Deep Variational Free Energy Approach to Dense Hydrogen

Hao Xie[1,2] Zi-Hang Li[1,2] Han Wang[3,*] Linfeng Zhang,[4,5,†] and Lei Wang[1,6,‡]

[1]*Beijing National Laboratory for Condensed Matter Physics and Institute of Physics,*
Chinese Academy of Sciences, Beijing 100190, China
[2]*School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China*
[3]*Laboratory of Computational Physics, Institute of Applied Physics and Computational Mathematics,*
Fenghao East Road 2, Beijing 100094, China
[4]*DP Technology, Beijing 100080, China*
[5]*AI for Science Institute, Beijing 100080, China*
[6]*Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China*

We developed a deep generative model-based variational free energy approach to the equations of state of dense hydrogen. We employ a normalizing flow network to model the proton Boltzmann distribution and a fermionic neural network to model the electron wave function at given proton positions. By jointly optimizing the two neural networks we reached a comparable variational free energy to the previous coupled electron-ion Monte Carlo calculation. The predicted equation of state of dense hydrogen under planetary conditions is denser than the findings of *ab initio* molecular dynamics calculation and empirical chemical model. Moreover, direct access to the entropy and free energy of dense hydrogen opens new opportunities in planetary modeling and high-pressure physics research.

Hydrogen is the most abundant element in the visible universe. It is also the first and simplest element in the periodic table, consisting of only a proton and an electron. Despite its simplicity, high-pressure dense hydrogen exhibits rich physical phenomena [1] such as metallization [2] and high-temperature superconductivity [3]. A thorough understanding of these phenomena is of fundamental importance to a broad range of disciplines, including planetary physics [4] and nuclear fusion [5]. For these reasons, accurate prediction of the equations of state and phase diagram of dense hydrogen has been a touchstone for computational methods.

Dense hydrogen is a quantum many-body system consisting of coupled protons and electrons. The Fermi temperatures of proton and electron are well separated due to their large mass difference. Therefore, for a wide temperature range that falls in between these two scales, one can safely treat protons as classical particles and assume the electrons stay in their instantaneous ground state with fixed proton positions. Solving the electronic Hamiltonian provides an effective potential for the protons. However, since the energy scale of electrons is much higher than that of protons, even a tiny error in the electronic calculation could significantly affect the predicted proton configurations. Standard *ab initio* molecular dynamics (MD) [6] solves the electronic structure problem using density functional theory calculations, whose reliability depends on the specific choice of density functionals [7]. Using machine-learned potential energy surfaces can push such MD simulations to much larger system sizes and longer times [8,9]. However, such an approach at most reflects the accuracy of the underlying electronic structure model that generates the training data, and the reliability issue still persists [10–14].

A more reliable method to solve the many-electron ground state problem is quantum Monte Carlo [15]. In the context of dense hydrogen, one can sample the proton configurations according to stochastic estimates of the energy or force acting on the protons, as were previously done in the coupled electron-ion Monte Carlo (CEIMC) [16] method and Langevin MD [17], respectively. However, these nested Monte Carlo approaches have two unsatisfactory drawbacks. First, the statistical noises in the estimated energy or force hamper an unbiased sampling of the protons, similar to the case of Bayesian inference with noisy log-likelihood functions [18]. There have been three remedies in the literature: (a) the noisy Monte Carlo approach [19] assumes the noises are sufficiently small and treats the acceptance rate using the von Neumann-Ulam method; (b) the penalty method [20] assumes that the noisy energy estimates follow the Gaussian distribution and reduces the acceptance rate with an empirically estimated variance; (c) the stochastic gradient Langevin dynamics [17,21,22] relies on sufficiently small integration steps with noisy forces to sample from the correct Boltzmann distribution. In all cases, statistical uncertainties in the energy functions deteriorate the sampling efficiency and may even introduce bias to the results.

126501-1

These downsides have partially reduced the reliability advantage of employing more sophisticated quantum Monte Carlo solver for dense hydrogen. For example, suppose one has obtained different results with different flavors of quantum Monte Carlo methods or initial proton configurations, it is rather difficult to tell which results to trust. Second, the nested nature of the approach makes the computation very demanding: one has to make the inner electronic calculation fully converge to ensure correct sampling of the Born-Oppenheimer potential energy surface of protons. Such a stringent requirement makes it rather tedious to ensure convergence of the method [23,24] or limits one to manually crafted variational wave functions with a few or even no variational parameters [25,26].

In light of these difficulties faced by the nested Monte Carlo approaches [16,17], we introduce a deep generative model-based variational free energy approach for the dense hydrogen problem. We will minimize the variational free energy with respect to a trial density matrix [27]

$$F = k_B T \mathrm{Tr}(\rho \ln \rho) + \mathrm{Tr}(\rho H), \quad (1)$$

with the two terms being entropy and energy, respectively; k_B is the Boltzmann constant and T is the temperature. Although free energy minimization is a fundamental principle in quantum statistical mechanics, its practical application is inhibited by the intractable computational cost of entropy term [28]. Recent advances in deep generative models [30] have removed this roadblock. Variational free energy calculations based on deep generative models have been applied to a wide range of problems, including the Ising models [31–34], lattice field theories [35–37], atomic solids [38,39], quantum dots [40], and uniform electron gases [41]. Because of the intrinsic difficulties of alternative quantum Monte Carlo approaches, the deep generative model-based variational free energy methods have the potential to become an indispensable tool for many-fermion problems at finite temperature, such as Refs. [40,41] and the hydrogen problem considered here.

We represent the density matrix ρ using two neural networks as shown in Fig. 1, one for the proton Boltzmann distribution and one for the electron wave function. Variational free energy calculation of Eq. (1) then amounts to solving a stochastic optimization problem [42]. In such a formulation, the statistical noises in the estimated energy will not be as catastrophic as in the Monte Carlo sampling [16,17]. In this respect, the present approach trades the shortcomings of nested Monte Carlo approaches [16,17] with a variational bias. However, by making use of deep neural network ansatzes one can largely overcome this issue by progressively lowering the variational free energy. With further systematic improvements of variational ansatz and optimization scheme, we anticipate one will reach a
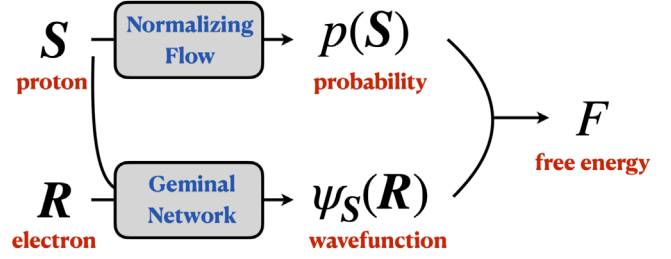

FIG. 1. A sketch of computational graph for the dense hydrogen problem. The model consists of a normalizing flow [Eq. (3)] for the proton Boltzmann distribution and a geminal neural network [Eq. (4)] for the electron wave function with fixed proton positions. We jointly optimize the two neural networks to minimize the variational free energy [Eq. (1)].

reliable description for the whole phase diagram of dense hydrogen with the variational free energy approach [43].

Consider N protons and N electrons in a periodic cubic box of volume L^3. The system is unpolarized so there are $N/2$ spin-up and spin-down electrons, respectively. The density of the system is specified by the dimensionless parameter $r_s = (3/4\pi N)^{1/3} L/a_0$, where a_0 is the Bohr radius. In the atomic units, the Hamiltonian of the hydrogen system reads

$$H = \sum_i \frac{-\nabla_i^2}{2} + \sum_{i<j} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{I<J} \frac{1}{|\boldsymbol{s}_I - \boldsymbol{s}_J|} - \sum_{i,I} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{s}_I|}, \quad (2)$$

where $\boldsymbol{S} = \{\boldsymbol{s}_I\}$ and $\boldsymbol{R} = \{\boldsymbol{r}_i\}$ denote the proton and electron coordinates, respectively. Note here we have omitted the proton kinetic energy term, whose effect is considered separately in the Supplemental Material [43].

The variational density matrix of dense hydrogen under consideration is diagonal with respect to the proton degrees of freedom, and can be written as $\rho = \int d\boldsymbol{S} p(\boldsymbol{S}) |\boldsymbol{S}, \psi_S\rangle \langle \boldsymbol{S}, \psi_S|$. $p(\boldsymbol{S})$ is a normalized probability density for the proton coordinates, and $|\boldsymbol{S}, \psi_S\rangle \equiv |\boldsymbol{S}\rangle \otimes |\psi_S\rangle$ is a basis of the whole system's Hilbert space, where $|\psi_S\rangle$ is the electronic ground state at fixed proton configuration $\boldsymbol{S}$. The variational free energy $F = \int d\boldsymbol{S} p(\boldsymbol{S}) [k_B T \ln p(\boldsymbol{S}) + \langle \psi_S | H | \psi_S \rangle]$ then consists of the entropy of the proton Boltzmann distribution and the electronic expected energy weighted over the proton configurations.

We parametrize the proton Boltzmann distribution $p(\boldsymbol{S})$ using a normalizing flow network. Normalizing flow is a class of deep generative model that represents high-dimensional probability density using change-of-variables transformations [44]. Specifically, assuming $\boldsymbol{\zeta}$ is a set of independent and uniformly distributed "collective" coordinates, we use a neural network to implement a learnable bijective mapping between $\boldsymbol{\zeta}$ and the original proton

coordinates S. The resulting probability density for protons then reads

$$p(S) = \frac{1}{L^{3N}} \left| \det\left(\frac{\partial \zeta}{\partial S}\right) \right|. \tag{3}$$

Note this expression is normalized, which facilitates a straightforward and easy computation of the entropy term in the variational free energy objective [Eq. (1)]. We construct the bijective transformation between ζ and S as a residual network using the fermionic neural network layers [45]. Since each layer is permutation equivariant, Eq. (3) is invariant to the permutation of proton coordinates [46]. Moreover, the construction also ensures translational invariance and periodicity of the probability density [43]. We compute the transformation Jacobian in Eq. (3) using forward-mode automatic differentiation [47].

Next, we design a neural network for the electronic ground-state trial wave function $\psi_S(R) \propto \langle R|\psi_S\rangle$ at fixed proton positions. We concatenate the proton and electron coordinates together and feed them into a fermionic layer [45] that accounts for the periodic boundary conditions and translational invariance [41,48–51]. The layer outputs $f^S \in \mathbb{R}^{N \times M}$ and $f^\uparrow, f^\downarrow \in \mathbb{R}^{N/2 \times M}$, which are features that transform equivariantly under permutation of protons or electrons of the same spin. Using these equivariant features, we construct an unnormalized Jastrow-geminal-type wave function [43]

$$\psi_S(R) = e^J \det(G \circ D), \tag{4}$$

where $J = \sum_{i,\mu} a_\mu f^S_{i\mu}$ serves as a Jastrow factor, and $\circ$ denotes an element-wise product between two $N/2 \times N/2$ geminal matrices. $G_{ij} = \sum_{\mu\nu} \chi^\uparrow_{i\mu} W_{\mu\nu} \chi^\downarrow_{j\nu}$ depends on a learnable real-valued matrix W, and $\chi^\uparrow, \chi^\downarrow \in \mathbb{R}^{N/2 \times M}$ are simply another set of features that are connected with $f^\uparrow, f^\downarrow$ via a linear map. On the other hand, $D_{ij} = \sum_k \lambda_k e^{i k \cdot (z^\uparrow_i - z^\downarrow_j)}$ is formed by usual plane-wave orbitals with backflow coordinates $z^\uparrow$ and $z^\downarrow$. λ_k are learnable positive parameters representing the occupation number for the momenta $k = 2\pi n/L$ ($n \in \mathbb{Z}^3$). We use a large number of k points in the summation so that the wave function can easily capture oscillatory features.

Note the feature size M relevant to the matrix G plays essentially the same role as the number of plane-wave orbitals in such a geminal-type wave function ansatz. When $M > N/2$, a single geminal determinant would correspond to a summation of combinatorially large $\binom{M}{N/2}$ number of determinants according to the Cauchy-Binet formula. From this point of view, the fermionic neural network [45] construction of the features $\chi^\uparrow, \chi^\downarrow$ further boosts the expressibility of the ansatz compared to the well-known geminal states with Jastrow factors [52]. One can verify that besides being antisymmetric under permutation of

electrons of the same spin to account for their fermionic nature, the wave function [Eq. (4)] is also permutation *invariant* with respect to the proton coordinates. Note the architecture design here is more lightweight than Refs. [53,54] regarding the goal of respecting the permutation symmetry of nuclei.

By parametrizing the hydrogen density matrix using two neural networks, the variational free energy calculation of Eq. (1) reduces to the following stochastic optimization problem:

$$\min_{\phi,\theta} \mathbb{E}_{S \sim p(S)} \left[k_B T \ln p(S) + \mathbb{E}_{R \sim |\psi_S(R)|^2} \left[\frac{H\psi_S(R)}{\psi_S(R)} \right] \right], \tag{5}$$

where ϕ, θ are variational parameters of the proton Boltzmann distribution [Eq. (3)] and the electronic ground-state wave function [Eq. (4)], respectively. We employ the Markov chain Monte Carlo algorithm to draw proton and electron coordinate samples from the two models in an ancestral manner. As for optimization, we have used a generalized stochastic reconfiguration method [52] for density matrices, similar to the one employed in [41]. Note, however, that some subtle yet important modifications have to be made, which originate from the fact that the electron wave function ansatz [Eq. (4)] adopted in this Letter is unnormalized [43]. Since the proton Boltzmann distribution and electron wave function are optimized in a joint manner, one does not need to wait for an expensive inner optimization loop to fully converge before moving the protons. At a conceptual level, jointly moving the nuclear and electronic degrees of freedom is akin to the Car-Parrinello molecular dynamics [6], yet in a principled variational optimization framework.

We employ twist-averaged boundary conditions [55] to reduce the finite size effects, in particular those originating from the single-particle momentum shell structure. This amounts to adding some twist angle $q \in [-\pi/L, \pi/L]^3$ to the momenta k in the plane-wave geminal matrix D of Eq. (4). To further improve the accuracy of electronic calculations, we have also introduced some other explicit twist dependence in the wave function ansatz [43].

As a first application of the deep variational free energy approach, we focus on dense hydrogen at planetary conditions, where it is an ionized liquid of protons and electrons. Equations of state of dense hydrogen under such a condition can be used to construct models for giant planets' interiors [4]. For these applications, it is also crucial to compute the entropy so that one can follow the adiabatic curve from the planet's surface to its interior [56]. However, standard Monte Carlo methods do not have direct access to the entropy. Reference [57] employed the coupling-constant integration method based on CEIMC calculations to obtain the free energy and entropy of dense hydrogen at $r_s = 1.25$ and $T = 6000$ K. This result was
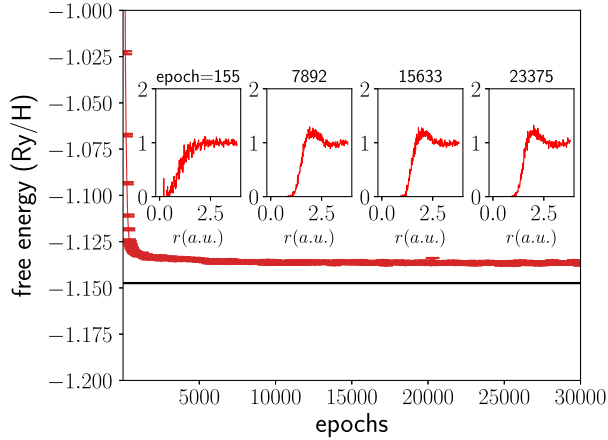
FIG. 2. Variational free energy per atom versus optimization epochs of $N = 54$ hydrogen atoms at $r_s = 1.25$ and $T = 6000$ K. The horizontal line shows the free energy obtained by coupling-constant integration of CEIMC energies [57]. The inset shows the proton-proton radial distribution functions at several different optimization epochs.
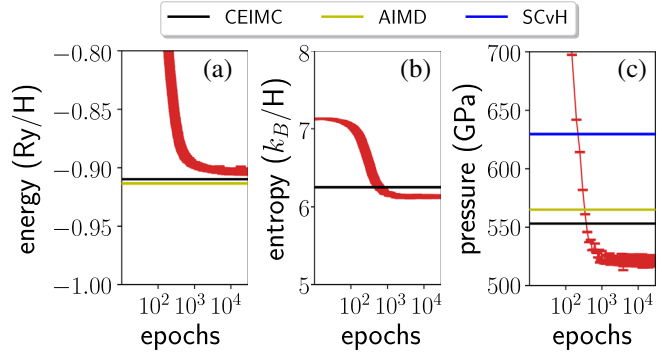

FIG. 3. (a) Internal energy per atom, (b) entropy per atom, and (c) pressure versus the optimization epochs. The system parameters are the same as Fig. 2. The horizontal black lines show the CEIMC results of [57]. The yellow and blue lines show the *ab initio* MD and SCvH chemical model [59] predictions, respectively, also taken from [57].

then used as the anchor point to obtain the full equations of state of dense atomic hydrogen.

We carry out deep variational free energy calculation at the same point ($r_s = 1.25$, $T = 6000$ K) so it is possible to benchmark with Ref. [57]. We choose the system size $N = 54$ and perform the twist average over a $4 \times 4 \times 4$ Monkhorst-Pack mesh as in [57], expecting a similar amount of finite size errors as detailed in the Supplemental Material [43]. We initialize parameters in the neural networks [Eqs. (3) and (4)] such that the protons are uniformly distributed and the electrons start from an itinerant plane-wave state. Figure 2 shows the variational free energy converges to a slightly higher value than that reported in [57]. We note, however, that the reference data may be nonvariational due to the nature of coupling-constant integration. The insets of Fig. 2 show the proton-proton radial distribution functions develop a structure as the optimization proceeds. Similar to Ref. [57], we have found that the proton-proton correlation functions show very little size effect at the considered density and temperature [43].

Figure 3 shows the energy, entropy (per atom), and pressure of the very same system as Fig. 2. Notice the entropy decreases upon training, since the protons will develop a more informative distribution than the initially uniform one. The converged values for the energy (entropy) are slightly higher (lower) than the estimates of Ref. [57], respectively, even after accounting for the finite size corrections [43]. The pressure is computed using the virial theorem as $(2K + V)/(3L^3)$ [58], where K and V are the total kinetic and potential energy of the system, respectively.

After accounting for the finite size correction [43], our estimated pressure 529(4) GPa is slightly smaller than the

CEIMC prediction 553(1) GPa [57]. Nevertheless, both values are smaller than *ab initio* MD calculation with local density approximation, which is even further away from the Saumon–Chabrier–van Horn (SCvH) equations of state based on a chemical model as a mixture of hydrogen molecules, atoms, protons, and electrons [59]. Predicting an even denser equation of state than the SCvH chemical model is in line with previous quantum Monte Carlo calculations [57,60]. Combined with the ability to directly access entropy, the deep variational free energy approach developed in this Letter can be a valuable tool for planetary modeling.

The fact that we have reached higher internal energy ($K + V$) but lower pressure ($\propto 2K + V$) indicates our calculation obtains lower kinetic energy and higher potential energy than the CEIMC results of Ref. [57]. It is known that the accuracy of internal energy is often greater than pressure, as the latter requires high accuracy both in the kinetic and potential energies whose errors tend to cancel out in the internal energy [61]. Likewise, we are thus more confident about our calculated internal energy than the pressure because the virial estimator for the pressure omits Pulay-like corrections due to incomplete optimizations [62]. We have released our codes and trained models [63]. By inspecting and further reducing the variational errors one can verify our findings reported here.

Unlike many previous studies on dense hydrogen [16,17,57,60,65], our calculation starts from a rather uninformative point with minimal physical constraints. In this regard, it is rather satisfying that the present calculation has yielded compatible equations of state for such an intensively studied system. In the future, one can also put prior knowledge such as the empirical or machine-learned potential [66] into the flow model [Eq. (3)]. One can use them either to pretrain the flow model or replace the uniform base distribution. In the latter case, the flow transformation is intended only to learn a small correction

and our variational objective function will be the free energy difference with the base model.

The results of Figs. 2 and 3 run on 32 Nvidia A100 GPUs for about 200 hours. It will just be a matter of parameter scan to produce the full equations of state of dense hydrogen in the atomic phase with direct access to entropy and free energy data. For higher temperatures, one can take into account the thermal effect of electrons using the neural canonical transformation approach [40,41]. Note that unlike the path integral Monte Carlo methods [67–70] such an approach will not suffer from the fermion sign problem. At lower temperatures, intriguing physics such as quantum liquid and superconducting order may come into play. To take into account nuclear quantum effects in this case, we envision one can either generalize the proton probabilistic model to the path integral representation [71] or adopt the neural canonical transformation approach [40,41] for protons. Either way, we anticipate the computational cost brought by the zero-point motion of protons does not increase significantly, in contrast to conventional approaches like path integral MD.

In this Letter, we have directly computed the Jacobian of the flow transformation in Eq. (3). To scale up to even larger system sizes, one may employ more efficient permutation equivariant flow models [38,72–74]. Moreover, one may also consider more scalable optimization schemes [75] for a larger number of variational parameters. In the meantime, it may be useful to explore optimization schemes beyond the score function gradient estimator [76] for the flow model. For example, one can use the pathwise gradient estimator to exploit information about the nuclear force [77,78]. It is unclear whether this alternative choice will make training of the flow model more efficient. Last but not least, in conjunction with an alternative optimization approach, one could also explore the possibility of direct sampling of proton configurations with the normalizing flow.

It is believed there is a first-order atomic-to-molecular transition in the phase diagram of hydrogen [1]. Near the phase transition, there are significant difficulties related to slow equilibrium time or even lack of ergodicity in the Monte Carlo sampling of multimodal probability distributions [24]. This difficulty does not magically disappear in the variational free energy approach as they will show up as metastable states in the optimization landscape. Fortunately, the variational nature of our approach provides a clear guidance to judge and improve the qualities of various calculations, as one can always choose to believe the solution with the lowest free energy. We regard this as the most appealing feature of the present approach over the previous ones based on nested Monte Carlo sampling [16,17].

Deep variational free energy optimization presented in this Letter is a general computational framework. Both the nuclear probability distribution and the electronic wave function are open to further extensions. Thus, the framework holds the promise to be applied to a broad range of finite-temperature quantum matters beyond dense hydrogen.

[*]wang_han@iapcm.ac.cn
[†]linfeng.zhang.zlf@gmail.com
[‡]wanglei@iphy.ac.cn

[1] Jeffrey M. McMahon, Miguel A. Morales, Carlo Pierleoni, and David M. Ceperley, The properties of hydrogen and helium under extreme conditions, Rev. Mod. Phys. **84,** 1607 (2012).

[2] E. Wigner and H. B. Huntington, On the possibility of a metallic modification of hydrogen, J. Chem. Phys. **3,** 764 (1935).

[3] N. W. Ashcroft, Metallic Hydrogen: A High-Temperature Superconductor?, Phys. Rev. Lett. **21,** 1748 (1968).

[4] Ravit Helled, Guglielmo Mazzola, and Ronald Redmer, Understanding dense hydrogen at planetary conditions, Nat. Rev. Phys. **2,** 562 (2020).

[5] Setsuo Ichimaru, Nuclear fusion in dense plasmas, Rev. Mod. Phys. **65,** 255 (1993).

[6] R. Car and M. Parrinello, Unified Approach for Molecular Dynamics and Density-Functional Theory, Phys. Rev. Lett. **55,** 2471 (1985).

[7] Sam Azadi and W. M. C. Foulkes, Fate of density functional theory in the study of high-pressure solid hydrogen, Phys. Rev. B **88,** 014115 (2013).

[8] Thomas B. Blank, Steven D. Brown, August W. Calhoun, and Douglas J. Doren, Neural network models of potential energy surfaces, J. Chem. Phys. **103,** 4129 (1995).

[9] Weile Jia, Han Wang, Mohan Chen, Denghui Lu, Lin Lin, Roberto Car, Weinan E, and Linfeng Zhang, Pushing the limit of molecular dynamics with *ab initio* accuracy to 100 million atoms with machine learning, in *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis* (IEEE, New York, 2020), pp. 1–14.

[10] Bingqing Cheng, Guglielmo Mazzola, Chris J. Pickard, and Michele Ceriotti, Evidence for supercritical behaviour of high-pressure liquid hydrogen, Nature (London) **585,** 217 (2020).

[11] Valentin V. Karasiev, Joshua Hinz, S. X. Hu, and S. B. Trickey, On the liquid–liquid phase transition of dense hydrogen, Nature (London) **600,** E12 (2021).

[12] Hongxiang Zong, Heather Wiebe, and Graeme J. Ackland, Understanding high pressure molecular hydrogen with a hierarchical machine-learned potential, Nat. Commun. **11,** 5014 (2020).

[13] Andrea Tirelli, Giacomo Tenti, Kousuke Nakano, and Sandro Sorella, High-pressure hydrogen by machine learning and quantum Monte Carlo, Phys. Rev. B **106,** L041105 (2022).

[14] Hongwei Niu, Yubo Yang, Scott Jensen, Markus Holzmann, Carlo Pierleoni, and David M. Ceperley, Stable Solid Molecular Hydrogen above 900 K from a Machine-Learned Potential Trained with Diffusion Quantum Monte Carlo, Phys. Rev. Lett. **130,** 076102 (2023).

[15] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73,** 33 (2001).

[16] Carlo Pierleoni, David M. Ceperley, and Markus Holzmann, Coupled Electron-Ion Monte Carlo Calculations of Dense Metallic Hydrogen, Phys. Rev. Lett. **93,** 146402 (2004).

[17] Claudio Attaccalite and Sandro Sorella, Stable Liquid Hydrogen at High Pressure by a Novel *Ab Initio* Molecular-Dynamics Calculation, Phys. Rev. Lett. **100,** 114501 (2008).

[18] Rémi Bardenet, Arnaud Doucet, and Chris Holmes, On Markov Chain Monte Carlo methods for tall data, J. Mach. Learn. Res. **18,** 1 (2017), https://jmlr.org/papers/v18/15-205.html.

[19] A. D. Kennedy and J. Kuti, Noise without Noise: A New Monte Carlo Method, Phys. Rev. Lett. **54,** 2473 (1985).

[20] D. M. Ceperley and M. Dewing, The penalty method for random walks with uncertain energies, J. Chem. Phys. **110,** 9812 (1999).

[21] Florian R. Krajewski and Michele Parrinello, Linear scaling electronic structure calculations and accurate statistical mechanics sampling with noisy forces, Phys. Rev. B **73,** 041105(R) (2006).

[22] Max Welling and Yee Whye Teh, Bayesian learning via stochastic gradient Langevin dynamics, in *Proceedings of the 28th International Conference on Machine Learning* (OmniPress, Madison, WI, USA, 2011), pp. 681–688.

[23] Ye Luo, Andrea Zen, and Sandro Sorella, *Ab initio* molecular dynamics with noisy forces: Validating the quantum Monte Carlo approach with benchmark calculations of molecular vibrational properties, J. Chem. Phys. **141,** 194112 (2014).

[24] Guglielmo Mazzola and Sandro Sorella, Accelerating *ab initio* Molecular Dynamics and Probing the Weak Dispersive Forces in Dense Liquid Hydrogen, Phys. Rev. Lett. **118,** 015703 (2017).

[25] M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler, Backflow correlations for the electron gas and metallic hydrogen, Phys. Rev. E **68,** 046707 (2003).

[26] Carlo Pierleoni, Kris T. Delaney, Miguel A. Morales, David M. Ceperley, and Markus Holzmann, Trial wave functions for high-pressure metallic hydrogen, Comput. Phys. Commun. **179,** 89 (2008).

[27] Albrecht Huber, Variational principles in quantum statistical mechanics, in *Mathematical Methods in Solid State and Superfluid Theory*, edited by R. C. Clark and G. H. Derrick (Springer US, Boston, MA, 1968), pp. 364–392, 10.1007/978-1-4899-6435-9_14.

[28] Note that Ref. [29] opted to integrate the Bloch equation for Gaussian density matrix instead of optimizing the variational free energy [Eq. (1)].

[29] Burkhard Militzer and E. L. Pollock, Variational density matrix method for warm, condensed matter: Application to dense hydrogen, Phys. Rev. E **61,** 3470 (2000).

[30] Lei Wang, Generative models for physicists (2018), http://wangleiphy.github.io/lectures/PILtutorial.pdf.

[31] Shuo-Hui Li and Lei Wang, Neural Network Renormalization Group, Phys. Rev. Lett. **121,** 260601 (2018).

[32] Dian Wu, Lei Wang, and Pan Zhang, Solving Statistical Mechanics Using Variational Autoregressive Networks, Phys. Rev. Lett. **122,** 080602 (2019).

[33] Linfeng Zhang, Weinan E, and Lei Wang, Monge-Ampère flow for generative modeling, arXiv:1809.10188.

[34] Kim A. Nicoli, Shinichi Nakajima, Nils Strodthoff, Wojciech Samek, Klaus-Robert Müller, and Pan Kessel, Asymptotically unbiased estimation of physical observables with neural samplers, Phys. Rev. E **101,** 023304 (2020).

[35] M. S. Albergo, G. Kanwar, and P. E. Shanahan, Flow-based generative models for Markov Chain Monte Carlo in lattice field theory, Phys. Rev. D **100,** 034515 (2019).

[36] Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan, Equivariant Flow-Based Sampling for Lattice Gauge Theory, Phys. Rev. Lett. **125,** 121601 (2020).

[37] Kim A. Nicoli, Christopher J. Anders, Lena Funcke, Tobias Hartung, Karl Jansen, Pan Kessel, Shinichi Nakajima, and Paolo Stornati, Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models, Phys. Rev. Lett. **126,** 032001 (2021).

[38] Peter Wirnsberger, George Papamakarios, Borja Ibarz, Sébastien Racanière, Andrew J. Ballard, Alexander Pritzel, and Charles Blundell, Normalizing flows for atomic solids, Mach. Learn. Sci. Technol. **3,** 025009 (2022).

[39] Rasool Ahmad and Wei Cai, Free energy calculation of crystalline solids using normalizing flows, Model. Simul. Mater. Sci. Eng. **30,** 065007 (2022).

[40] Hao Xie, Linfeng Zhang, and Lei Wang, Ab-initio study of interacting fermions at finite temperature with neural canonical transformation, J. Mach. Learn. **1,** 38 (2022).

[41] Hao Xie, Linfeng Zhang, and Lei Wang, m^* of two-dimensional electron gas: A neural canonical transformation study, SciPost Phys. **14,** 154 (2023).

[42] Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley, Stochastic variational inference, J. Mach. Learn. Res. **14,** 1303 (2013), https://jmlr.org/papers/v14/hoffman13a.html.

[43] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.126501 for (a) proton kinetic energy contributions to the thermodynamic quantities, (b) model architectures, (c) finite-size analysis, (d) optimization details, and (e) illustrative calculation of the molecular phase.

[44] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, J. Mach. Learn. Res. **22,** 1 (2021), https://jmlr.org/papers/v22/19-1028.html.

[45] David Pfau, James S. Spencer, Alexander G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2,** 033429 (2020).

[46] Jonas Köhler, Leon Klein, and Frank Noé, Equivariant flows: Sampling configurations for multi-body systems with symmetric energies, arXiv:1910.00753.

[47] Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind, Automatic differentiation in machine learning: A survey, J. Mach. Learn. Res. **18,** 1 (2018), https://jmlr.org/papers/v18/17-468.html.

[48] Gabriel Pescia, Jiequn Han, Alessandro Lovato, Jianfeng Lu, and Giuseppe Carleo, Neural-network quantum states for periodic systems in continuous space, Phys. Rev. Res. **4,** 023138 (2022).

[49] Max Wilson, Saverio Moroni, Markus Holzmann, Nicholas Gao, Filip Wudarski, Tejs Vegge, and Arghya Bhowmik, Neural network ansatz for periodic wave functions and the homogeneous electron gas, Phys. Rev. B **107,** 235139 (2023).

[50] Gino Cassella, Halvard Sutterud, Sam Azadi, N. D. Drummond, David Pfau, James S. Spencer, and W. M. C. Foulkes, Discovering Quantum Phase Transitions with Fermionic Neural Networks, Phys. Rev. Lett. **130,** 036401 (2023).

[51] Xiang Li, Zhe Li, and Ji Chen, *Ab initio* calculation of real solids via neural network ansatz, Nat. Commun. **13,** 7895 (2022).

[52] Federico Becca and Sandro Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, England, 2017), 10.1017/9781316417041.

[53] Nicholas Gao and Stephan Günnemann, Ab-initio potential energy surfaces by pairing GNNs with neural wave functions, in *International Conference on Learning Representations* (2022).

[54] Michael Scherbela, Rafael Reisenhofer, Leon Gerard, Philipp Marquetand, and Philipp Grohs, Solving the electronic Schrödinger equation for multiple nuclear geometries with weight-sharing deep neural networks, Nat. Comput. Sci. **2,** 331 (2022).

[55] C. Lin, F. H. Zong, and D. M. Ceperley, Twist-averaged boundary conditions in continuum quantum Monte Carlo algorithms, Phys. Rev. E **64,** 016702 (2001).

[56] Y. Miguel, T. Guillot, and L. Fayon, Jupiter internal structure: The effect of different equations of state, Astron. Astrophys. **596,** A114 (2016).

[57] Miguel A. Morales, Carlo Pierleoni, and D. M. Ceperley, Equation of state of metallic hydrogen from coupled electron-ion Monte Carlo simulations, Phys. Rev. E **81,** 021202 (2010).

[58] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz, *Statistical Physics: Volume 5* (Elsevier, New York, 2013).

[59] D. Saumon, G. Chabrier, and H. M. van Horn, An equation of state for low-mass stars and giant planets, Astrophys. J. Suppl. Ser. **99,** 713 (1995).

[60] Guglielmo Mazzola, Ravit Helled, and Sandro Sorella, Phase Diagram of Hydrogen and a Hydrogen-Helium Mixture at Planetary Conditions by Quantum Monte Carlo Simulations, Phys. Rev. Lett. **120,** 025701 (2018).

[61] B. Militzer and D. M. Ceperley, Path integral Monte Carlo simulation of the low-density hydrogen plasma, Phys. Rev. E **63,** 066404 (2001).

[62] Guglielmo Mazzola, Seiji Yunoki, and Sandro Sorella, Unexpectedly high pressure for molecular dissociation in liquid hydrogen by electronic simulation, Nat. Commun. **5,** 3487 (2014).

[63] See https://github.com/fermiflow/hydrogen for code written in Jax [64].

[64] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang, JAX: Composable transformations of Python + NumPy programs (2018), https://github.com/google/jax.

[65] Carlo Pierleoni, Miguel A. Morales, Giovanni Rillo, Markus Holzmann, and David M. Ceperley, Liquid–liquid phase transition in hydrogen by coupled electron–ion Monte Carlo simulations, Proc. Natl. Acad. Sci. U.S.A. **113,** 4953 (2016).

[66] Linfeng Zhang, Jiequn Han, Han Wang, Roberto Car, and Weinan E, Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics, Phys. Rev. Lett. **120,** 143001 (2018).

[67] D. M. Ceperley, Path integral Monte Carlo methods for fermions, in *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*, edited by K. Binder and G. Ciccotti (Editrice Compositori, Bologna, Italy, 1996).

[68] C. Pierleoni, D. M. Ceperley, B. Bernu, and W. R. Magro, Equation of State of the Hydrogen Plasma by Path Integral Monte Carlo Simulation, Phys. Rev. Lett. **73,** 2145 (1994).

[69] B. Militzer and D. M. Ceperley, Path integral Monte Carlo simulation of the low-density hydrogen plasma, Phys. Rev. E **63,** 066404 (2001).

[70] Maximilian Böhme, Zhandos A. Moldabekov, Jan Vorberger, and Tobias Dornheim, Static Electronic Density Response of Warm Dense Hydrogen: *Ab initio* Path Integral Monte Carlo Simulations, Phys. Rev. Lett. **129,** 066402 (2022).

[71] Richard P. Feynman, Albert R. Hibbs, and Daniel F. Styer, *Quantum Mechanics and Path Integrals* (Dover Publications, New York, 2010).

[72] Yang Li, Haidong Yi, Christopher Bender, Siyuan Shan, and Junier B Oliva, Exchangeable neural ode for set modeling, in *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc., 2020), Vol. 33, pp. 6936–6946.

[73] Marin Biloš and Stephan Günnemann, Scalable normalizing flows for permutation invariant densities, in *Proceedings of the 38th International Conference on Machine Learning*, Proceedings of Machine Learning Research Vol. 139, pp. 957–967.

[74] Peter Wirnsberger, Andrew J. Ballard, George Papamakarios, Stuart Abercrombie, Sébastien Racanière, Alexander Pritzel, Danilo Jimenez Rezende, and Charles Blundell, Targeted free energy estimation via learned mappings, J. Chem. Phys. **153**, 144112 (2020).

[75] James Martens and Roger Grosse, Optimizing neural networks with kronecker-factored approximate curvature, in *Proceedings of the 32nd International Conference on Machine Learning*, Proceedings of Machine Learning Research Vol. 37, edited by Francis Bach and David Blei (PMLR, Lille, France, 2015), pp. 2408–2417.

[76] Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih, Monte Carlo gradient estimation in machine learning, J. Mach. Learn. Res. **21**, 1 (2020), https://jmlr.org/papers/v21/19-346.html.

[77] Sandro Sorella and Luca Capriotti, Algorithmic differentiation and the calculation of forces by quantum Monte Carlo, J. Chem. Phys. **133**, 234111 (2010).

[78] Yubing Qian, Weizhong Fu, Weiluo Ren, and Ji Chen, Interatomic force from neural network based variational quantum Monte Carlo, J. Chem. Phys. **157**, 164104 (2022).