Machine learning for molecular dynamics with strongly correlated electrons

Hidemaro Suwa,[1, 2] Justin S. Smith,[3] Nicholas Lubbers,[4] Cristian D. Batista,[1, 5] Gia-Wei Chern,[6] and Kipton Barros[3]

[1]*Department of Physics and Astronomy, The University of Tennessee, Knoxville, TN 37996, USA*
[2]*Department of Physics, The University of Tokyo, Tokyo 113-0033, Japan*
[3]*Theoretical Division and CNLS, Los Alamos National Laboratory, Los Alamos, NM 87544*
[4]*CCS Division, Los Alamos National Laboratory, Los Alamos, NM 87544*
[5]*Quantum Condensed Matter Division and Shull-Wollan Center,*
Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[6]*Department of Physics, University of Virginia, Charlottesville, VA 22904*
(Dated: November 6, 2018)

We use machine learning to enable large-scale molecular dynamics (MD) of a correlated electron model under the Gutzwiller approximation scheme. This model exhibits a Mott transition as a function of on-site Coulomb repulsion U. Repeated solution of the Gutzwiller self-consistency equations would be prohibitively expensive for large-scale MD simulations. We show that machine learning models of the Gutzwiller potential energy can be remarkably accurate. The models, which are trained with $N = 33$ atoms, enable highly accurate MD simulations at much larger scales ($N \gtrsim 10^3$). We investigate the physics of the smooth Mott crossover in the fluid phase.

Understanding strongly correlated electron systems is an outstanding challenge in condensed matter theory. Some of the simplest many-body models remain unsolved. This challenge persists despite steady progress in theoretical and computational methods. Machine learning (ML) is emerging as a promising tool to help model various types of many-body phenomena.

Because the many-body quantum state space grows exponentially with system size, the cost of direct numerical solution quickly becomes intractable. Unbiased quantum Monte Carlo (QMC) can be very effective in special cases (e.g., for the square-lattice Hubbard model at half-filling). But generally one is plagued by the "sign problem" of resolving the delicate signal that remains after cancellations between samples with complex phases [1]. Many clever mitigation strategies have emerged [2–6]. An intriguing possibility is to use ML to extract relevant physics from QMC samples *without* reweighting [7, 8].

Alternatively, one may seek to represent the many-body state variationally, e.g. with the density matrix renormalization group (DMRG) [9, 10] or tensor network generalizations. ML is inspiring new variational ansatzes [11–14] that compare favorably with previous ones.

To make quantitative predictions for real correlated electron materials, one commonly employs physically-motivated approximations, such as variational and fixed node QMC [15] and dynamical mean-field theory (DMFT) [16]. These methods remain computationally demanding. Again, ML presents new opportunities. For example, ML may be useful as a low cost surrogate model for the impurity solver within DMFT [17] or even the full DMFT calculation itself [18].

An emerging research area is the molecular dynamics (MD) of strongly correlated electron materials. Developing such toolkits not only is of fundamental importance, but also has important technological implications. While quantum MD methods based on density functional theory (DFT) have been successfully applied to a wide variety of materials, they have limited validity in their treatment of electron correlations. On the other hand, most of the many-body techniques mentioned above are computationally too costly for MD simulations.

ML offers the possibility of large-scale MD simulations by emulating the time-consuming quantum calculations required at each time-step. Indeed, ML has already proven extremely effective in modeling MD potentials for chemistry and materials applications [19–30]. An ML model might be trained from a dataset containing 10^4–10^6 individual atomic forces, often calculated with DFT. In organic chemistry, ML now routinely predicts molecular energies that agree with new DFT calculations to within 0.04 eV (1 kcal/mol) [22, 26, 27], whereas DFT itself is almost certainly not this accurate. This success has spurred recent efforts to calculate the training data at levels of quantum theory significantly beyond DFT [31–33].

Here we show that ML can be similarly effective for building fast, linear-scaling MD potentials that capture correlated electron physics. Specifically, we use ML to enable large-scale Gutzwiller MD simulations of a liquid Hubbard model [34]. The correlated electronic state is computed using an efficient Gutzwiller method at every time-step. Contrary to DFT, the Gutzwiller approach captures crucial correlation effects such as Mott metal-insulator transition [34]. For our neural network model [22] running on a single modern GPU, a typical force evaluation costs $\sim 10\,\mu$s per atom. For the systems considered in the present study, ML is up to 6 orders of magnitude faster than direct quantum calculations.

The present work opens a path toward large-scale dynamical simulation of realistic models of correlated materials. Future studies could train ML models on data generated from small scale QMC or DMFT calculations. ML works well assuming *locality*, i.e., that the total energy can be decomposed as a sum of local contributions [35]. Our success in this study offers evidence that locality can remain valid in the presence of strong electron cor-

relations.

Following earlier work [34], we consider a single-orbital model Hamiltonian

$$\mathcal{H}[\mathbf{r}, \mathbf{p}] = \mathcal{H}_{\text{Hubbard}}[\mathbf{r}] + V_{\text{pair}}[\mathbf{r}] + E_{\text{kin}}[\mathbf{p}], \qquad (1)$$

where $\{\mathbf{r}_i\}$ and $\{\mathbf{p}_i\}$ are the positions and momentum of the nuclei. The electronic part

$$\mathcal{H}_{\text{Hubbard}} = -\sum_{i \neq j} \sum_{\sigma} t_{ij} c_{i\sigma}^{\dagger} c_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow}, \qquad (2)$$

has hopping and on-site Coulomb repulsion terms. The operator $c_{i\sigma}^{\dagger}$ creates an electron with spin $\sigma \in \{\uparrow, \downarrow\}$ on the ith atom, and $n_{i\sigma} = c_{i\sigma}^{\dagger} c_{i\sigma}$ is the number operator. We take the system to be half filled (one electron per nucleus). The hoppings $t_{ij} = t(|\mathbf{r}_i - \mathbf{r}_j|)$ decay exponentially with distance between nuclei, $t(r) = t_0 \exp(-r/\xi_1)$. The pair repulsions $V_{\text{pair}} = \sum_{i \neq j} \phi_{ij}/2$ also decay exponentially, $\phi_{ij} = \phi(|\mathbf{r}_i - \mathbf{r}_j|)$ with $\phi(r) = \phi_0 \exp(-r/\xi_2)$. Selecting $t_0 = 24\,\text{eV}$, $\xi_1 = 0.526\,\text{Å}$, $\phi_0 = 100\,\text{eV}$, and $\xi_2 = 0.283\,\text{Å}$ gives a highly simplified model of hydrogen. With these choices, at $U = 0$ the dimer molecule is bound with energy $-4.58\,\text{eV}$ at distance $r_0 = 0.83\,\text{Å}$, in loose agreement with the physical values of $-4.52\,\text{eV}$ and $0.74\,\text{Å}$. Our model clearly departs from hydrogen, however, in the large range of Hubbard U values that we consider: $0\,\text{eV} \leq U \leq 17\,\text{eV}$ [36]. Finally, our Hamiltonian includes a kinetic energy term $E_{\text{kin}} = \sum_i |\mathbf{p}_i|^2/2m$, where $m \approx 1\,\text{amu}$ is the mass of the proton.

Gutzwiller method. To estimate the electronic free energy at nonzero U, we employ a finite temperature generalization of the Gutzwiller projection method [37, 38]. In this approach, we seek a variational approximation $\rho_G = \mathcal{P}\rho_0\mathcal{P}$ for the density matrix $\exp(-\beta\mathcal{H}_{\text{Hubbard}})$. Here ρ_0 is the Boltzmann distribution of free quasiparticles, and the so-called Gutzwiller projection operator $\mathcal{P} = \prod_i \mathcal{P}_i$ effectively reweights electron occupation numbers at each site. We consider the paramagnetic phase of the Hubbard model, which means $n_{i,\uparrow} = n_{i,\downarrow}$. The variational target is to minimize a free energy $F_G = \langle\mathcal{H}_{\text{Hubbard}}\rangle_G - TS_G$, where $S_G = \text{Tr}\,(\rho_G/Z_G)\ln(\rho_G/Z_G)$, subject to the Gutzwiller constraint $\langle n_i\rangle_G = \langle n_i\rangle_0$. We use $\langle\mathcal{O}\rangle_G$ and $\langle\mathcal{O}\rangle_0$ to denote expectations of $\mathcal{O}$ computed from density matrices ρ_G and ρ_0, respectively. Importantly, the expectation values of the hopping terms can be efficiently computed using the so-called Gutzwiller approximation [39] (exact in infinite dimensions): $\langle c_i^{\dagger} c_j\rangle_G = \mathcal{R}_i\mathcal{R}_j \langle c_i^{\dagger} c_j\rangle_0$, where the renormalization factor $\mathcal{R}_i$ is uniquely determined from the electron density $n_i^0 = \langle n_i\rangle_G$ and double occupancy $d_i = \langle n_{i\uparrow} n_{i\downarrow}\rangle_G$ [40].

Although S_G cannot be evaluated exactly, a good approximation to the free energy is $F_G = F_0 + U\sum_i d_i - T\Delta S$. The term $F_0 = -k_B T \ln \text{Tr}\rho_0$ is the free energy of quasi-particles and ΔS is the entropy correction due to the projector $\mathcal{P}$ [37].

Self-consistent minimization of F_G with respect to the variational parameters produces the electronic free energy of interest. This minimization is performed by cycling between two steps [40]. (1) For fixed n_i^0 and d_i, the renormalized Hamiltonian

$$\mathcal{H}_0 = -\sum_{i \neq j} \mathcal{R}_i\mathcal{R}_j t_{ij} c_i^{\dagger} c_j - \sum_i \mu_i n_i \qquad (3)$$

is diagonalized to obtain ρ_0. The Lagrange multipliers μ_i impose the density constraint. (2) For fixed ρ_0, one adjusts n_i^0 and d_i to minimize F_G.

Once converged, we treat $V_{\text{elec}} = \min F_G$ as the electronic part of the total MD potential, $V = V_{\text{elec}} + V_{\text{pair}}$. The corresponding forces,

$$-\frac{\partial V}{\partial \mathbf{r}_i} = 2\sum_j \frac{\partial t_{ij}}{\partial \mathbf{r}_i}\mathcal{R}_i\mathcal{R}_j \langle c_i^{\dagger} c_j\rangle_0 - \sum_j \frac{\partial \phi_{ij}}{\partial \mathbf{r}_i}, \qquad (4)$$

drive MD simulations of the nuclei under the Born Oppenheimer approximation. To derive Eq. (4), we used the fact that F_G is minimized with respect to the variational parameters.

The scheme here is largely similar to that of our previous Gutzwiller-MD study [34]. The primary difference is that, here, we *reinitialize* the variational parameters at each MD time-step before iteratively optimizing them. In the prior version of our code, we selected the initial guess for (n_i^0, d_i) as the self-consistent solution obtained in the previous MD time-step. That lack of reinitialization leads to weakly stable solution branches of the Gutzwiller equations, and strong hysteresis. In the present study, we enforce that V_{elec} is single-valued by reinitializing (n_i^0, d_i) to default values at each time-step before iteratively solving the self-consistency equations. This scheme eliminates hysteresis and simultaneously lowers the Gutzwiller variational free energy.

Machine learning. Solving the above Gutzwiller equations for the MD potential $V_{\text{elec}}[\mathbf{r}]$ at each time-step can be computationally expensive. We use an ML model to estimate $\hat{V}_{\text{elec}} \approx V_{\text{elec}}$, while treating V_{pair} exactly. A key assumption is the (non-unique) decomposition of energy as a sum of local contributions, $\hat{V}_{\text{elec}} = \sum_{i=1}^{N} \hat{V}_{\text{elec};i}$, where $V_{\text{elec};i}$ is a function only of the atomic environment near atom i. We use a neural network to model the local potential $V_{\text{elec};i}$. We have explored two established neural network architectures, HIP-NN [26] and ANI [22]. Although HIP-NN may yield slightly better accuracies, we selected ANI for our MD simulations because of its highly optimized NeuroChem implementation.

ANI constructs a rotationally and translationally invariant representation of the environment near atom i from two types of information: (1) the pairwise distances $\{r_{ij}\}$ for all atoms j satisfying $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j| < 2.8\,\text{Å}$, and (2) the set of three-body angles $\{\theta_{jik}\}$, where $\cos\theta_{jik} = \mathbf{r}_{ij} \cdot \mathbf{r}_{ik}/r_{ij}r_{ik}$, provided that both r_{ij} and r_{ik} are less than $2.0\,\text{Å}$. The values $\{r_{ij}\}$ and $\{\theta_{jik}\}$ transformed into a fixed-length "feature vector" $\{G_{i,m}\}_{m=1...M_0}$ using continuous binning. In this study, $G_{i,m}$ contains $M_0 = 96$ scalar components. There will typically be about 5–15
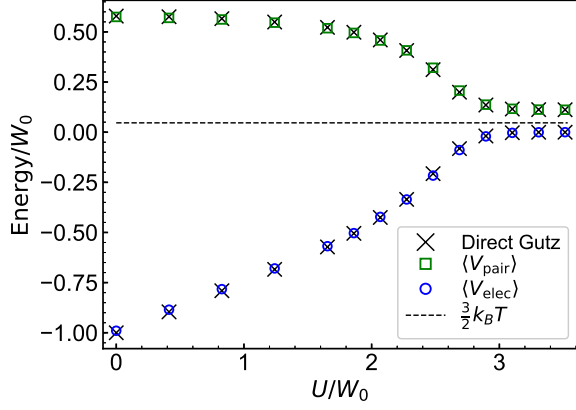
Figure 1. MD averaged electronic free energy $\langle V_{\mathrm{elec}} \rangle$ and pair repulsion energy $\langle V_{\mathrm{pair}} \rangle$, for varying Hubbard U. At every MD time-step, forces are produced by an ML model that emulates the Gutzwiller calculation. The crosses represent reference simulations that use direct Gutzwiller calculations rather than ML. The dashed line represents the kinetic energy. The energy scale is $W_0 = -\langle V_{\mathrm{elec}} \rangle|_{U=0} = 4.84\,\mathrm{eV}$. Tests show that these results with $N = 100$ atoms are representative of $N \to \infty$.
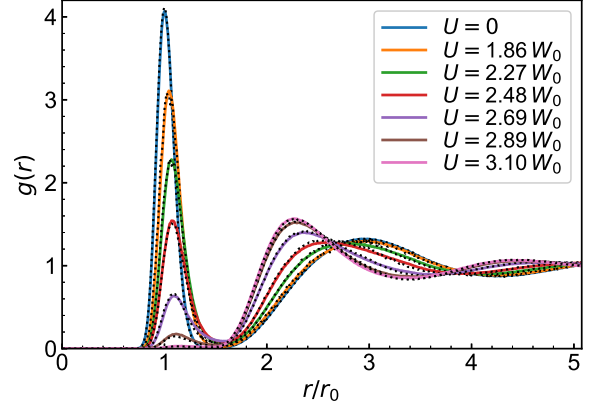

Figure 2. Radial distribution functions $g(r)$ for varying Hubbard U, and the same simulation parameters as in Fig. 1. The results using ML (colored lines) are in strong agreement with direct Gutzwiller calculations (black dots). The typical bond length at $U = 0$ is $r_0 = 0.83\,\text{Å}$.

atoms within a distance $2.8\,\text{Å}$ of atom i, and $G_{i,m}$ describes this environment.

The neural network $\hat{V}_{\mathrm{elec},i}$ is composed of multiple real-valued activations $\{z_m^\ell\}_{m=1\ldots M_\ell}$ for layer index $\ell = 0 \ldots L$. The input to the neural network is the feature vector, $z_m^{\ell=0} = G_{i,m}$. Activations are evaluated using fully connected layers, $z_m^{\ell+1} = f_{\mathrm{activ}}(\sum_{n=1}^{M_\ell} w_{mn}^\ell z_n^\ell + b_m^\ell)$. The matrix elements w_{mn}^ℓ and offsets b_m^ℓ are learnable parameters. We select $f_{\mathrm{activ}}(x)$ to be the CELU activation function [41]. A linear combination of the final activations yields the local potential, $\hat{V}_{\mathrm{elec};i} = \sum_{n=1}^{M_L} w_n^L z_n^L + b^L$. We select $L = 3$ and layer sizes $M_0 = 96$, $M_1 = 48$, $M_2 = 32$, and $M_3 = 16$. There are thus approximately 10^4 learnable parameters in this neural network.

Training of the model parameters involves optimizing a loss function $\mathcal{L}$ that quantifies the deviation between the model $\hat{V}_{\mathrm{elec}}$ and direct Gutzwiller calculations V_{elec}, evaluated on a training dataset. The loss function incorporates errors in the potential, $\hat{V}_{\mathrm{elec}} - V_{\mathrm{elec}}$ and forces $\nabla_{\mathbf{r}_i}(\hat{V}_{\mathrm{elec}} - V_{\mathrm{elec}})$. We optimize the parameters w_{mn}^ℓ and b_m^ℓ using the ADAM variant of stochastic gradient descent [42]. To mitigate overfitting, we employ weight decay regularization with $\alpha = 10^{-6}$ [43] and early stopping according to a validation dataset.

To produce our training dataset, we ran MD simulations with forces obtained from direct-Gutzwiller calculations. We used Verlet integration with a time-step of $0.2\,\mathrm{fs}$ to evolve the atoms. The simulation box has periodic boundary conditions, and its volume V_0 is set according to a fixed density $\varrho_0 = N/V_0 = 0.153\,\text{Å}^{-3}$. Just $N = 33$ atoms are sufficient to train an extensible ML potential, which remains valid for much larger N. To fix the temperature $k_B T = 0.15\,\mathrm{eV}$, we employ a Langevin thermostat with friction coefficient $\gamma = 5 \times 10^{-3}\,\mathrm{amu/fs}$. We generated independent training data sets for U values ranging from 0 to 17 eV. For each U, we collected 3×10^3 snapshots, one per 200 MD time-steps. Every snapshot contains the electron free energy and associated forces.

We found that an ML model trained on this dataset alone would lead to unstable MD simulations. To improve the robustness of our ML model, we collected additional data that sampled a broader range of the atomic configurations. Specifically, we augmented our training dataset by running additional direct-Gutzwiller MD with $k_B T = 0.075\,\mathrm{eV}$ and $k_B T = 0.3\,\mathrm{eV}$ for the Langevin thermostat only, *without* changing the temperature used in the Gutzwiller equations. Our total training dataset, per U value, thus contains 9×10^3 MD snapshots and about 3×10^5 atomic forces.

We reduce variance by averaging over an ensemble of eight neural networks, each trained on a subset of the data [44].

Results. We use our ML models to drive efficient and accurate MD simulations at large scales. Figure 1 shows mean energies for an MD simulation with $N = 100$ atoms (keeping $k_{\mathrm{B}}T = 0.15\,\mathrm{eV}$) and 3×10^6 time-steps. The mean electronic energy evaluated at $U = 0$ serves as a convenient energy scale, $W_0 = 4.84\,\mathrm{eV}$. At large U the electrons become localized, the electronic energy V_{elec} goes to zero, and atoms repel according to V_{pair}. At small U the system is metallic and V_{elec} is an attractive interaction that tends to bond atoms into dimer molecules, counterbalanced by V_{pair}.

For validation, we also performed reference simulations using forces from direct-Gutzwiller calculations at each MD time-step, and only $\approx 10^5$ time-steps. The ML and reference simulations in Fig. 1 are barely distinguishable, which is remarkable given that the training data was
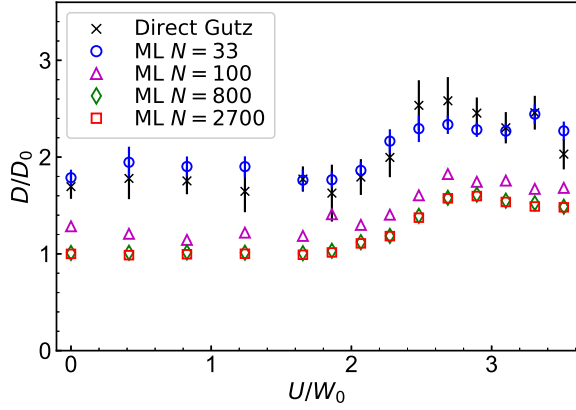
Figure 3. The diffusivity D for varying Hubbard U, and the same simulation parameters as in Fig. 1. At $U = 0$ with $N = 2700$ atoms, we measure $D_0 = 0.0088\,\text{Å}^2/\text{fs}$. Direct Gutzwiller calculations (black crosses) employ five times fewer time-steps than used in the ML-based simulations. Approximately $N = 800$ atoms are required to reach convergence.

Figure 4. A volume collapse occurs with decreasing U. Here we employ an ensemble with constant pressure $P = 12, 20, 40,$ and $80\,\text{GPa}$; the pressure of the previous simulations at fixed volume $V_0(N) = N/\varrho_0$ corresponds to $P = 20\,\text{GPa}$ when $U \approx 2.4\,W_0$. At this temperature, there is a relatively smooth Mott cross-over.

generated using only $N = 33$ atoms. We also directly compare the ML predictions $\hat{V}_{\text{elec}}$ with reference energy calculations V_{elec} for $N = 100$ and $N = 800$. The mean absolute error (MAE) scales as $0.02\sqrt{N}\,W_0$. The factor $\sqrt{N}$ appears because the errors in local contributions $\hat{V}_{\text{elec};i}$ and $\hat{V}_{\text{elec};j}$ are essentially independent for atoms $i \neq j$. The MAE of electronic force $-\nabla \hat{V}_{\text{elec}}$ is approximately $0.009\,f_0$, where $f_0 = 17.7\,\text{eV/Å}$ is the mean force magnitude at $U = 0$.

Figure 2 shows the radial distribution functions $g(r)$. At $U = 0$, $g(r)$ has a characteristic peak at $r_0 = 0.83\,\text{Å}$, which reflects the bond length of a dimer molecule. This peak gradually decreases with increasing U, and disappears entirely at $U \gtrsim 3.1\,W_0$, where $\langle V_{\text{elec}} \rangle \approx 0$ according to Fig. 1. The ML-based simulations again appear almost identical to direct-Gutzwiller reference simulations.

Figure 3 shows a *dynamic* observable, the diffusivity

$$D = \frac{1}{3N} \sum_{i=1}^{N} \int_0^\infty \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle \mathrm{d}t. \tag{5}$$

Replacing ∞ with a finite time τ would yield a naive estimator of D. Instead we use an estimator $\hat{D}(\tau) = \frac{1}{2\tau} \frac{1}{3N} \sum_i |\int_0^\tau \mathbf{v}_i(t)dt|^2$ with reduced variance. Varying τ, we find that the bias of the estimator $D \approx \langle \hat{D}(\tau_0) \rangle$ becomes negligible at τ_0 corresponding to 10^5 MD time-steps. We collected more than 10 independent samples of $\hat{D}(\tau_0)$ from each MD simulation and estimated the error bars using bootstrapping. D converges surprisingly slowly with system size; about $N \approx 800$ atoms seem to be required. Such simulations would have been extremely challenging without ML.

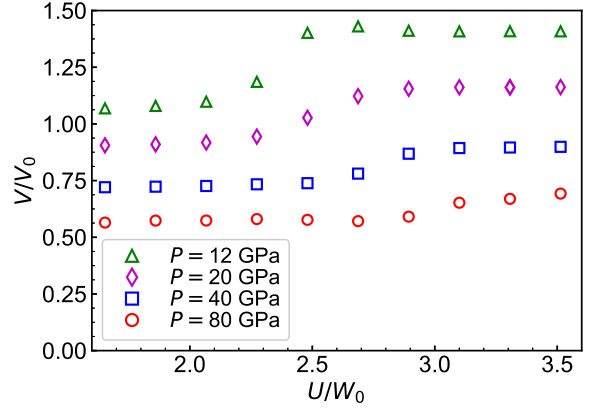Due to updates in our Gutzwiller solver, the present results deviate significantly from prior work [34]. The mean energies (Fig. 1) and $g(r)$ curves (Fig. 2) now vary smoothly with U, indicating a crossover rather than a first order transition. We also observe in Fig. 3 that D weakly rises throughout the *entire* Mott transition, whereas before we had observed a continuous rise in D followed by a sharp drop. The previous results were dependent on strong hysteresis effects. Here we reinitialize the Gutzwiller parameters at every time-step, which eliminates hysteresis and lowers the variational free energy in all instances we checked. We argue that the present approach is more consistent with the assumption (used in the finite-temperature Gutzwiller method) that the electrons are in equilibrium.

Finally, we investigate the nature of the Mott transition using large-scale MD with $N = 2700$ atoms. Here, we switch to an ensemble with constant pressure $P = 12, 20, 40,$ and $80\,\text{GPa}$, implemented using the Monte Carlo Barostat algorithm [45]. The pressure of our previous (fixed density) simulations matches $P = 20\,\text{GPa}$ when $U \approx 2.4\,W_0$. Figure 4 shows collapse of volume V with decreasing U, corresponding to the crossover from the Mott insulating phase to the metallic state.

We anticipate that the ML techniques applied here will also be relevant in the study of structural properties of real f-electron materials. It is surprising how accurately ML predicts Gutzwiller forces using only the atomic environment within a $2.8\,\text{Å}$ radius. The power of ML is that, under the locality assumption, only training data for relatively small system sizes is required. In future work, it should be feasible to generate training data from, e.g., quantum Monte Carlo or DMFT. Once the ML model has been trained, accurate MD simulations of unprecedented scale become practical. For MD studies with $N = 2700$ atoms, our ML model running on a Tesla P100 GPU can evaluate all forces in about 26 ms, which is $\approx 10^6$

times faster (extrapolated) than our reference Gutzwiller implementation running on a Intel Xeon CPU E5-2680. We anticipate that a distributed ML/MD implementation could readily enable simulations of f-electron materials with millions of atoms.

[1] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).
[2] G. Parisi, Physics Letters B **131**, 393 (1983).
[3] J. Bloch, Phys. Rev. D **95**, 054509 (2017).
[4] M. Cristoforetti, F. Di Renzo, and L. Scorzato (AuroraScience Collaboration), Phys. Rev. D **86**, 074506 (2012).
[5] M. Fukuma and N. Umeda, Prog. Theor. Exp. Phys. **2017**, 073B01 (2017).
[6] D. Frame, R. He, I. Ipsen, D. Lee, D. Lee, and E. Rrapaj, Phys. Rev. Lett. **121**, 032501 (2018).
[7] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, Sci. Rep. **7**, 8823 (2017).
[8] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Phys. Rev. X **7**, 031038 (2017).
[9] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[10] U. Schollwöck, Ann. Phys. **326**, 96 (2011).
[11] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[12] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Phys. Rev. B **96**, 205152 (2017).
[13] Z. Cai and J. Liu, Phys. Rev. B **97**, 035116 (2018).
[14] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, ArXiv e-prints (2018), arXiv:1803.09780 [quant-ph].
[15] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[16] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006).
[17] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Phys. Rev. B **90**, 155136 (2014).
[18] L.-F. Arsenault, O. Anatole von Lilienfeld, and A. J. Millis, ArXiv e-prints (2015), arXiv:1506.08858 [cond-mat.str-el].
[19] M. Rupp, Int. J. Quantum Chem. **115**, 1058 (2015).
[20] J. Behler, Int. J. Quantum Chem. **115**, 1032 (2015).
[21] A. Shapeev, Multiscale Model. Simul. **14**, 1153 (2016).
[22] J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. **8**, 3192 (2017).
[23] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Sci. Adv. **3**, e1701816 (2017).
[24] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, ArXiv e-prints (2017), arXiv:1704.01212.
[25] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Nat. Commun. **8**, 13890 (2017).

[26] N. Lubbers, J. S. Smith, and K. Barros, J. Chem. Phys. **148**, 241715 (2018).
[27] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).
[28] M. J. Willatt, F. Musil, and M. Ceriotti, ArXiv e-prints (2018), arXiv:1807.00236 [physics.chem-ph].
[29] D. Dragoni, T. D. Daff, G. Csányi, and N. Marzari, Phys. Rev. Materials **2**, 013808 (2018).
[30] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Nature **559**, 547 (2018).
[31] R. T. McGibbon, A. G. Taube, A. G. Donchev, K. Siva, F. Hernández, C. Hargus, K.-H. Law, J. L. Klepeis, and D. E. Shaw, J. Chem. Phys. **147**, 161725 (2017).
[32] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. Roitberg, ChemRXiv e-prints (2018), 10.26434/chemrxiv.6744440.v1, ChemRXiv:6744440.
[33] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, ArXiv e-prints (2018), arXiv:1802.09238 [physics.chem-ph].
[34] G.-W. Chern, K. Barros, C. D. Batista, J. D. Kress, and G. Kotliar, Phys. Rev. Lett. **118**, 226401 (2017).
[35] Locality is not strictly required; long-range interactions of known form (i.e. classical Coulomb interactions) can be added to the ML potential by hand.
[36] G. Chiappe, E. Louis, E. SanFabián, and J. A. Verges, Phys. Rev. B **75**, 195104 (2007).
[37] W.-S. Wang, X.-M. He, D. Wang, Q.-H. Wang, Z. D. Wang, and F. C. Zhang, Phys. Rev. B **82**, 125105 (2010).
[38] M. Sandri, M. Capone, and M. Fabrizio, Phys. Rev. B **87**, 205108 (2013).
[39] M. C. Gutzwiller, Phys. Rev. **137**, A1726 (1965).
[40] N. Lanatà, H. U. R. Strand, X. Dai, and B. Hellsing, Phys. Rev. B **85**, 035133 (2012).
[41] J. T. Barron, ArXiv e-prints (2017), arXiv:1704.07483.
[42] D. P. Kingma and J. Ba, ArXiv e-prints (2014), arXiv:1412.6980.
[43] I. Loshchilov and F. Hutter, ArXiv e-prints (2017), arXiv:1711.05101.
[44] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, J. Chem. Phys. **148**, 241733 (2018).
[45] R. Faller and J. J. de Pablo, J. Chem. Phys. **116**, 55 (2002).