Supplemental Information: On the liquid-liquid phase transition of dense hydrogen

Valentin V. Karasiev,[1,*] Joshua Hinz,[1] S. X. Hu,[1] and S.B. Trickey[2]

[1] *Laboratory for Laser Energetics, University of Rochester, 250 East River Road, Rochester, New York 14623 USA*

[2] *Quantum Theory Project, Dept. of Physics, University of Florida, Gainesville FL 32603 USA*

ARISING FROM Bingqing Cheng et al. *Nature* https://doi.org/10.1038/s41586-020-2677-y (2020)

For clarity yet brevity, in the following we refer to the Supplemental Information to Cheng et al. [1] as "Cheng-SI". Our discussion focuses on their machine-learnt potential (MLP).

Begin with several discrepancies between the MLP results and those from molecular dynamics driven by density functional theory forces (MD-DFT) with the PBE exchange-correlation (XC) functional [2]. (PBE XC was used in Ref. [1] to train the version of the MLP discussed in their paper.)

Fig. S9 of Cheng-SI shows H-H pair correlation functions (PCFs) from MD-DFT and from MD with the MLP, both in the NVT ensemble. There are clear, substantive, undiscussed differences especially for higher densities $r_s = 1.38$, 1.34, and 1.30 for distances inside roughly 2 Å.

Another discrepancy is displayed in Fig. S8 of the Cheng-SI. It shows that the molecular fraction at $T =$ 1000K in MD-DFT for a 128 atom system remains below ≈ 0.5 up to about $r_s = 1.43$, then jumps to above 0.8 at $r_s \approx 1.45$ and stays there almost unchanged to lower densities. The MLP is qualitatively different: it gives a molecular fraction that goes smoothly through the region $1.4 \leq r_s \leq 1.5$.

A possible, but apparently untested inconsistency involves the solid-like structures that Cheng et al. found in roughly 2 ps from MD-DFT with PBE XC for 128 atoms. They attributed formation of those structures to finite-size effects. As far as can be determined, the MLP was not tested to see if it would reproduce this size effect.

At larger sizes, information on page 8 of Cheng-SI suggests that MLP transferability from small training systems to large system applications may be problematic. In NVT simulations with 500 atoms using MLP at 900 K, they state that "...solid-like structures appeared after 50 ps." Our MD-/DFT calculations show no obvious evidence of such structure formation for that size system or larger in much shorter times (few ps).

The issue in all four of these is not whether a particular effect is physical. Rather, the issue is the fidelity of the MLP to the underlying electronic structure, in this case DFT/PBE. Does the MLP replicates what the PBE XC gives? If it does not, then the MLP has introduced something artifactual in the course of the machine learning that is not in the reference DFT electronic structure. From the evidence available in Ref. 1 and the Cheng-SI, that fidelity is questionable, at the least.

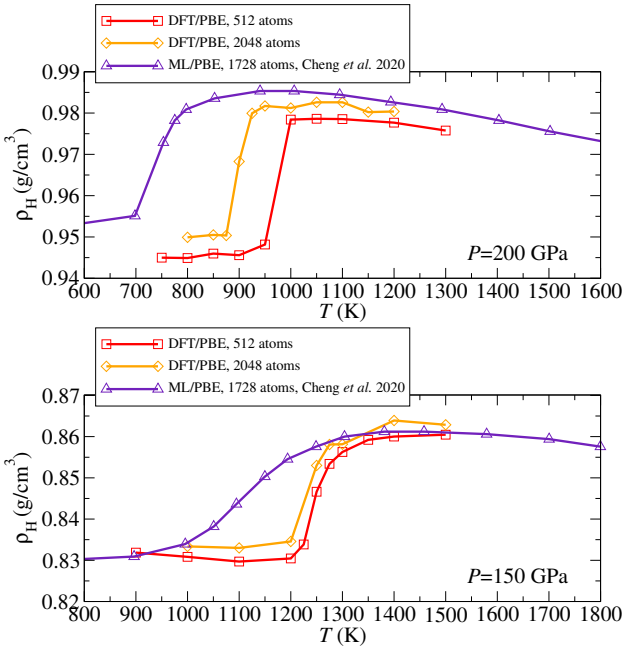

FIG. S1: High-P hydrogen density along two isobars for *ab initio* MD-DFT (DFT/PBE) on 512 and 2048 atom systems and MD-MLP (MLP/PBE) for 1728 atoms.

The results of our *ab initio* MD-DFT simulations for a system of 2048 atoms presented in the main text further supports the foregoing assessment, as can be seen by revisiting Fig. S1. It displays the bulk density as a function of T along two different isobars (same data as shown in Fig. 1 of the main text). The MD-MLP (PBE) curves (labeled "MLP/PBE") are smooth and featureless, while the MD-DFT (PBE) results (labeled "DFT/PBE") show clear density jumps. The interpolation done by the MLP is particularly striking, as is the uncontrolled nature of the MLP approximation. Above the transition temperature, for $P = 200$ GPa the 1728-atom MD-MLP density matches reasonably well with the 2048-atom MD-DFT results. But for temperatures below the transition, the 1728-atom MD-MLP density is far off the 2048-atom MD-DFT values. In contrast, for the $P = 150$ GPa isobar,

the 1728-atom MD-MLP density curve matches reasonably at both low and high T with both the 512-atom and 2048-atom MD-DFT curves. In between, the MD-MLP curve amounts to an incorrectly smoothed connection of the two correct regions

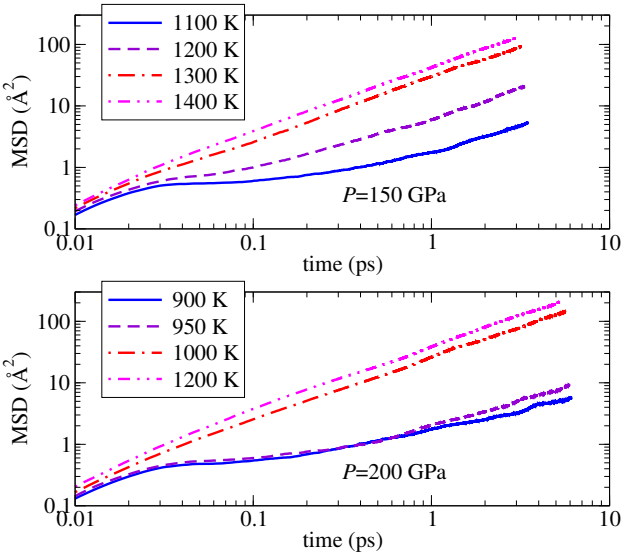

FIG. S2: Mean-squared displacement of 512-atom system as function of time for two $P = 150$GPa at $1100 \leq \mathrm{T} \leq 1400$K (upper panel) and for $P = 200$GPa at $900 \leq \mathrm{T} \leq 1200$K (lower panel).

Regarding the possibility that our simulations might in fact be for defective solids, we have calculated the mean-squared displacement (MSD)

$$D_m(t) = \frac{1}{N}\sum_{i=1}^{N} |\mathbf{R}_i(t) - \mathbf{R}_i(0)|^2 \qquad \text{(S1)}$$

as a function of time for 512-atom systems at $P = 150$ GPa and $1100 \leq \mathrm{T} \leq 1400$K and at 200 GPa and $900 \leq \mathrm{T} \leq 1200$K. Here $\mathbf{R}_i(t)$ is the position of nucleus i at time t and there are N nuclei. The MSD is computed over a fixed time interval (17,500 MD steps) with six initial times sampled along each MD-DFT trajectory. The results are displayed in Fig. S2. One sees clearly the continuous, near-linear increase of D_m with time that is an evident liquid signature. The distinction between molecular (lower MSD) and atomic liquids (higher MSD) also is evident.

We now turn our attention towards examining the most obvious suspect for the discrepancies between the MLP and our DFT simulations. Within any ML formulation it is essential that the data comprising the input vectors contains all of the features needed to describe the corresponding output vector fully. For predicting the energy of a given ionic configuration $\mathbb{R}$, the input vector

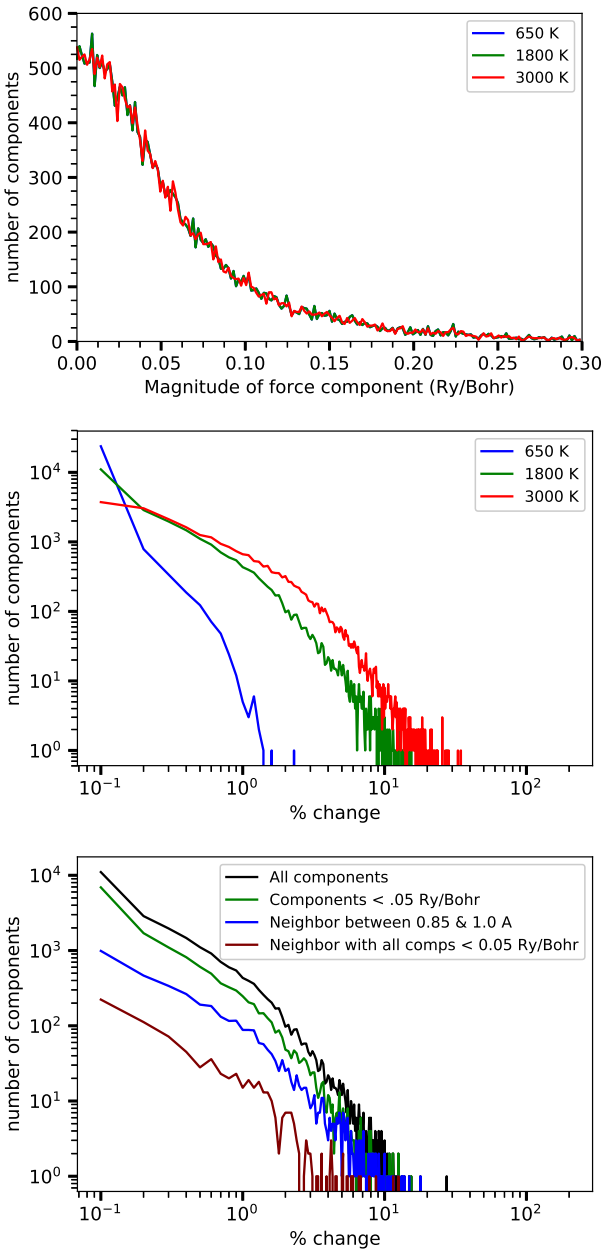

FIG. S3: Force component distributions resulting from the 35 snapshots. Top panel: distribution of the absolute value of the ionic force components, $|F_i|$. Note, the bin width is 0.0015 Ry/bohr and the curve represents the number of force components in each bin plotted with respect to the value of the left edge of each bin. Middle panel: distribution of the relative change in the ionic force components as defined in the text. Here the bin width is 0.1% and the curves are functions of the value of the right edge of each bin. Bottom panel: relative change distributions for $T_{\mathrm{electron}} = 1800$ K. The black curve is the distribution of all force components regardless of their magnitude. The green curve is the distribution for those force components that have a value less than 0.05 Ry/bohr; and the blue and red curves are the distributions of the atoms with a nearest neighbor between 0.85 and 1.0 Å and ions satisfying the neighbor conditions and that have all force components with a magnitude less than 0.05 Ry/bohr respectively.

prior to the construction of the descriptors should contain information on $\mathbb{R}$, T, V, and N. That corresponds to the MD-DFT energies and forces having an explicit dependence on the electron temperature (as well as N and V). In our work it is assumed to be the same as the ion temperature: $T_{\text{electron}} = T_{\text{ion}} \equiv T$. In the MLP, however, that input vector (again prior to the construction of the descriptors) is projected onto a vector containing only $\mathbb{R}$. The MLP training configurations thus were sampled in both the molecular liquid and atomic liquid regimes without explicit T-dependence (either ionic or electronic) or consideration of chemical state. Instead, the iterative procedure (Cheng-SI, page 9) for 38,716 training configurations of small systems (8-100 atoms) picks a set of ion positions $\mathbb{R}$ from a newly fitted approximate potential (from the previous step), evaluates the single-point DFT ground-state energy ($T_{\text{electron}} = 0$ K) at that configuration, refines the search potential, and repeats. The result is an unspecified thermal distribution of reference $\mathbb{R}$ values with no chemical specificity in what, in fact, is a two-component system.

An MLP trained on molecular fluid configurations reachable at various ionic temperatures and on atomic fluid configurations at various ionic temperatures (with the reference DFT data properly calculated for each temperature or approximated as the ground-state DFT data) may have inappropriate implicit temperature dependence. We have investigated the energy, pressure, and force component magnitudes as a function of T_{electron} that can arise from such thermally indifferent configuration sampling. For example, we took 35 independent ionic configurations from MD along the $\rho_{\text{H}} = 1.0$ g/cm^3 isochore at T_{ion} ranging from 800 to 2500 K (with SCAN-L XC [3, 4]). For each of those snapshots, we did single-point DFT/SCAN-L calculations for $600 \leq T_{\text{electron}} \leq 3000$ K at 50 K intervals (details of those simulations are consistent with our previous study [7]). Force components from all 35 configurations, each consisting of 256 atoms, were calculated at each electronic temperature and used to form force component distributions. The distribution of the magnitude of the force components is shown in Fig. S3 top panel for three different electronic temperatures. It is clear that this distribution as a whole is relatively insensitive to T_{electron}.

For further insight, we examine the relative change of the ionic force components, Fig. S3 middle panel. We define this relative change as $|F_i(T) - F_i(600\text{K})|/|F_i(600\text{K})|_{\text{average}}$, $i = \{x, y, z\}$. Here we use the average of the force component magnitude distribution in the denominator in an attempt to avoid rather large but meaningless relative changes in the smallest force components from skewing the distribution. Furthermore, we investigated the conditional distributions of the shift in component magnitudes for only those ions that have a neighbor within 0.85-1.0 Å (hence, may be involved in molecular dissociation or formation) and for those ions which satisfy the neighbor condition and that have all three force component magnitudes below 0.05 Ry/bohr. Of the 8960 ions considered in our 35 snapshots, roughly 14.9% satisfy the neighbor condition and 8.3% satisfy both conditions. In the context of the relative changes, at $T_{\text{electron}} = 1800$ K, Fig. S3 bottom panel, only 2.7% of the ions satisfying the neighbor condition see a relative change larger than 5% and only 1.7% of the ions satisfying both conditions see shifts larger than 5%.

It is implausible that this rather small population would have a major qualitative impact on the outcome of MD runs. To confirm that assessment, MD simulations along the 200 GPa isobar with PBE XC were performed with T_{electron} set to 800 K for a system of 512 atoms. The resulting density as a function of the ionic temperature was nearly identical to the curve for which $T_{\text{electron}} = T_{\text{ion}}$, confirming the inconsequential effect such a deviation of the electronic temperature has on the LLPT. On the basis of tests of this sort, it appears that the lack of explicit T-dependence in the MLP is not material in causing its qualitative deviation from the underlying DFT energetics. High accuracy delineation of the details of the phase transition should, however, include T_{electron} effects explicitly [7].

Acknowledgements: This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

V.V.K., J.H., and S.X.H. were supported by the Department of Energy National Nuclear Security Administration Award Number DE-NA0003856 and US National Science Foundation PHY Grant No. 1802964. Partial funding for S.X.H. was provided by NSF Physics Frontier Center Award PHY-2020249. S.B.T. was supported by Department of Energy Grant DE-SC0002139. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science

User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Part of the computations were performed on the Laboratory for Laser Energetics HPC systems.

* Electronic address: vkarasev@lle.rochester.edu

[1] B. Cheng, G. Mazzola, C.J. Pickard, and M. Ceriotti, Nature **585**, 217 (2020).
[2] J.P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); erratum *ibid.* **78**, 1396 (1997).
[3] D. Mejía-Rodríguez and S.B. Trickey, Phys. Rev. A **96**, 052512 (2017).
[4] D. Mejía-Rodríguez and S.B. Trickey, Phys. Rev. B **98**, 115161 (2018).
[5] H.Y. Geng, Q. Wu, M. Marqués, and G.J. Ackland, Phys. Rev. B **100**, 134109 (2019).
[6] Gérald Faussurier and Christophe Blancard, Phys. Rev. E **99**, 053201 (2019).
[7] Joshua Hinz, Valentin V. Karasiev, S.X. Hu, Mohamed Zaghoo, Daniel Mejía-Rodríguez, S.B. Trickey, and L. Calderín, Phys. Rev. Research **2**, 032065(R) (2020).