Pure density functional for strong correlation and the thermodynamic limit from machine learning

Li Li (□□), Thomas E. Baker, Steven R. White, and Kieron Burke

Phys. Rev. B **94**, 245129 — Published 21 December 2016

DOI: 10.1103/PhysRevB.94.245129

Pure density functional for strong correlation and the thermodynamic limit from machine learning

Li Li (李力),[1] Thomas E. Baker,[1] Steven R. White,[1] and Kieron Burke[2,1]

[1] *Department of Physics and Astronomy, University of California, Irvine, CA 92697*
[2] *Department of Chemistry, University of California, Irvine, CA 92697*

(Dated: November 11, 2016)

We use the density-matrix renormalization group, applied to a one-dimensional model of continuum Hamiltonians, to accurately solve chains of hydrogen atoms of various separations and numbers of atoms. We train and test a machine-learned approximation to $F[n]$, the universal part of the electronic density functional, to within quantum chemical accuracy. We also develop a data-driven, atom-centered basis set for densities which greatly reduces the computational cost and accurately represents the physical information in the machine learning calculation. Our calculation (a) bypasses the standard Kohn-Sham approach, avoiding the need to find orbitals, (b) includes the strong correlation of highly-stretched bonds without any specific difficulty (unlike all standard DFT approximations) and (c) is so accurate that it can be used to find the energy in the thermodynamic limit to quantum chemical accuracy.

I. INTRODUCTION

Although widely used in solid-state physics, chemistry, and materials science [1], Kohn-Sham density functional theory (KS-DFT) with standard approximations fails for strong correlation [2, 3]. The prototype is the H_2 molecule. When stretched, the electrons localize on each site while remaining in a singlet state, but this is not captured by such calculations [4]. These kinds of difficulties have led to the popularity of many beyond-DFT schemes, ranging from the simple addition [5] of a Hubbard U to the use of dynamical mean field theory [6] as well as many others.

But even KS-DFT is too slow for many large calculations, such as those using classical MD or continuum mechanics [7]. The original DFT, first suggested in the Thomas-Fermi approximation [8, 9] and later justified by the Hohenberg-Kohn theorem [10], uses only pure functionals of the total density, $n(\mathbf{r})$. This 'orbital-free' version has the potential to be much faster than even the most efficient KS implementations, because the KS equations need not be solved [11]. Several recent attempts have constructed machine learning (ML) kinetic energy functionals specifically to bypass this step [12–15]. These are designed to be used in conjunction with standard KS approximations to speed up such calculations, but not to improve their accuracy.

Meanwhile, beyond the world of DFT, density matrix renormalization group (DMRG) has become a standard tool for finding extremely accurate solutions to strongly correlated lattice problems [16–19]. In recent years, a one-dimensional analog of ab-initio Hamiltonians has been developed, using typically about 20 grid points per atom and interactions involving many grid points, with the express purpose of rapidly exploring both conceptual and practical issues in DFT [3, 20–23]. A particular advantage is that, since 2000 grid points is routinely accessible, this includes up to 100 atoms, and extrapolations to the thermodynamic limit are much easier than in three dimensions. Applications include a demonstration of the behavior of the KS gap in a Mott-Hubbard insulator [20] and a proof of convergence of the KS equations with the exact functional, regardless of the starting point or strength of correlation [21].

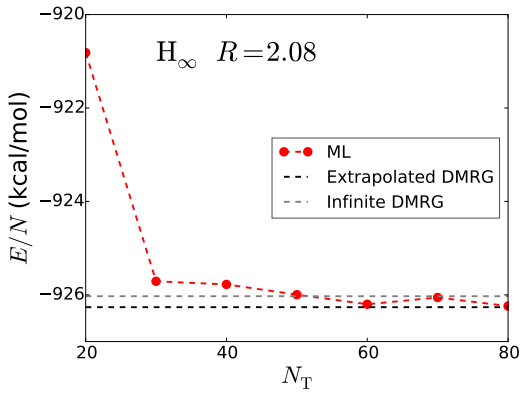

FIG. 1. (Color online) Electronic energy of infinite chain from model learned from extrapolated chain densities and energies. The accurate value was calculated with infinite DMRG (see text).

In the present work, we combine all these methodologies to demonstrate several important features. We perform DMRG calculations on a variety of one-dimensional hydrogen atom chains, with from 2 to 20 atoms, and whose interatomic spacing R varies from 1 to 10 Bohr radii, and use these to train a ML model of $F[n]$, the 'universal' part of the density functional identified by Hohenberg-Kohn. This simulanteously includes both the non-interacting kinetic energy sought in orbital-free DFT and the exchange-correlation energy that is approximated in KS calculations. We demonstrate that, with reasonable amounts of training, we can *self-consistently* calculate densities and energies for these chains at new values of R, outside the training set, with quantum chemical accuracy. This includes highly stretched systems which are strongly correlated, and where all popular

DFT approximations fail. We furthermore extrapolate the DMRG densities from the center of finite chains to the infinite chain limit, i.e., a 1d solid. We train a new ML model and find we can solve self-consistently the solid problem at the same level of accuracy. Fig. 1 shows the convergence of our ML method for a typical separation of the infinite chain with respect to the number of training points. The horizontal lines show two independent DMRG estimates of the energy.

II. BACKGROUND

A. DFT

The Hohenberg-Kohn theorem [10] establishes that the ground-state energy and density of a many-body problem may be found by minimizing a density functional:

$$E = \min_n \left\{ F[n] + \int d^3r\, n(\mathbf{r})\, v(\mathbf{r}) \right\}, \tag{1}$$

where $n(\mathbf{r})$ is the single-particle density, normalized to N particles, and $v(\mathbf{r})$ is the one-body potential. The functional F can be defined via a constrained search as [24]

$$F[n] = \min_{\Psi \to n} \langle \Psi | \hat{T} + \hat{V}_{\rm ee} | \Psi \rangle \tag{2}$$

where $\hat{T}$ is the kinetic energy operator and $V_{\rm ee}$ is the electron-electron repulsion operator, while Ψ is a normalized antisymmetric (for fermions) wavefunction. While this showed that the old Thomas-Fermi theory [8, 9, 25] was an approximation to an exact formulation, few modern calculations perform such a direct minimization. In practice, almost all calculations use the famous Kohn-Sham (KS) scheme, which uses an auxillary set of non-interacting orbitals in a single, multiplicative potential whose density is defined to match that of the original system, and in terms of which we can write

$$F[n] = T_{\rm S}[n] + U[n] + E_{\rm XC}[n], \tag{3}$$

where $T_{\rm S}$ is the non-interacting kinetic energy of the KS electrons, U is the Hartree self-repulsion, and $E_{\rm XC}$ is the exchange-correlation energy (defined by this equation).

The genius of the KS formulation is that $E_{\rm XC}$ is typically a small fraction of F, so that much higher accuracy can be achieved by approximating only this component. The cost of the KS scheme is formally N^3, the cost of solving for the orbitals. Much of modern DFT research is devoted to improving approximations to $E_{\rm XC}$, within which all quantum-many body effects are contained (by definition). The smaller field of pure DFT, also known as 'orbital-free', aspires to approximate $T_{\rm S}[n]$ directly, as in the old TF theory [8, 26], and thus bypass the need to solve the KS equations.

Many modern XC approximations are local or semilocal, i.e., use the density and its gradient to approximate the XC energy density at a point. While remarkably useful results can be obtained with such approximations, there remains a classic failure that can be understood in terms of the simple H_2 molecule [27]. Those approximations work well in the vicinity of the equilibrium bond length, but as the bond is stretched, they fail more and more badly. In the limit of a large but finite bond length (which we call stretched), a spin-restricted calculation yields the highly inaccurate energy of two unpolarized H atoms. On the other hand, an unrestricted calculation yields an accurate stretched energy, but has broken spin symmetry. Neither situation is satisfactory [4], most modern semi-local approximations fail in this way [27, 28] and efforts to overcome such difficulties are ongoing [29, 30]. An analogous failure occurs for semilocal approximations to $T_{\rm S}[n]$ when bonds are stretched in orbital-free DFT. Fig. 2 illustrates the failure of semilocal XC, by comparing the blue restricted LDA curve with the black DMRG curve. There is a huge error in the stretched limit.

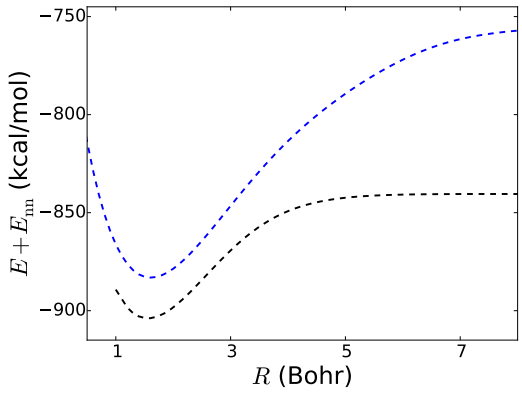

FIG. 2. (Color online) Binding curve for a 1d H_2 molecule. Black: highly accurate, converged DMRG results. Blue: LDA result restricted to a spin singlet [23].

B. DMRG benchmark data

It is difficult to overemphasize the utility of benchmark quantum chemical calculations for the development of DFT. The DFT revolution in quantum chemistry was made possible by the existence of the well-tested G2 data set for small molecules, and databases in quantum chemistry have proliferated ever since. On the other hand, calculations of 'quantum chemical' accuracy, i.e., errors below 1 kcal/mol, are much more difficult and rarer for solids. A recent heroic effort [31] was made for benzene, a molecular crystal.

For the present study, we need to consider chains of up to 20 H atoms, with many different values of the interatomic spacing ranging from about 1 to 10 Bohr. Extracting this large amount of data at the required level of accuracy from a quantum chemical code would be extremely demanding, if not impossible, given the strong correlation effects when the bonds are stretched.

Recently, DMRG has been applied to a one-dimensional analog of real-space Coulomb-interacing Hamiltonians, for precisely the purpose of performing demanding, highly accurate benchmark calculations of strongly correlated systems. In particular, the interaction is modeled as

$$v_{\rm ee}(u) = A\,\exp(-\kappa\,|u|) \tag{4}$$

where $A = 1.071$ and $\kappa^{-1} = 2.385$ [23] and u is the separation. This choice best mimics a popular soft Coulomb interaction, while having a single exponential allows DMRG to run very fast [23]. The one-body potential is then taken as $v(x) = -Zv_{\rm ee}(x)$, where Z is the 'charge' on a nucleus. Here $Z = 1$ for each H atom in the chain. This 1d analog allows rapid testing of novel ideas in electronic structure, especially those involving the bulk limit. Fig. 2 is in fact for 1d H_2 with these parameters, and illustrates that the failures of standard DFT approximations such as LDA mimic those of 3d Coulomb systems. The DMRG curve plateau is at twice the ground-state energy of one of these 1d H atoms.

C. Machine learning of the KS kinetic energy functional

ML is a set of algorithms developed to find hidden insights in data. It is widely used especially when the pattern behind complicated data is difficult to deduce explicitly. Successful applications include computer vision [32], cybersecurity [33], ancient abstract strategy games [34], etc.

Recently, in chemistry and materials science, machine-learning has become a popular tool for analyzing properties of molecules and materials, and finding specific functions from large data sets [35, 36]. But it has also been applied to the problem of finding density functionals, constructed by interpolation from accurate examples. To date, the focus has been on the KS kinetic energy, $T_{\rm S}[n]$, partially because of the ready availability of data (every cycle of every one of the 30,000 KS-DFT calculations each year [1] produces an accurate example of $T_{\rm S}[n]$) and because of the enormous potential for speeding up routine DFT calculations.

The ML algorithm we used for modeling $T_{\rm S}[n]$ is kernel ridge regression (KRR). It is a nonlinear regression method with an L2 regularization [37]. The density functional is represented as

$$T_{\rm S}^{\rm ML}[n] = \sum_{i=1}^{N_{\rm T}} \alpha_i k[n, n_i], \tag{5}$$

where $N_{\rm T}$ is the number of training data, $n_i(x)$ are the training data and $k[n, n_i]$ is a kernel, some measure of the "similarity" between densities. Throughout this work, we use a Gaussian kernel,

$$k[n, n'] = \exp(-\|n - n'\|^2/2\sigma^2), \tag{6}$$

where

$$\|n - n'\| = \int dx (n(x) - n'(x))^2. \tag{7}$$

Such a kernel is standard in KRR, and has yielded excellent results in previous studies of $T_{\rm S}[n]$ [14]. The weights $\alpha = (\alpha_1, \cdots, \alpha_{N_{\rm T}})$ are found by optimizing the cost function

$$\mathcal{C}(\alpha) = \sum_{i=1}^{N_{\rm T}} (T_{\rm S}^{\rm ML}[n_i] - T_{\rm S}[n_i])^2 + \lambda \alpha^{\rm T} \mathbf{K} \alpha \tag{8}$$

where $\mathbf{K}$ is the kernel matrix, $K_{ij} = k[n_i, n_j]$. The regularization strength λ and length scale σ are hyperparameters which are found via cross validation. A crucial principle in kernel ridge regression is the separation of the training data from the test data. A test set is constructed entirely independently from the training set. The cross-validation to find the hyperparameters occurs using only training data. The resulting approximate functional is tested only on the test data.

While highly accurate results for $T_{\rm S}[n]$ can be found with relatively little data [12], it was immediately realized that the corresponding functional derivative is highly *inaccurate*. This is unfortunate, as the practical usefulness of an accurate model for $T_{\rm S}[n]$ is in finding the density via solution of the Euler equation (for the KS system):

$$\frac{\delta T_{\rm S}}{\delta n(x)} = -v_{\rm S}(x), \tag{9}$$

where $v_{\rm S}(x)$ is the KS potential. This difficulty has been surmounted in a sequence of increasingly sophisticated methods [13–15], each of which constrains the density search to only the manifold of densities spanned by the data, avoiding searching in directions for which there is insufficient data to evaluate $T_{\rm S}$ accurately. With such techniques, it has been possible to demonstrate an ML $T_{\rm S}$ functional that correctly mimics the KS solution even as a bond stretches [13], something impossible for any local or semilocal approximation to $T_{\rm S}$. The value of this is to cut down the computational cost of large, repetitive KS calculations, but one still uses some standard XC approximation. Thus a machine-learned functional for $T_{\rm S}$ can, at best, reproduce the incorrect LDA curve of Fig. 2.

III. METHOD

In all applications in this work, we generate a large data set of highly accurate results generated using DMRG. We use a real-space grid with spacing 0.04, which has previously been shown to be sufficient to converge the results [23]. We calculate the energies and densities of chains of even numbers of atoms, from 2 to 20, with interatomic separations between 1 and 10 Bohr. Higher accuracy is achieved when every atom is centered on a grid

point, discretizing the set of allowed separations. The specific separations used are listed in the supplemental information.

Then a subset of these data are left out as test set. The training set, with $N_{\rm T}$ values of R, are collected from the remaining data. These are chosen to be as close to equally spaced as practical. The test set is excluded from the data where the training set is sampled from, to avoid contamination via the cross-validation process.

A. Machine-learned $F[n]$ for a given molecule

The first improvement on previous work is to apply ML to $F[n]$ itself, not $T_{\rm s}[n]$ as in earlier work [13]. All the equations of Sec. II C apply directly, by replacing $T_{\rm s}[n]$ with $F[n]$ and $v_{\rm s}(x)$ with $v(x)$. It is not *a priori* obvious that one might not encounter some difficulty, as $F[n]$ contains all the many-body physics of the ground state.

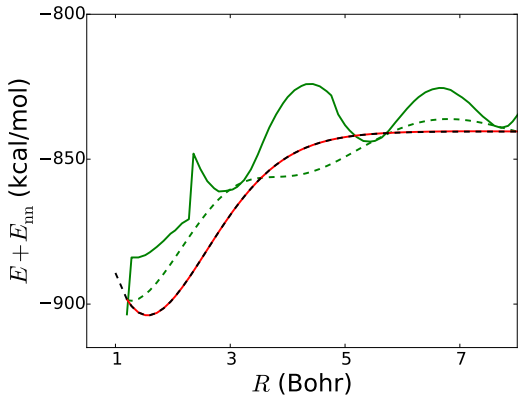

FIG. 3. (Color online) Same as Fig. 2. The green curves are ML with $N_{\rm T}$ = 5 on both the exact (dashed) and ML-optimized (solid) densities. The red solid curve is the ML with N_T=20 (number of training points) on ML-optimized (solid) densities. Black dashed curve is the exact DMRG curve, matching nearly exactly the N_T=20 on ML line.

We continue to use the H_2 molecule to illustrate our method. Contrary to previous work, we apply KRR algorithms to ML the interacting functional $F[n]$ itself, by training on highly accurate DMRG energies and densities at various values of R. In Table I, we list the errors for H_2 as a function of $N_{\rm T}$, both on the exact density and on an optimally constrained density found by the methods of Ref. [38].

To illustrate the procedure, in Fig. 3, we show the energies with only 5 training points, $R = 1.00, 3.20, 5.48, 7.76, 10.00$, yielding the smooth, green dashed curve, when evaluated on the exact densities. The curve (almost) exactly matches at the training points, but is noticeably inaccurate inbetween. But note that, in contrast to all previous studies, we are fitting the full $F[n]$, not just $T_{\rm s}[n]$, so that, e.g., our inaccurate curve dissociates H_2 correctly, while no standard DFT calculation, with a standard XC approximation, can.

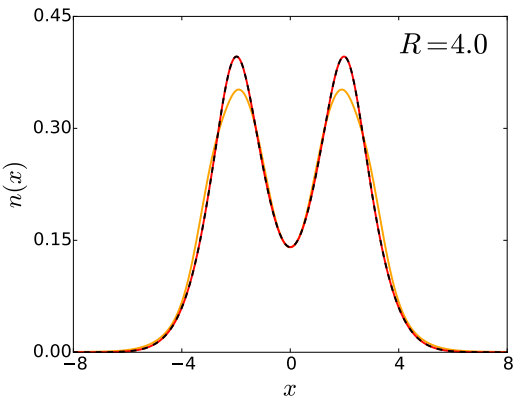

FIG. 4. (Color online) Optimal densities for 1d H_2 molecule in the test set: DMRG (black dashed), ML with N_T=5 (orange solid), ML with N_T=20 (red solid).

The problem is actually much greater than even the smooth dashed green curve would suggest. In practice, we not only need the energy functional, but also its derivative, at least in the vicinity of a solution density. This is because we use the functional to find the density itself, via the Euler equation

$$\frac{\delta F}{\delta n(x)} = -v(x). \tag{10}$$

In fact, the derivatives of ML functionals such as that of Eq (6) are highly inaccurate and cannot be used to find the minimizing density. Methods have been developed to constrain the search to the manifold of training data via non-linear gradient denoising (NLGD) [38]. For our H_2 with $N_{\rm T}$ = 5, these lead to the (even worse) solid green curve of Fig. 3. The optimal density is shown in Fig. 4. We clearly see that (a) the accuracy is not high enough and (b) the error is dominated by the error in the densities. (This is called a density-driven error [39] in a DFT calculation.)

However, when we increase to 20 data points, the ML curve (red solid) is indistinguishable from the exact one, and the error at equilibrium is only 0.007 kcal/mol, and shrinks with increasing R. This calculation applies all the principles discussed in Ref. 13, but is now applying them to the many-body problem, not just the KS problem. Even in the stretched limit, where the system is strongly correlated, there is no loss of accuracy. Note that we are not just fitting the binding curve, as we are reproducing the many-body density at every value of R, starting from data at a limited number of values. In Fig. 4, we plot the optimally-constrained densities at R = 4.0 (outside all training sets) for $N_{\rm T}$ = 5 and $N_{\rm T}$ = 20, compared with the exact density.

Here, ML has entirely bypassed the difficulty of solving the many-fermion problem. The machine learns the characteristics of the solution without ever solving the differential equation. Moreover, the HK theorem is a

statement of the minimal information needed to characterize the ground-state of the system. In some ways, this ML approach is the purest embodiment of the HK theorem.

We note that the Euler equation (10) is merely the same as in KS theory but with T_s replaced by F. Thus, the entire algorithm for learning is synonymous with the previous works [13] but with T_s replaced by $F[n]$.

B. Finding a data-driven optimal basis for longer chains

The cost of optimal gradient descent methods, evaluated on a spatial grid, grows very rapidly with the number of grid points, and rapidly becomes unfeasible as the number of H atoms grows. Thus a simpler representation of the density is required. To overcome those difficulties, we introduce a basis set. Inspired by the localized atomic bases used in most quantum chemical codes, we developed a data-driven basis set using Hirshfeld partitioning [40] and principal component analysis (PCA).

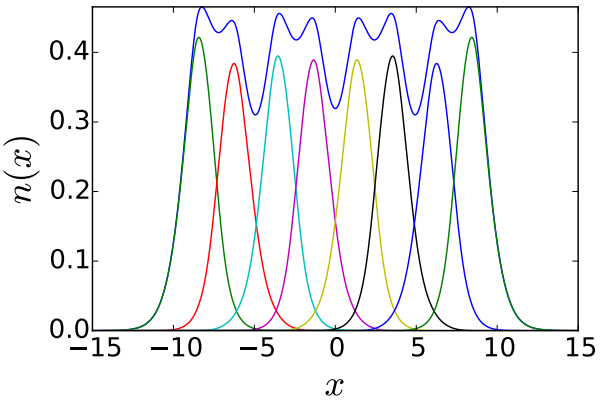

FIG. 5. Partition density of each H atom in H_8.

To partition a molecular density via the Hirshfeld scheme, begin with the protomolecule of overlapped atomic densities at the nuclear positions of the real molecule. If $n_i^0(x) = n_1^0(x-(i-1)R)$ is an isolated atomic density at the i-th nuclear center, spaced R apart, then

$$n^0(x) = \sum_{i=1}^{N} n_i^0(x) \tag{11}$$

is the density of the protomolecule, where R is the interatomic spacing. We define a weight

$$w_i(x) = n_i^0(x)/n^0(x), \tag{12}$$

associated with each atom, and then define the density of each Hirshfeld atom within the real molecule as

$$n_i(x) = w_i(x)n(x), \tag{13}$$

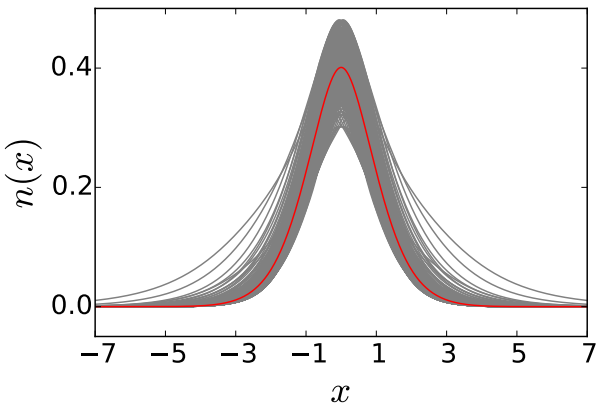

FIG. 6. Single H atom densities for H atoms in different chains and atomic distance (gray). The average density is plotted in red.

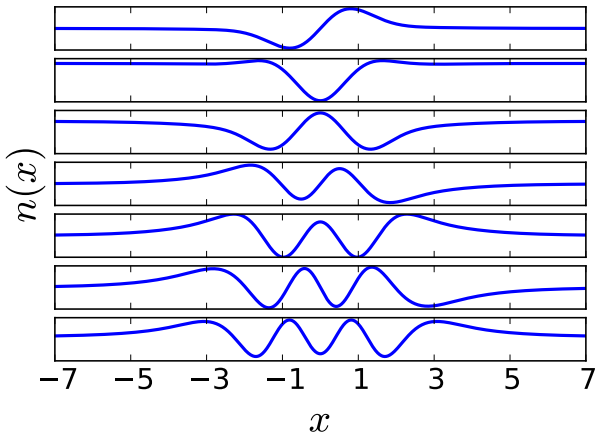

FIG. 7. First 7 principal components of the densities shown in Fig. 6, from top to bottom.

where $n(x)$ is the exact molecular density. The ground state density of a single hydrogen atom $n_i^0(x)$ is reported in Ref. 23. Fig. 5 shows partition densities $n_i(x)$ of atoms in one H_8.

Next, for a specific chain length N, we consider a range of interatomic separations R, and consider the collection of every atomic density within the chain for every value of R in a training set, each centered on the origin, as shown in Fig. 6. These individual atomic partition densities reflect the diverse behaviors caused by the interaction between the hydrogen atoms inside the chains. A principal component analysis is applied to these densities, and the eigenvalues are ordered in decreasing magnitude to find a subspace with the maximum variance. Each atomic density can be accurately represented by the base density $f_0(x)$ (red in Fig. 6) and 7 principal components (Fig.

7),

$$n_i(R,x) = f_0(x) + \sum_{p=1}^{7} c_{i,p}(R) f_p(x). \qquad (14)$$

Thus the total density of H_N with separations R is $\sum_i^N n_i(R,x)$, and is described by just $7N$ coefficients. Note that $f_0(x)$ is very close to an isolated atom density, but we use the average to center our data for the PCA analysis. Our representation greatly reduces the number of variables in the density representation for a given chain length, and saves a significant amount of computational cost when solving for the ground state density of the system. This new basis set is completely data-driven and physically meaningful. Further, the derivation of this method is not limited to 1d.

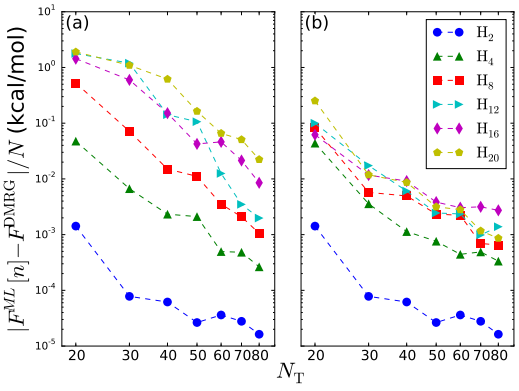

FIG. 8. (Color online) Learning curves for several 1d H chains. (a) ML using the total density. (b) ML using the bulk partition densities (see text).

We next repeated these calculations for a sequence of chains of increasing length. In each case, we train $F^{\mathrm{ML}}[n]$ on a limited training set, and then compare on a test set (see supplementary material), with the accurate results supplied by DMRG. The learning curves, i.e., error as a function of N_{T}, of chains of differing length, are shown in Fig. 10(a). The error typically decreases with amount of training data, but for fixed N_{T}, longer chains display substantially larger errors. This is because the binding energy curve changes more rapidly when the chain length is increased.

In applied machine learning, feature engineering, which uses domain knowledge of the data to improve the efficiency of ML algorithms, is a crucial step. Here, we know that as the chain length increases, the central density should converge to a fixed value (thermodynamic limit). We therefore choose the central two atomic densities alone to use as a minimal input feature for learning the energy of a given finite chain. The learning curves for models trained only on this central partition density are shown in Fig. 10(b). For chain lengths greater than or equal to 12, substantially greater accuracy is reached for a fixed amount of training data. Here we still use the total density for $N \leq 8$ and the bulk density for $N \geq 12$. The model performance and hyperparameters are presented in Table I.

C. Extrapolation to the thermodynamic limit

Our ultimate goal is to use ML to find the energy of the infinite chain to within chemical accuracy, for all interatomic separations. To do this, we first build a set of infinite chain energies and densities. For each value of R, we extrapolate both the density and energy of our finite chains as a function of N. This then gives us a set of data for the infinite chain that we can both train and test on and gave rise to Fig. 1.

In an entirely separate calculation, we also performed DMRG directly for the infinite chain, using the method of McCulloch [41] for a four atom unit cell [42]. The system is initialized by solving the equivalent finite size system with box edges at $R/2$. As a part of the iDMRG algorithm [41], a single unit cell is then inserted into the center of the finite system and 15 sweeps are performed over the inserted unit cell. The sequence is repeated–after adding another unit cell–until convergence. We compare these energies with the extrapolated values, finding agreement to within 1 kcal/mol for all values of R. This agreement validates our extrapolation procedure. We find that, with 50 data points, the ML result, on the optimized density, also agrees to within 1 kcal/mol. Thus, armed with the 50-data-point machine learned functional, one can self-consistently find the density and energy of the infinite chain to quantum chemical accuracy.

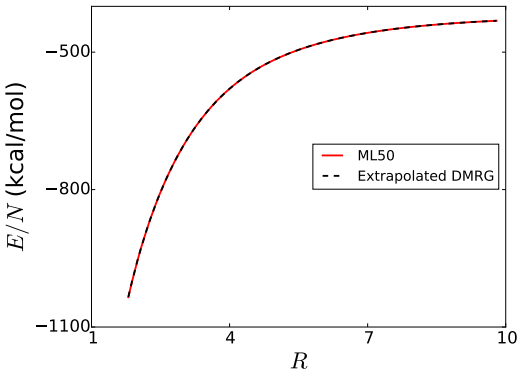

FIG. 9. (Color online) Electronic energy per atom in the thermodynamic limit, both via DMRG chains (extrapolated to infinity) and using machine learning with 50 data points per chain.

Our final figure simply demonstrates that the error for the infinite chain (and for all the ML calculations) is almost entirely due to the error in the optimized density. The functional-driven error [39] is the energy error made on the exact density:

$$\Delta E_F = E^{\mathrm{ML}}[n] - E[n] = F^{\mathrm{ML}}[n] - F[n]. \qquad (15)$$

We see that, at any level of training, ΔE_F is an order

N	$N_{\rm T}$	λ	σ	$\overline{\lvert\Delta E_{\rm F}\rvert}/N$	$\max\lvert\Delta E_{\rm F}\rvert/N$	$\overline{\lvert\Delta E\rvert}/N$	$\max\lvert\Delta E\rvert/N$	$E^{\rm ML}_{R=9.8}/N$	$E^{\rm DMRG}_{R=9.8}/N$
2	5	1.0×10^{-8}	1000	2.54	7.02	9.74	20.3	-421.291	-425.797
2	20	4.6×10^{-10}	2.15	0.00121	0.00802	0.005	0.013	-425.785	-425.797
2	50	1.0×10^{-12}	0.70	0.00003	0.00034	0.050	0.304	-425.798	-425.797
4	50	2.2×10^{-11}	46.4	0.0021	0.016	0.005	0.017	-428.617	-428.620
8	50	1.0×10^{-4}	2.15	0.011	0.31	0.28	1.68	-430.011	-430.032
12	50	1.0×10^{-12}	0.46	0.0031	0.010	0.24	0.88	-430.502	-430.503
16	50	2.2×10^{-11}	0.46	0.0042	0.012	0.08	0.41	-430.738	-430.738
20	50	2.2×10^{-11}	0.46	0.0042	0.014	0.26	0.88	-430.880	-430.880
∞	50	1.0×10^{-8}	0.46	0.012	0.050	0.073	0.27	-431.447	-431.444

TABLE I. ML performance on different chains H_N. N_T is the size of training set. Regularization strength λ and kernel length scale σ is the model hyperparameters selected by cross validation [14]. The functional driven error $\Delta E_F/N$ [39] is tested on the entire test set to show the overall accuracy. The total error $\Delta E/N$ is tested on the equilibrium test set to emphasize accuracy around equilibrium position. $E_{R=9.8}/N$ shows that ML can get very accurate dissociation limit. All errors are given in kcal/mol.

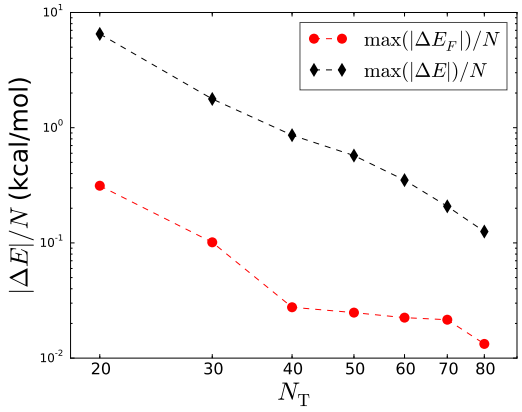

FIG. 10. (Color online) For a given training set with N_T training points, the functional driven error, ΔF_F per atom is shown in red (lower curve). The upper curve is the total energy error per atom evaluated self-consistently.

of magnitude smaller than the final energy error on the optimized density. Thus the error is density-driven but, nonetheless, can be forced down to quantum chemical limits with enough data.

IV. DISCUSSION

We have shown that it is in principle possible to construct, via machine learning, the entire interacting functional of Hohenberg and Kohn, $F[n]$, so accurately that optimized densities and energies evaluated on them are within quantum chemical accuracy. We have done this using the 1d simulation of continuum Hamiltonians established over the last several years, and using DMRG as an efficient solver. We apply the ML methods previously developed for approximating the non-interacting kinetic energy. Here, because we have precise energies for the interacting system to train on, we are able to construct the interacting functional, including all exchange and correlation effects. Our ML functional has no difficulties when bonds are stretched so that correlations become strong. We have even managed to apply this methodology to chains extrapolated to the thermodynamic limit, producing chemically accurate results for solids. This level of accuracy is far beyond that of any existing DFT calculation of a solid.

We conclude with a discussion of the steps needed to generalize this calculation to realistic solids. The first point is that, while we have performed the present calculations in 1d for both computational and programming efficiency, there is absolutely no reason they could not be repeated for real 3d hydrogen chains. These can readily be treated using DMRG [43, 44] and the ML algorithms are independent of the dimensionality. The extrapolation to an infinite chain limit should behave in a similar fashion. While the algorithm for generating a model of $F[n]$ was already developed, merely replacing $T_{\rm S}$ in the algorithm in Refs. 13 and 14 with $F[n]$, this is the first demonstration that this algorithm works for interacting electrons. The ground-state density is solved by the techniques of Ref. 38. However, instead of using a spatial grid, a data-driven atom-centered basis set for the density is developed. The distance metric, derivatives, and second derivatives are calculated on these basis functions. This greatly reduces the number of variables in the algorithm. For H_{20}, the calculation in this new basis can easily be performed on a personal laptop.

We note also that the (relatively) large amount of data needed to achieve chemical accuracy is solely because we have chosen to approximate the entire HK functional $F[n]$ and also need to find its derivative sufficiently accurately to produce an accurate energy. If, instead, we had used the KS scheme with a standard approximation for $E_{\rm XC}$, we could used ML simply for the error in that approximation, yielding inherently much more accurate densities, and requiring much less data for the same level of accuracy in the energy.

More generally, an accurate general purpose solver

such as QMC or accurate quantum chemical methods could be used to provide the highly-accurate data needed to train the machine learning. For 2d or 3d solids, extrapolation to the limit requires many more atoms. But since e.g., a 20x20x20 array is emminently practicable within KS-DFT, this is accessible with a machine-learned correction. Thus, at least within KS-DFT, there is no reason that an ML-constructed functional could not be created from QMC data to extract the bulk energy of a solid.

It has also recently been shown [45] that the amount of data needed to bypass the KS equations can be greatly reduced by learning the density as a functional of the KS potential, so that the functional derivative of the KS kinetic energy is never used. This was demonstrated for 3d molecules.

Lastly, we mention that the geometries used here are rather simple. We have not attempted to create ML-functionals that apply to many different atoms in many diverse bonding situations, as has been done in other work, and our functionals do not apply outside the domain they have been trained on. But since the energy curve of a bulk solid does not require such a functional, our ML approximation is sufficient for the purpose here.

Ulitmately, any ML method can be limited by the need for excessive training. But our work here shows that this is possible in principle, and there is no reason to think it more difficult in practice.

V. ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under award #DE-SC008696. T.E.B. also thanks the gracious support of the Pat Beckman Memorial Scholarship from the Orange County Chapter of the Achievement Rewards for College Scientists Foundation.

VI. SUPPLEMENTARY MATERIAL

A. Description of Data

The density matrix renormalization group (DMRG) [16–19] has become the gold standard for calculations in one dimension. The ansatz made for the wavefunction is that of a matrix product state (MPS). This ansatz allows for a site-by-site determination of the wavefunction by concentrating on a small number (in our implementation, two) lattice sites at a time. Once the wavefunction is updated on those two sites, the next two sites are treated. The entire system is swept back and forth until convergence which usually occurs very quickly in one dimension.

To evaluate the Hydrogen chains in this work, an extended Hubbard model [20, 21, 23],

$$\mathcal{H} = \sum_{j,\sigma} \frac{-1}{2a^2} \left(\hat{c}^\dagger_{j,\sigma} \hat{c}_{j+1,\sigma} + \text{h.c.} \right) - \tilde{\mu} n_{j\sigma} \tag{16}$$

$$+ \sum_j v^j n_j + \frac{1}{2} \sum_{ij} v^{ij}_{\text{ee}} n_i (n_j - \delta_{ij}), \tag{17}$$

can be constructed to recover the continuum limit in the limit of many sites. The prefactor on the kinetic energy terms is chosen to match the finite difference approximation for the kinetic energy with grid spacing a. An external potential is applied in the variable v^j while $\tilde{\mu} = \mu - \frac{1}{a^2}$ for chemical potential μ. Also, an electron-electron term, v^{ij}_{ee} is represented by an exponential function [23], This exponential mimicks the soft-Coulomb interaction, which itself is an approximation of the Coulomb interaction in 3d but spherically averaged [23]. The similarity between these functions gives the similar behaviors of the 1d atoms and their 3d counterparts when the symmetry is high.

Systems are calculated with open boundary conditions ("box" boundary conditions). The limit where the box boundary is far from the nearest atomic center is always taken, so no finite size effects appear.

A complication is apparent in 1d that does not appear in 3d. There is no angular momentum in 1d. Thus, not all neutral atoms bind their electrons. One can see this in a reduced example as follows: Consider a delta function interaction in 1d of the form $-\delta(x - R/2) - \delta(x + R/2)$ [46]. When $R = 0$, there is only one solution. At any finite R, the number of electrons that will bind increases from two. The same effect occurs for the exponential interaction, though it is not as easy to see.

This implies that a lower cutoff in the exponentially interaction hydrogen chains will impose a lower limit on suitable chain length. We are interested in systems that do bind all electrons, so a systems below a critical R are ignored. Table II lists the range of interatomic distances used for each chain. For each Hydrogen chain data generated by DMRG, first sample 80 data from the *entire test set range* in Table II equi-distantly. This test set is inaccessible in the training process. The rest of data in *training set range* in in Table II are used as grand training set, where the N_T training data are uniformly sampled to train the model. The *equilibrium test set range* is a subset of entire test set range, emphasizing the performance around equilibrium positions. The upper bound is around twice the equilibrium position given by DMRG result.

N	training set range	entire test set range	equilibrium test set range
2	$1.0 \le R \le 10$ (146)	$1.2 \le R \le 9.8$ (80)	$1.2 \le R \le 3.12$ (19)
4	$1.4 \le R \le 10$ (136)	$1.6 \le R \le 9.8$ (80)	$1.6 \le R \le 4.08$ (25)
8	$1.4 \le R \le 10$ (136)	$1.6 \le R \le 9.8$ (80)	$1.6 \le R \le 4.28$ (27)
12	$1.6 \le R \le 10$ (131)	$1.8 \le R \le 9.8$ (80)	$1.8 \le R \le 4.32$ (26)
16	$1.6 \le R \le 10$ (131)	$1.8 \le R \le 9.8$ (80)	$1.8 \le R \le 4.32$ (26)
20	$1.6 \le R \le 10$ (131)	$1.8 \le R \le 9.8$ (80)	$1.8 \le R \le 4.4$ (27)

TABLE II. Hydrogen chain data. N is the number of Hydrogen atoms in the chain. R is the atomic distance between atoms. The number of DMRG data in each range is in parenthese.

[1] A Pribram-Jones, DA Gross, and K Burke, "Dft: A theory full of holes?" Annual review of physical chemistry **66**, 283 (2015).

[2] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**, A1133–A1138 (1965).

[3] Lucas O. Wagner, Thomas E. Baker, M. Stoudenmire, E., Kieron Burke, and Steven R. White, "Kohn-sham calculations with the exact functional," Phys. Rev. B **90**, 045109 (2014).

[4] John P. Perdew, Andreas Savin, and Kieron Burke, "Escaping the symmetry dilemma through a pair-density interpretation of spin-density functional theory," Phys. Rev. A **51**, 4531–4541 (1995).

[5] AI Liechtenstein, VI Anisimov, and J Zaanen, "Density-functional theory and strong interactions: Orbital ordering in mott-hubbard insulators," Physical Review B **52**, R5467 (1995).

[6] Antoine Georges, Gabriel Kotliar, Werner Krauth, and Marcelo J. Rozenberg, "Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions," Rev. Mod. Phys. **68**, 13–125 (1996).

[7] Radu Iftimie, Peter Minary, and Mark E. Tuckerman, "Ab initio molecular dynamics: Concepts, recent developments, and future trends," Proceedings of the National Academy of Sciences of the United States of America **102**, 6654–6659 (2005).

[8] L. H. Thomas, "The calculation of atomic fields," Math.

Proc. Camb. Phil. Soc. **23**, 542–548 (1927).
[9] E. Fermi, Rend. Acc. Naz. Lincei **6** (1927).
[10] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev. **136**, B864–B871 (1964).
[11] F. Tran and T.A. Wesolowski, "Introduction of the explicit long-range nonlocality as an alternative to the gradient expansion approximation for the kinetic-energy functional," Chem. Phys. Lett. **360**, 209 (2002).
[12] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke, "Finding density functionals with machine learning," Phys. Rev. Lett. **108**, 253002 (2012).
[13] John C. Snyder, Matthias Rupp, Katja Hansen, Leo Blooston, Klaus-Robert Müller, and Kieron Burke, "Orbital-free bond breaking via machine learning," J. Chem. Phys. **139**, 224104 (2013).
[14] Li Li, John C. Snyder, Isabelle M Pelaschier, Jessica Huang, Uma-Naresh Niranjan, Paul Duncan, Matthias Rupp, Klaus-Robert Müller, and Kieron Burke, "Understanding machine-learned density functionals," International Journal of Quantum Chemistry **116**, 819–833 (2016).
[15] Kevin Vu, John C. Snyder, Li Li, Matthias Rupp, Brandon F. Chen, Tarek Khelif, Klaus-Robert Mller, and Kieron Burke, "Understanding kernel ridge regression: Common behaviors from simple functions to density functionals," International Journal of Quantum Chemistry **115**, 1115–1128 (2015).
[16] F. Sols, "Gauge-invariant formulation of electron linear transport," Phys. Rev. Lett. **67**, 2874 (1991).
[17] Steven R. White, "Density matrix formulation for quantum renormalization groups," Phys. Rev. Lett. **69**, 2863–2866 (1992).
[18] Ulrich Schollwöck, "The density-matrix renormalization group," Reviews of modern physics **77**, 259 (2005).
[19] Ulrich Schollwöck, "The density-matrix renormalization group in the age of matrix product states," Annals of Physics **326**, 96–192 (2011).
[20] E. M. Stoudenmire, Lucas O. Wagner, Steven R. White, and Kieron Burke, "One-dimensional continuum electronic structure with the density-matrix renormalization group and its implications for density-functional theory," Phys. Rev. Lett. **109**, 056402 (2012).
[21] Lucas O. Wagner, E.M. Stoudenmire, Kieron Burke, and Steven R. White, "Reference electronic structure calculations in one dimension," Phys. Chem. Chem. Phys. **14**, 8581 – 8590 (2012).
[22] Lucas O. Wagner, E. M. Stoudenmire, Kieron Burke, and Steven R. White, "Guaranteed convergence of the kohn-sham equations," Phys. Rev. Lett. **111**, 093003 (2013).
[23] Thomas E. Baker, E. Miles Stoudenmire, Lucas O. Wagner, Kieron Burke, and Steven R. White, "One-dimensional mimicking of electronic structure: The case for exponentials," Phys. Rev. B **91**, 235141 (2015).
[24] Mel Levy, "Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem," Proceedings of the National Academy of Sciences of the United States of America **76**, 6062–6065 (1979).
[25] Elliott H Lieb and Barry Simon, "The Thomas-Fermi theory of atoms, molecules and solids," Advances in Mathematics **23**, 22 – 116 (1977).
[26] E. Fermi, "Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente (a statistical method for the determination of some atomic properties and the application of this method to the theory of the periodic system of elements)," Zeitschrift für Physik A Hadrons and Nuclei **48**, 73–79 (1928).
[27] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang, "Insights into current limitations of density functional theory," Science **321**, 792–794 (2008).
[28] Paula Mori-Sánchez, Aron J Cohen, and Weitao Yang, "Localization and delocalization errors in density functional theory and implications for band-gap prediction," Physical review letters **100**, 146401 (2008).
[29] Jannis Erhard, Patrick Bleiziffer, and Andreas Görling, "Power series approximation for the correlation kernel leading to kohn-sham methods combining accuracy, computational efficiency, and general applicability," Physical Review Letters **117**, 143002 (2016).
[30] Tianyu Zhu, Piotr de Silva, Helen van Aggelen, and Troy Van Voorhis, "Many-electron expansion: A density functional hierarchy for strongly correlated systems," Physical Review B **93**, 201108 (2016).
[31] Jun Yang, Weifeng Hu, Denis Usvyat, Devin Matthews, Martin Schütz, and Garnet Kin-Lic Chan, "Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy," Science **345**, 640–643 (2014).
[32] O. M. Parkhi, A. Vedaldi, C. V. Jawahar, and A. Zisserman, "The truth about cats and dogs," in *IEEE International Conference on Computer Vision* (2011).
[33] X. Zhao, J. Luan, and M. Wolff, "Evaluating randomness in cyber attack textual artifacts," in *2016 APWG Symposium on Electronic Crime Research (eCrime)* (2016) pp. 1–5.
[34] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, *et al.*, "Mastering the game of go with deep neural networks and tree search," Nature **529**, 484–489 (2016).
[35] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O.A. von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," Phys. Rev. Lett. **108**, 058301 (2012).
[36] Matthias Rupp, "Special issue on machine learning and quantum mechanics," International Journal of Quantum Chemistry **115**, 1003–1004 (2015).
[37] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin, "The elements of statistical learning: data mining, inference and prediction," The Mathematical Intelligencer **27**, 83–85 (2005).
[38] John C. Snyder, Matthias Rupp, Klaus-Robert Mller, and Kieron Burke, "Nonlinear gradient denoising: Finding accurate extrema from inaccurate functional derivatives," International Journal of Quantum Chemistry **115**, 1102–1114 (2015).
[39] Min-Cheol Kim, Eunji Sim, and Kieron Burke, "Understanding and reducing errors in density functional calculations," Phys. Rev. Lett. **111**, 073003 (2013).
[40] F.L. Hirshfeld, "Bonded-atom fragments for describing molecular charge densities," Theoretica chimica acta **44**, 129–138 (1977).
[41] Ian P McCulloch, "Infinite size density matrix renormalization group, revisited," arXiv preprint arXiv:0804.2509

(2008).
[42] "Calculations were performed using the itensor library: http://itensor.org/," .
[43] Steven R White and Richard L Martin, "Ab initio quantum chemistry using the density matrix renormalization group," The Journal of chemical physics **110**, 4127–4130 (1999).
[44] Garnet Kin-Lic Chan and Martin Head-Gordon, "Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group," The Journal of chemical physics **116**, 4462–4476 (2002).
[45] Felix Brockherde, Li Li, Kieron Burke, and Klaus-Robert Müller, "By-passing the kohn-sham equations with machine learning," arXiv preprint arXiv:1609.02815 (2016).
[46] R.J. Magyar and K. Burke, "A density functional approach to one-dimensional interacting fermions," Phys. Rev. A **70**, 032508 (2004).