SUPPLEMENTARY INFORMATION

Robust Training of Machine Learning Interatomic Potentials with Dimensionality Reduction and Stratified Sampling

Ji Qi,[†] Tsz Wai Ko,[‡] Brandon C. Wood,[¶,§] Tuan Anh Pham,[*,¶,§] and Shyue Ping Ong[*,‡]

†*Materials Science and Engineering Program, University of California San Diego, 9500 Gilman Dr, Mail Code 0448, La Jolla, CA 92093-0448, United States*

‡*Department of NanoEngineering, University of California San Diego, 9500 Gilman Dr, Mail Code 0448, La Jolla, CA 92093-0448, United States*

¶*Quantum Simulations Group, Lawrence Livermore National Laboratory, Livermore, CA, 94550, USA*

§*Laboratory for Energy Applications for the Future (LEAF), Lawrence Livermore National Laboratory, Livermore, CA, 94550, USA*

E-mail: pham16@llnl.gov; ongsp@ucsd.edu

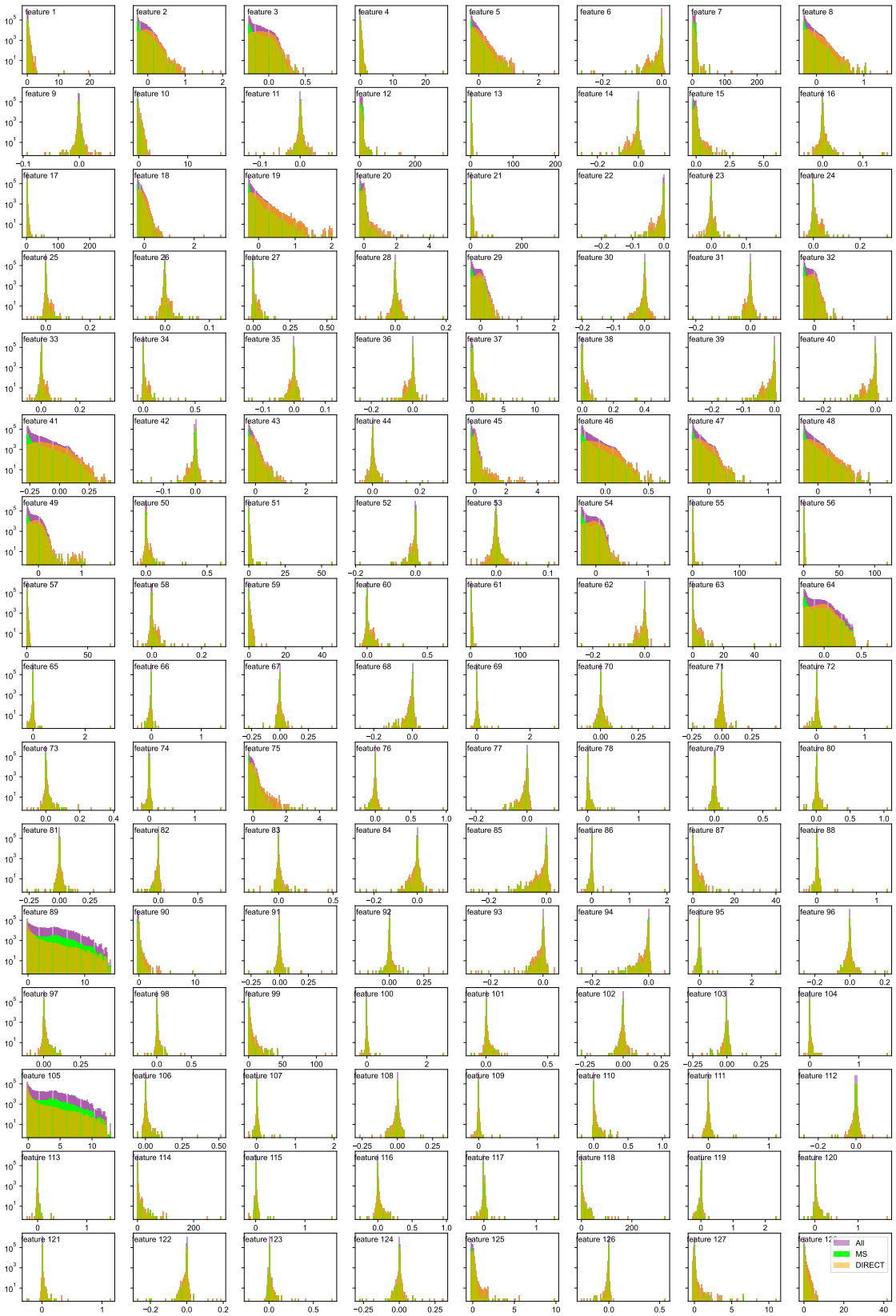

Figure 1: Distribution of all the 128 elements in M3GNet structural features of structures in DIRECT, MS and the entire MPF.2021.2.8 dataset.

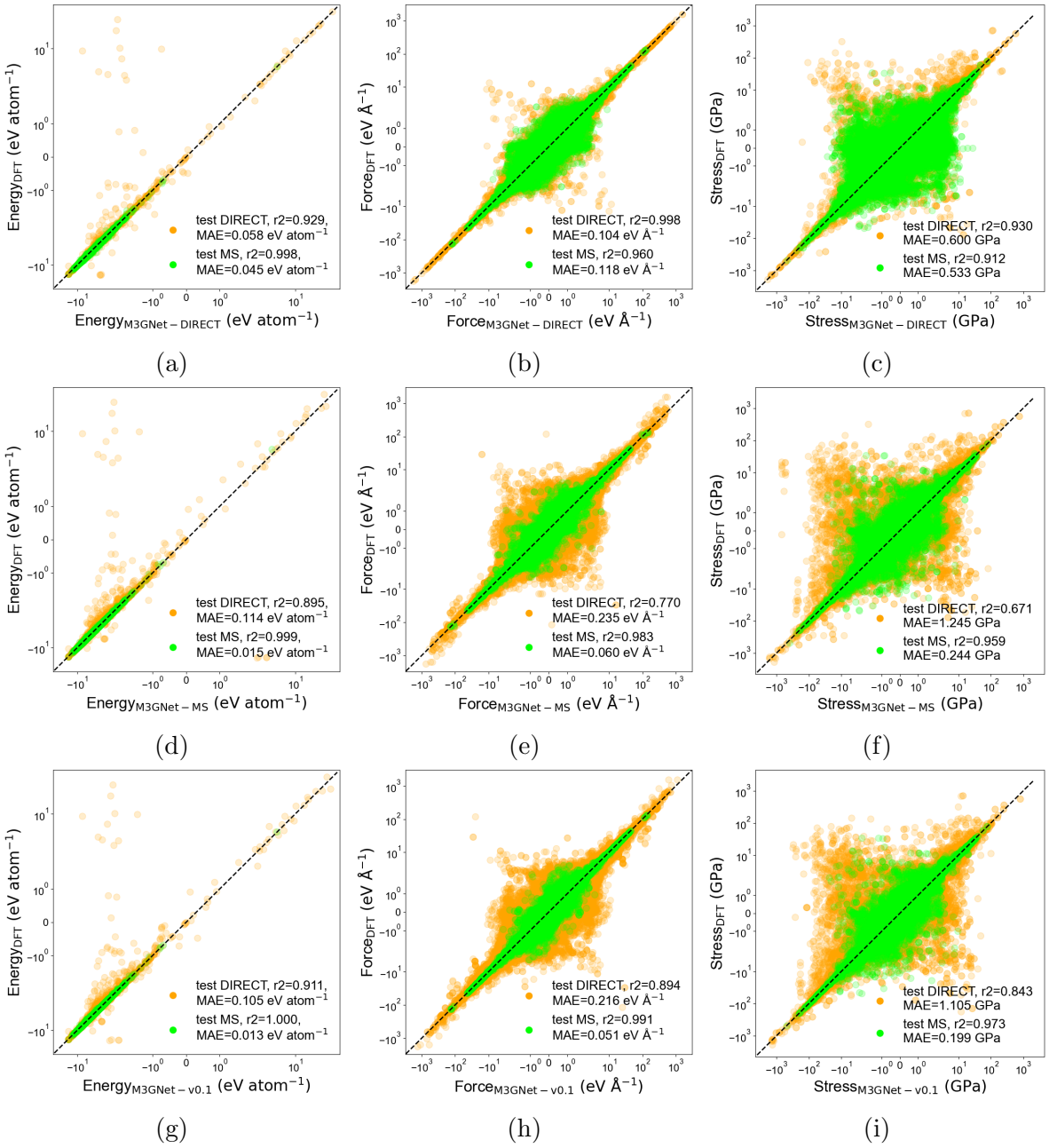

Figure 2: Performance of M3GNet UPs with the same model complexity as that of the pre-trained M3GNet-v0.1 by Chen and Ong[1]. (See details in Methods) Parity plots for (a) energies, (b) forces and (c) stresses for the M3GNet UP trained on the DIRECT set. The equivalent plots for the M3GNet UP trained on the MS set is shown in plots (d)-(f). The respective plots for the pre-trained M3GNet-v0.1 are also provided for comparison.

References

(1) Chen, C.; Ong, S. P. A Universal Graph Deep Learning Interatomic Potential for the Periodic Table. *Nature Computational Science* **2022**, *2*, 718–728.