Supplementary Information

Molecular dynamics simulation of nitrogen diffusion in iron and iron nitrides using ab initio data trained machine learning potentials

Peijie Feng[1], Aditya Dilip Lele[2,3], Minhyeok Lee[1], Yiguang Ju[3,4], Yuji Suzuki[1*]

1 Department of Mechanical Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

2 Department of Mechanical Engineering, Rowan University
Glassboro, New Jersey, 08028, USA

3 Mechanical & Aerospace Engineering, Princeton University,
Princeton, New Jersey, 08544, USA

4 Institute of Fluid Science, Tohoku University,
Sendai, Miyagi 980-8576, Japan

*Corresponding Author
Yuji Suzuki

Department of Mechanical Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
Phone: +81-3-5841-6411
E-mail: ysuzuki@mesl.t.u-tokyo.ac.jp

Table S1: MLP training and validation AIMD snapshots count for thermalized systems of iron and iron nitride.

ID		Number of atoms	Number of snapshots for Training/Validation		
			AIMD @ 500 [K]	AIMD @ 1000 [K]	AIMD @ 1500 [K]
01	Fe (mp-13)	250	160/40	160/40	160/40
02	Fe (mp-150)	256	160/40	160/40	160/40
03	Fe8N (mp-555)	486	160/40	160/40	160/40
04	Fe4N (mp-535)	320	160/40	160/40	160/40
05	Fe3N (mp-1804)	512	160/40	160/40	160/40
06	Fe12N5 (mp-27908)	408	160/40	160/40	160/40
07	Fe2N (mp-248)	243	160/40	160/40	160/40
08	Fe2N (mp-21476)	216	160/40	160/40	160/40
09	FeN (mp-12120)	128	160/40	160/40	160/40
10	FeN (mp-6988)	512	160/40	160/40	160/40

Table S2: MLP training and validation AIMD snapshots count for systems with nitrogen atoms addition and removal.

ID		Nitrogen atoms Add/Removal	Number of atoms	AIMD @1000 [K] Number of snapshots for Training/Validation
01	Fe (mp-13)	+1	251	160/40
		+2	252	160/40
02	Fe (mp-150)	+2	258	160/40
		+4	260	160/40
03	Fe8N (mp-555)	+2	488	160/40
		–2	484	160/40
04	Fe4N (mp-535)	+2	322	160/40
		–2	318	160/40
05	Fe3N (mp-1804)	+2	218	160/40
		–2	514	160/40
06	Fe12N5 (mp-27908)	+4	412	160/40
		–4	404	160/40
07	Fe2N (mp-248)	+2	241	160/40
		–2	245	160/40

Table S3: MLP training and validation AIMD snapshots count for systems with cell size expansion and compression.

ID		Expansion or Compression	Number of atoms	AIMD @1000 [K] Number of snapshots for Training/Validation
01	Fe (mp-13)	+5%	250	80/20
		−5%	250	80/20
02	Fe (mp-150)	+5%	256	80/20
		−5%	256	80/20
03	Fe8N (mp-555)	+5%	486	80/20
		−5%	486	80/20
04	Fe4N (mp-535)	+5%	320	80/20
		−5%	320	80/20
05	Fe3N (mp-1804)	+5%	218	80/20
		−5%	514	80/20
06	Fe12N5 (mp-27908)	+3%	412	80/20
		−5%	404	80/20
07	Fe2N (mp-248)	+5%	241	80/20
		−5%	245	80/20
08	Fe2N (mp-21476)	+5%	216	80/20
		−5%	216	80/20
09	FeN (mp-12120)	+5%	128	80/20
		−5%	128	80/20
10	FeN (mp-6988)	+5%	512	80/20
		−5%	512	80/20

Statistical uncertainty of MSD-based diffusivities and sampling limits.

We estimate the self-diffusion coefficient D_{self} from the long-time slope of the mean-squared displacement (MSD) via the Einstein relation in three dimensions,

$$\langle \Delta r^2(t) \rangle = 6 D_{self} t$$

Here the angle brackets denotes an average over diffusion atoms and over time. t is the lag time.

So that the self-diffusion coefficient D_{self} is obtained from a linear fit of $\langle \Delta r^2(t) \rangle$ over a suitable chosen lag-time window.

At low temperatures, diffusion events (interstitial hops) becomes rare on MD timescales, and statistical noise in the MSD slope dominates the uncertainty in the estimated D_{self}. We quantify this sampling limit by interpret diffusion as a lattice random walk with effective jump length a, jump frequency per atom Γ,

The standard jump-diffusion relation is

$$D_{self} = \frac{1}{6} \Gamma a^2$$

$$\Gamma = \frac{6 D_{self}}{a^2}$$

The mean waiting time between successive hops of a given atom is then

$$\tau_{hop} = \Gamma^{-1} = \frac{a^2}{6 D_{self}}$$

Over a trajectory of length t_{traj} with N diffusion atoms, the expected total number of hops observed in the whole system is

$$E(N_{hops}) \approx \frac{N \cdot t_{traj}}{\tau_{hop}} = \frac{6 N D_{self} \cdot t_{traj}}{a^2}$$

Plug-in the computed self-diffusion coefficient as Table S4, and the interstitial-site spacing between the octahedral sites in the ε-iron nitride a=2.7Å (assuming temperature independent), N=4464 diffusing nitrogen atoms. The expected total number of hops are also shown in Table S4, which rationalized the observed trend in Figure 8: uncertainty grows below 1300 K because the trajectory contains fewer hopping events, making the MSD slope statistically noisy. The hop count in Table S4 is reported as a qualitative sampling diagnostic (i.e., to indicate whether

Table S4: Self-diffusion coefficient computed using molecular dynamics driven by MLP[All] with their standard deviation and the thermodynamic factors (Φ).

	1500 K	1400 K	1300 K	1200 K	1100 K	1000 K	830 K
D_{self} [m^2/s]	9.49e-11	2.88e-11	8.72e-12	2.54e-12	8.11e-13	1.52e-13	5.53e-15
Standard deviation of D_{self} [m^2/s]	6.83e-12	9.99e-13	1.82e-13	1.65e-13	6.52e-14	2.19e-14	2.81e-15
Φ [-] computed by Eq.1	9.98	10.18	10.42	10.70	11.03	11.42	12.31
Simulation time [ns]	> 1.25	> 1.25	> 1.25	> 1.25	> 1.50	> 3.00	>5.00
MSD fitting window [ns]	>0.75	>1.15	>1.25	>1.15	>1.00	>1.50	>3.00
Expected number of hops within the fitting window	26152.4	12169.7	4003.3	1073.6	297.8	61.9	6.10

diffusion events are sparse), and is not used as the criterion for diffusion-coefficient validity; the latter is assessed from the long-time linearity of the ensemble-averaged MSD (see Section 3.4 on MSD fitting criteria).

Therefore at 830 K, the MLP[All]-driven MDs capture diffusion events, but the event rate is so low that the MSD estimation is event-limited. Longer simulation time or equivalent independent replicas is required to reach a stable long-time linear MSD state with enough independent hops for a reliable slope. Premature fitting of the MSD in this sparse event state inflates variance and can bias the apparent slope upward, resulting in an overestimation of diffusion coefficient with wide confidence intervals. In addition, the initial solute placement (randomly assigning nitrogen atoms on octahedral site) can generate non-equilibrium occupancy correlations. These occupancies relax faster than the mean hop time but can still imprint a transient, non-stationary MSD. Including early transients in the fitting will bias D_{self} upwards. Given the computational cost (6 ns, 14,832 atoms, $\approx$120 node-hours per run on 128 CPUs/Node, four replicas), direct low-temperature MSD calculation are impractical. Because the MSD estimation are robust between 1000–1500 K and the 830 K data are qualitatively consistent with that Arrhenius-like trend, we did not target an accurate, direct computation of diffusion coefficient within experimental temperature range. Notably, the expected number of hops within the fitting window at 830 K is only 6.10; therefore, even a 2–3 times longer trajectory would increase the hop count only to 12–18 times, which is not expected to substantially reduce the uncertainty in the fitted MSD slope. If an accurate low-temperature value is required, we recommend using our MLP[All] with the non-equilibrium MD approach described in Section 2.4.1, rather than further MSD fitting in the sparse-event regime.
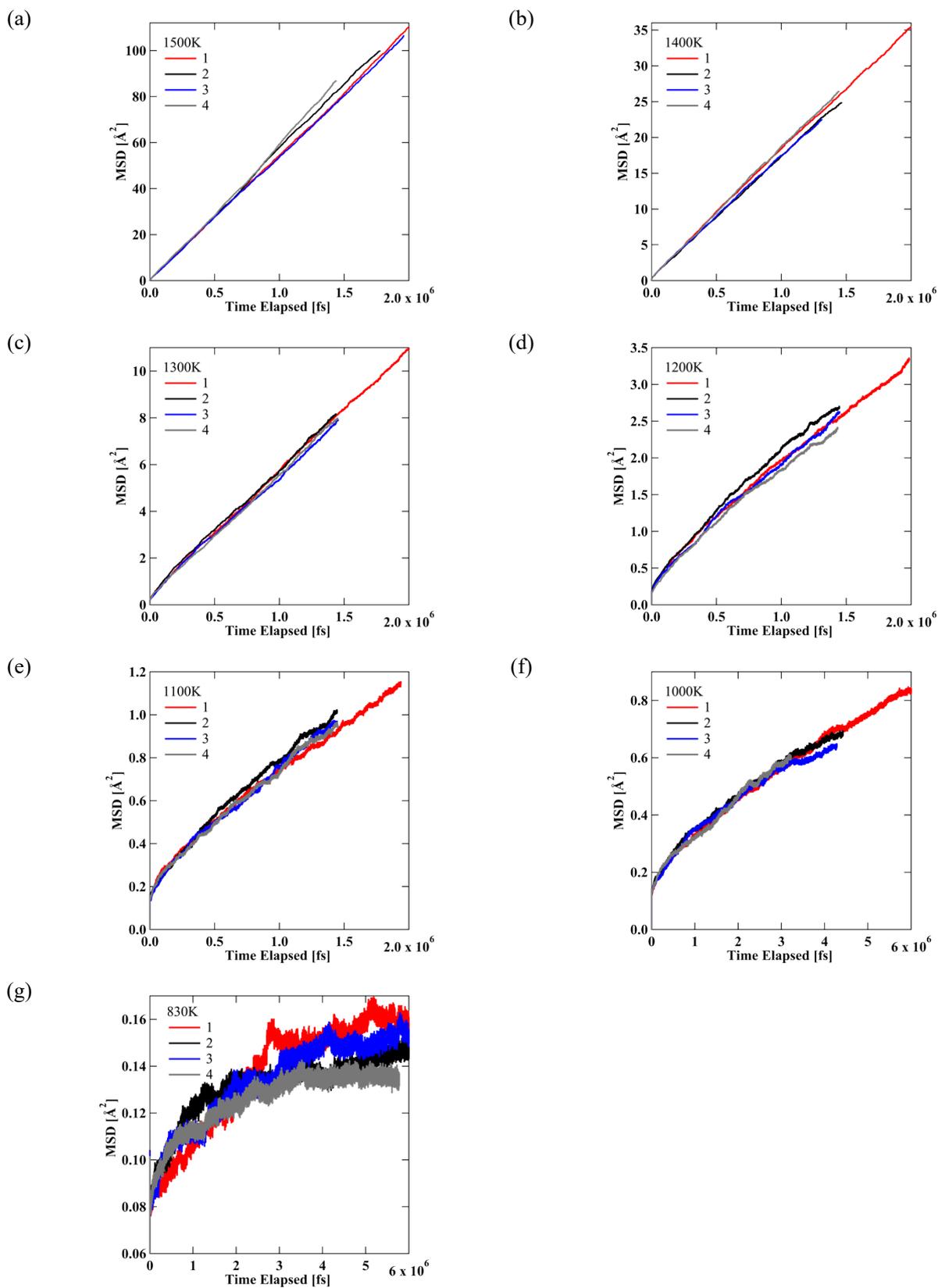
Figure S1: Raw data of the MLP[All] driven MD computed MSD at different temperatures to estimate the D_{self}
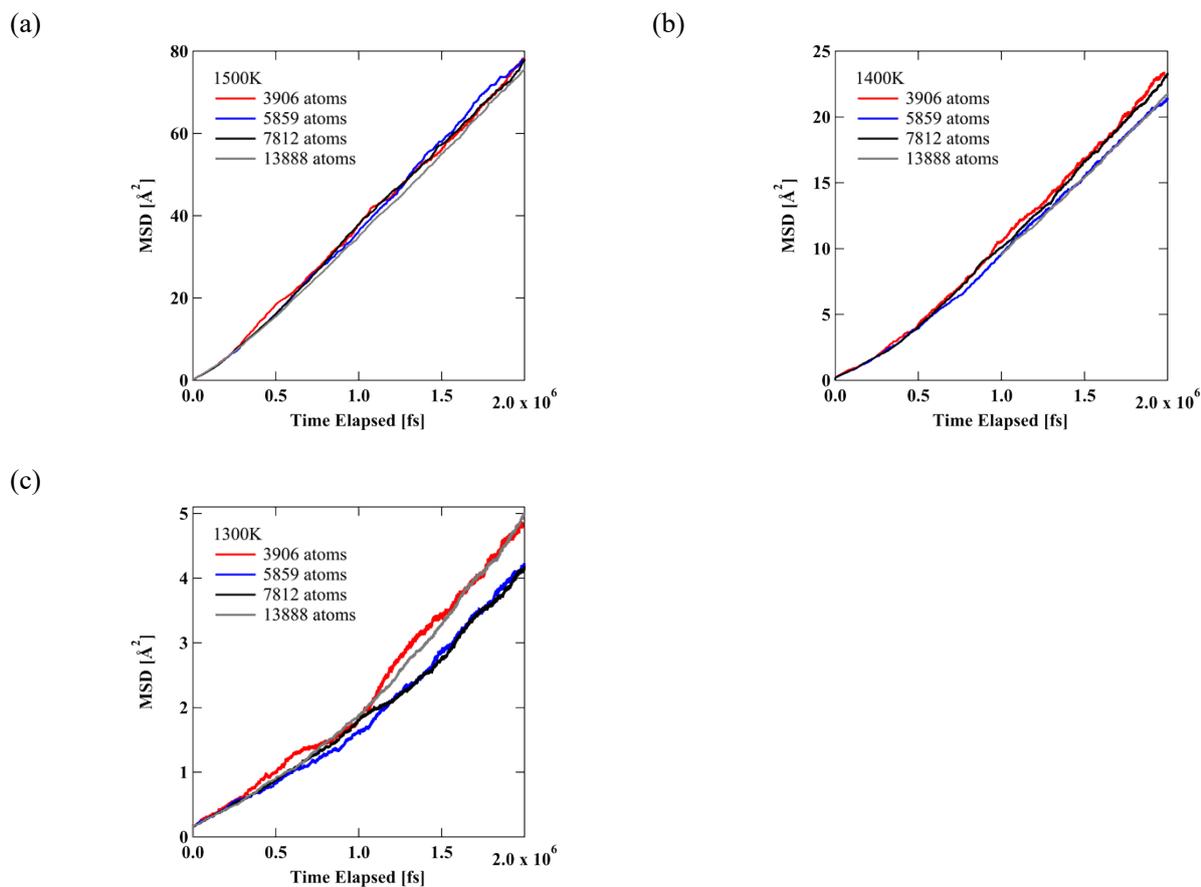
(a)

(b)

(c)

Figure S2: Raw data of the MLP[All] driven MD computed MSD at different temperatures with different system size. Showing no finite size effect (see the main text Section 3.4). Systems all consist of nitrogen contents approximately 9.51 wt%.
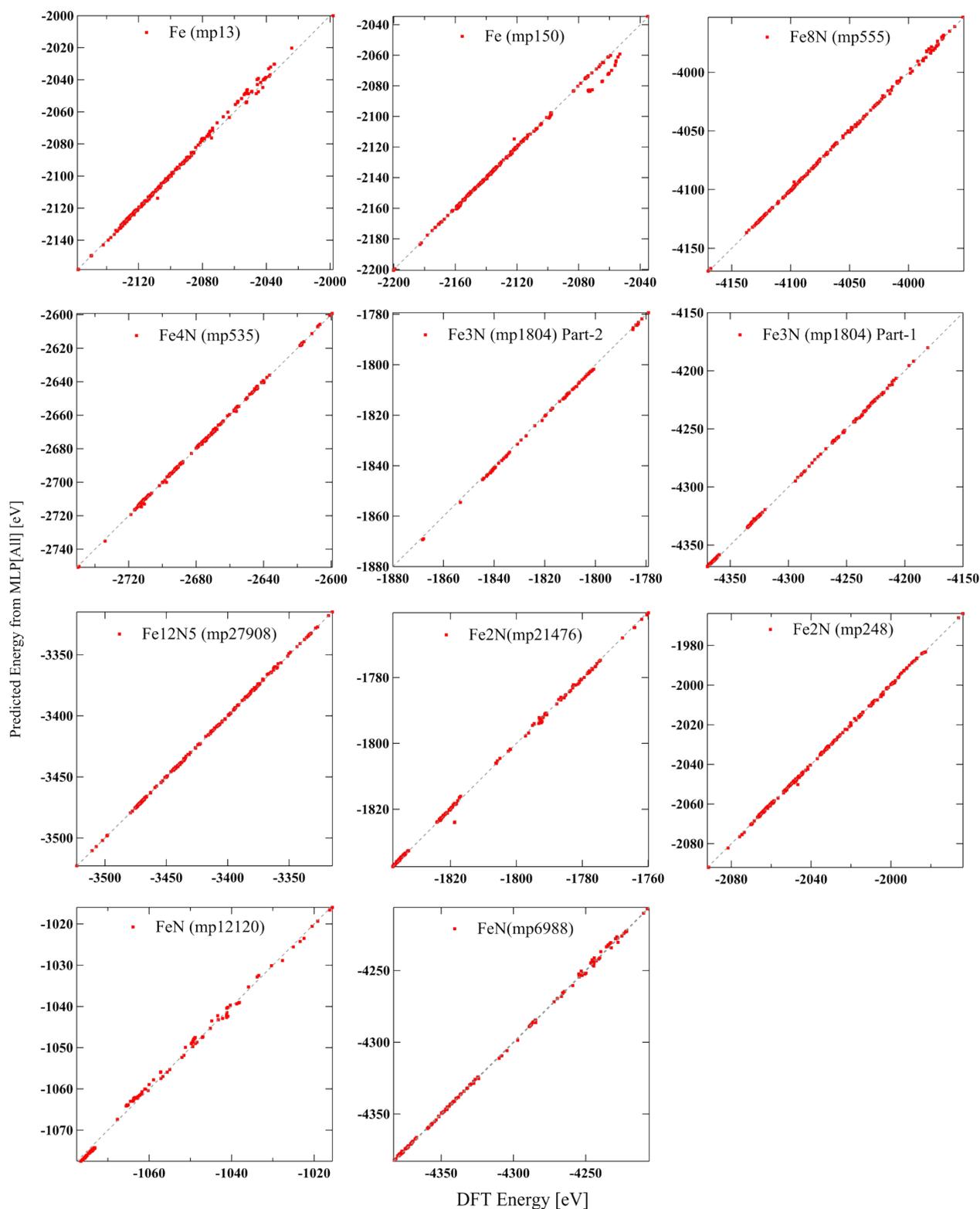
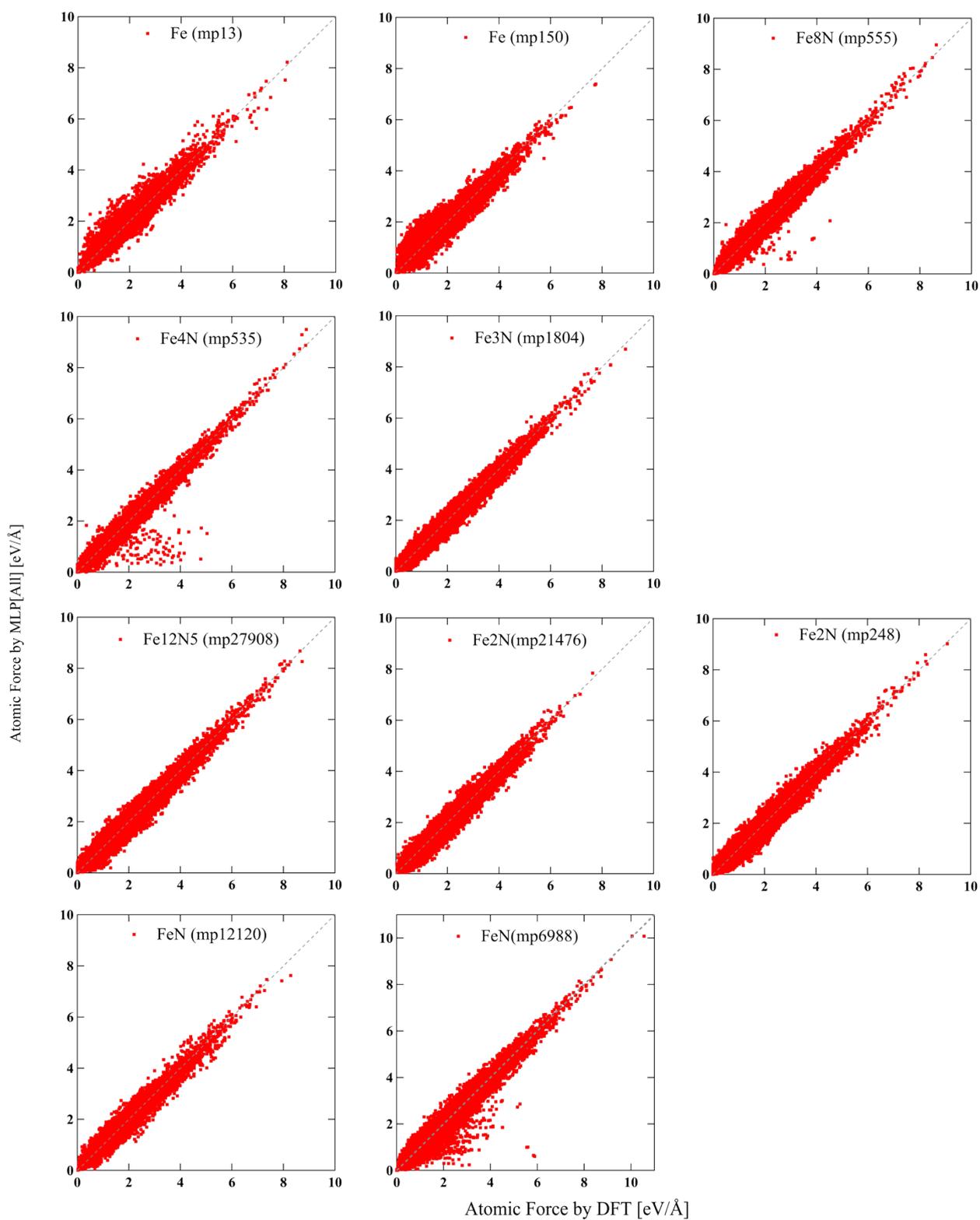Figure S3: The parity plots for energies; comparing the MLP[All] predictions against the DFT data.

Figure R4: The parity plots for atomic forces; comparing the MLP[All] predictions against the DFT data.

Table S5: Tested force mean absolute error (MAE) of N atoms in eV/Å.

System Name	Remove 2N	Remove 4N	Add 1N	Add 2N	Add 4N	500K	1000K	1500K	Cell Compression	Cell Expansion
Fe_mp13	-	-	0.0172	0.0295	-	-	-	-	-	-
Fe_mp150	-	-	0.0844	0.0420	0.0372	-	-	-	-	-
Fe8N_mp555	0.0168	-	-	0.0186	-	0.0070	0.0149	0.0401	0.0177	0.0266
Fe4N_mp535	0.0222	-	-	0.0178	-	0.0090	0.0145	0.0258	0.0374	0.0186
Fe3N_mp1804	0.0160	-	-	0.0202	-	0.0059	0.0127	0.0361	0.0328	0.0219
Fe12N5_mp27908	-	0.0245	-	-	0.0253	0.0096	0.0220	0.0515	0.0303	0.0413
Fe2N_mp21476	-	-	-	-	-	0.0155	0.0241	0.0601	0.0347	0.0590
Fe2N_mp248	0.0275	-	-	0.0271	-	0.0135	0.0265	0.0549	0.0398	0.0655
FeN_mp12120	-	-	-	-	-	0.0098	0.0469	0.0469	0.0480	0.0460
FeN_mp6988	-	-	-	-	-	0.0118	0.0173	0.0372	0.0569	0.0460

Table S6: Tested force root mean square error (RMSE) of N atoms in eV/Å.

System Name	Remove 2N	Remove 4N	Add 1N	Add 2N	Add 4N	500K	1000K	1500K	Cell Compression	Cell Expansion
Fe_mp13	-	-	0.1310	0.1717	-	-	-	-	-	-
Fe_mp150	-	-	0.2906	0.2048	0.1927	-	-	-	-	-
Fe8N_mp555	0.1294	-	-	0.1362	-	0.0839	0.1222	0.2002	0.1332	0.1632
Fe4N_mp535	0.1488	-	-	0.1336	-	0.0947	0.1204	0.1606	0.1933	0.1362
Fe3N_mp1804	0.1267	-	-	0.1421	-	0.0765	0.1129	0.1899	0.1811	0.1481
Fe12N5_mp27908	-	0.1566	-	-	0.1590	0.0982	0.1482	0.2268	0.1742	0.2031
Fe2N_mp21476	-	-	-	-	-	0.1244	0.1554	0.2452	0.1863	0.2429
Fe2N_mp248	0.1658	-	-	0.1646	-	0.1163	0.1627	0.2344	0.1995	0.2560
FeN_mp12120	-	-	-	-	-	0.0990	0.2165	0.2165	0.2192	0.2144
FeN_mp6988	-	-	-	-	-	0.1085	0.1315	0.1929	0.2386	0.2145

Table S7: Tested energy mean absolute error (MAE) in meV/atom.

System Name	Remove 2N	Remove 4N	Add 1N	Add 2N	Add 4N	500K	1000K	1500K	Cell Compression	Cell Expansion
Fe_mp13	-	-	2.196	2.463	-	1.539	2.136	8.366	6.113	17.746
Fe_mp150	-	-	1.791	1.250	2.123	0.948	1.099	2.075	1.610	36.985
Fe8N_mp555	1.194	-	-	1.292	-	0.470	0.868	2.269	0.908	3.979
Fe4N_mp535	0.993	-	-	1.214	-	0.423	0.642	1.144	1.179	2.847
Fe3N_mp1804	3.434	-	-	4.101	-	1.926	1.127	0.970	1.296	1.044
Fe12N5_mp27908	-	1.475	-	-	0.853	0.404	1.170	0.526	1.515	0.503
Fe2N_mp21476	-	-	-	-	-	0.992	1.050	1.541	1.826	7.057
Fe2N_mp248	1.429	-	-	1.628	-	1.089	0.847	1.840	1.412	2.014
FeN_mp12120	-	-	-	-	-	8.236	6.881	6.881	6.103	5.517
FeN_mp6988	-	-	-	-	-	0.822	0.666	2.316	2.946	0.225

Table S8: Tested energy root mean square error (RMSE) in meV/atom.

System Name	Remove 2N	Remove 4N	Add 1N	Add 2N	Add 4N	500K	1000K	1500K	Cell Compression	Cell Expansion
Fe_mp13	-	-	3.002	2.850	-	1.624	2.484	10.356	6.706	18.508
Fe_mp150	-	-	2.417	1.611	2.560	1.132	1.387	2.540	1.942	38.545
Fe8N_mp555	1.538	-	-	1.455	-	0.570	1.007	2.755	1.079	4.225
Fe4N_mp535	1.835	-	-	1.455	-	0.587	0.764	1.391	1.466	3.155
Fe3N_mp1804	3.632	-	-	4.337	-	1.958	1.217	1.244	1.517	1.433
Fe12N5_mp27908	-	2.051	-	-	1.038	0.507	1.358	0.694	1.844	0.631
Fe2N_mp21476	-	-	-	-	-	1.132	1.384	1.776	2.060	11.239
Fe2N_mp248	1.868	-	-	1.871	-	1.227	1.114	2.120	1.688	3.563
FeN_mp12120	-	-	-	-	-	8.275	7.533	7.533	6.868	6.335
FeN_mp6988	-	-	-	-	-	0.980	0.754	2.517	3.554	0.289