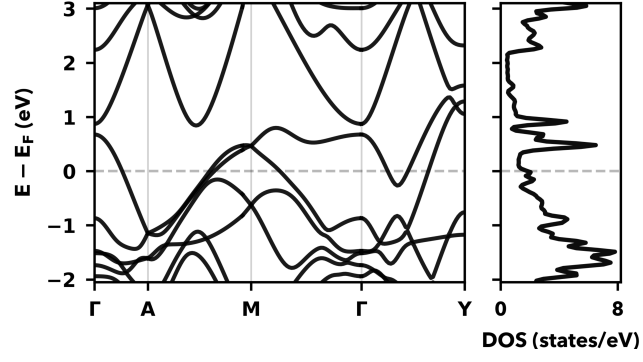


Fig. S7 DFT-calculated electronic band structure and density of states (DOS) of $C2/c$ - Li_3As_2 .

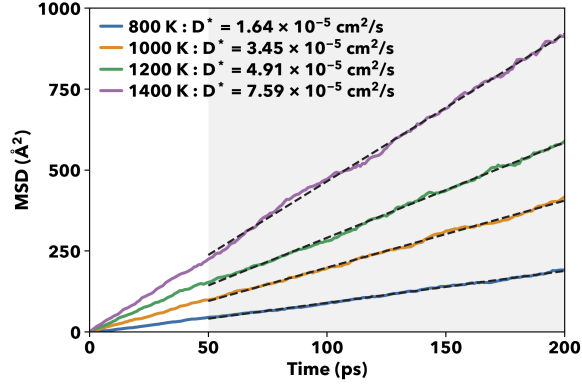


Fig. S8 Li-ion MSD curves and linear fits for $C2/c\text{-Li}_3\text{As}_2$ at different temperatures. The MSD profiles were obtained from MatterSim-MD trajectories, and the 50–200 ps lag-time window was used to fit the diffusive regime. The Li-ion self-diffusion coefficient D^* was calculated from the fitted MSD slope using the three-dimensional Einstein relation, $D^*(\text{cm}^2 \text{s}^{-1}) = \text{slope}(\text{\AA}^2 \text{ps}^{-1})/6 \times 10^{-4}$.