LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-828938

Accurate parameterization of the kinetic energy functional

S. Kumar, E. Landinez, B. Sadigh, S. Zhu, S. Hamel, B. Gallagher, V. Bulatov, J. Klepeis, A. Samanta

November 5, 2021

Journal of Chemical Physics

Disclaimer

Accurate parameterization of the kinetic energy functional

Shashikant Kumar,[1,2] Edgar Landinez Borda,[3] Babak Sadigh,[1] Siya Zhu,[1,4] Sebastian Hamel,[1] Brian Gallagher,[5] Vasily Bulatov,[6] John Klepeis,[1] and Amit Samanta[1,a)]

[1)]*Physics Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

[2)]*College of Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

[3)]*Chemistry Department, Brown University, Providence, Rhode Island 02912, USA*

[4)]*School of Engineering, Brown University, Providence, RI 02912, USA*

[5)]*Applications, Simulations and Quality Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

[6)]*Materials Science Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

(Dated: 8 November 2021)

The absence of a reliable formulation of kinetic energy density functional has hindered the development of orbital free density functional theory. Using the data-aided learning paradigm, we propose a simple prescription to accurately model the kinetic energy density of any system. Our method relies on a dictionary of functional forms for local and nonlocal contributions which have been proposed in the literature and the appropriate coefficients are calculated via a linear regression framework. To model the nonlocal contributions, we explore two new nonlocal functionals - a functional that captures fluctuations in electronic density and a functional that incorporates gradient information. Since, the analytical functional forms of the kernels present in these nonlocal terms are not known from theory, we propose a basis function expansion to model these seemingly difficult nonlocal quantities. This allows us to easily reconstruct kernels for any system using only a few structures. The proposed method is able to learn kinetic energy densities and total kinetic energies of molecular and periodic systems, such as H_2, LiH, LiF and a one-dimensional chain of 8 hydrogens using data from Kohn-Sham density functional theory calculations for only a few structures. For the ease of reproduction, codes used to generate the models are provided in the supporting materials.

I. INTRODUCTION

Kohn-Sham density functional theory (KSDFT) is a powerful technique that is commonly used to calculate the electronic structure of a wide variety of materials to predict, and/or to analyze their mechanical, optical, electronic, or magnetic properties. The relevance of KSDFT can be judged from the fact that tens of thousands of papers based on electronic structure calculations are published every year, but in spite of its widespread use, system sizes accessible in routine KSDFT are typically limited to a few hundred[1,2] or a few thousand atoms[3]. Orbital-free density functional theory (OFDFT) can potentially be used to simulate electronic structure of large, realistic systems which are currently not possible[4–11]. However, OFDFT calculations are often not as reliable as KSDFT and fundamental bottlenecks that have impeded the development of OFDFT (such as inability to capture bond formation) have been extensively discussed in the literature[9,12–17] and the foremost among them is the absence of an accurate kinetic energy functional. Nearly thirty years ago, in a seminal paper Wang and Teter[16] succinctly stated that "An accurate and computationally inexpensive kinetic-energy density functional is highly desirable in the simulation of condensed-matter systems, because, in that environment, the solution of the Kohn-Sham equation for a large number of bands and k-points can be reduced to the calculation of a single function by using the kinetic energy functional." Since then, many groups have proposed kinetic energy functionals and local pseudo-potentials that can be used to run orbital-free DFT calculations for large system sizes. However, in spite of these developments, the popularity and the applicability of OFDFT remains far behind that of KSDFT.

A typical kinetic energy functional used in OFDFT can be written as a sum of semilocal and nonlocal contributions, such as

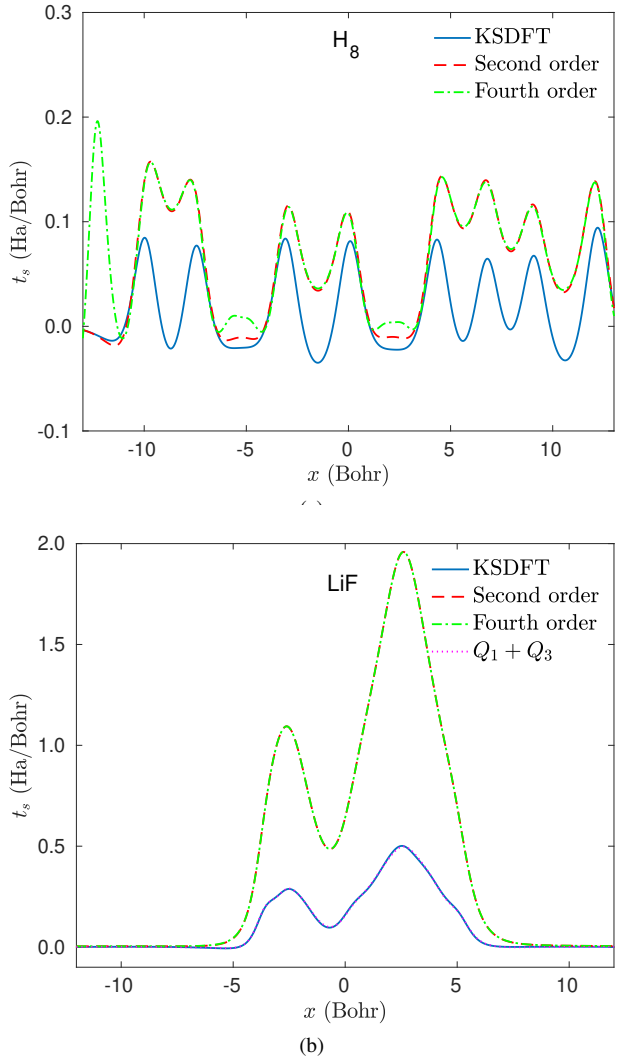
$$T_{\mathrm{KE}}[\rho(\mathbf{r})] = \int \Big(t_{\mathrm{loc}}(\mathbf{r}) + t_{\mathrm{nl}}(\mathbf{r}) \Big) d\mathbf{r}, \tag{1}$$

where the semilocal contributions (denoted by $t_{\mathrm{loc}}[\rho(\mathbf{r})]$) depend on the local electron density and its derivatives and the explicit functional form for non-interacting Fermions in one-dimension, derived by Samaj and Percus[18] is given by

$$\begin{aligned} \pi t_{\mathrm{loc}}(x) = {} & \frac{1}{6}[\pi\rho(x)]^3 - \frac{\pi[\rho'(x)]^2}{24\rho(x)} - \frac{\pi\rho''(x)}{12} \\ & - \frac{\rho^{(4)}(x)}{360\pi[\rho(x)]^2} + \frac{4[\rho''(x)]^2 + 7\rho'(x)\rho'''(x)}{360\pi[\rho(x)]^3}. \end{aligned} \tag{2}$$

In the above equation, the first and second terms on right hand side are the Thomas-Fermi and von Weizsäcker terms. Figure 1 shows the calculated kinetic energy densities from KSDFT for a few one-dimensional systems and those predicted by using Eq. 2.We add a regularization constant of magnitude 10^{-3} to the denominator of the last two terms of Eq. 2 to prevent divergence in the low density regime. The unphysical behavior seen in Fig. 1 in the low density limit is related to the behavior of the last two terms in Eq. 2: If density decays as $\rho(x) \sim \exp(-\lambda x)$ in the low density limit, then $[\rho''(x)]^2/[\rho(x)]^3 \sim 1/\rho(x)$ which diverges as $\rho(x) \to 0$. In

[a)]Electronic mail: samanta1@llnl.gov



(b)

FIG. 1. The performance of kinetic energy functional model developed for non-interacting Fermions by Samaj et al.[18] for two representative one-dimensional systems - a chain of eight hydrogens (shown in (a)) and a LiF molecule (shown in (b)) are compared with kinetic energy density calculated from a one-dimensional KSDFT code. Predictions from models containing only the first three terms on the right hand side of Eq 2 (marked "second order") are very close to predictions from models containing all the terms shown in Eq 2 (marked "fourth order"). This suggests that contributions from higher order corrections are negligible.

all of these systems, the model predictions are able to capture the trends in KSDFT kinetic energies, but the amplitudes are off. Thus, semilocal contributions alone are unable to capture the kinetic energy of these relatively simple one-dimensional systems.

The nonlocal term in Eq. 1 is typically written in literature[13,16,19] as

$$t_{\rm nl}^{(1)}(\mathbf{r}) = \rho^m(\mathbf{r}) \int \rho^n(\mathbf{r}') \, Q\left(\left|\mathbf{r}-\mathbf{r}'\right|, \mathcal{V}\right) d\mathbf{r}'. \qquad (3)$$

Here $\mathcal{V}$ is a function of electron density, m and n are exponents and taken to be m=n=1 and 5/6 in Chacón et al.[13] and Wang et al.[16], respectively. Different authors have used different forms of $\mathcal{V}$: In some cases, $\mathcal{V}$ depends only on $\rho(\mathbf{r})$[13], while in other cases it depends on both $\rho(\mathbf{r})$ and $\rho(\mathbf{r}')$[16]. This nonlocal term can be traced back to Hohenberg and Kohn[20] who showed that the response function is related to the second functional derivative of the kinetic energy, i.e. $K(\mathbf{r},\mathbf{r}') = \delta^2 T_{\rm KE}/\delta\rho(\mathbf{r})\,\delta\rho(\mathbf{r}')$ which for a homogeneous, non-interacting system (with density ρ_0) is related to the Lindhard susceptibility in the Fourier space[13,21] via $\hat{K}(k,\rho_0) = -1/\chi_0(\eta)$ where $\eta = k/2k_{\rm F}(\rho_0)$, $k_{\rm F}(\rho_0)$ is the Thomas-Fermi momentum and χ_0 is the Lindhard susceptibility, which for a one-dimensional system is given by

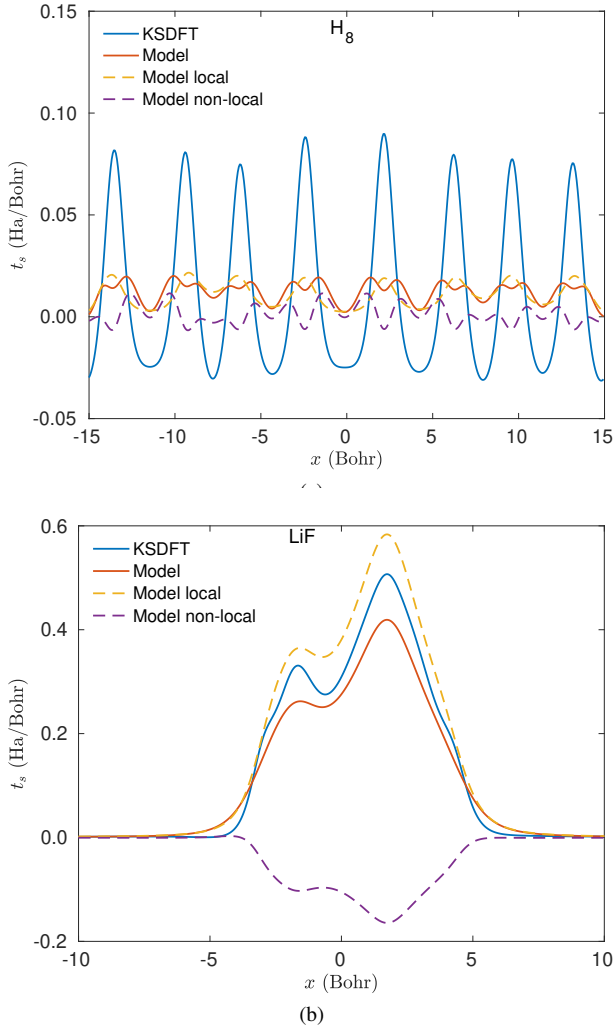
$$\chi_0(\eta) = -\frac{1}{\pi k_{\rm F}} \frac{1}{\eta} \log\left|\frac{1+\eta}{1-\eta}\right|. \qquad (4)$$

To gauge the importance of the nonlocal correction to the kinetic energy functional, in Fig. 2 we show results of one such 1D model proposed by Combariza et al.[22] where the expression of kinetic energy is written as

$$T[\rho(x)] = T_{\rm vW}[\rho(x)] - \frac{1}{3}T_{\rm TF}[\rho(x)] + \frac{4}{3}\int dx \rho(x) t(\bar{\rho}(x)),$$
$$\bar{\rho}(x) = \int dx' \rho(x') w(|x-x'|, \rho(x)). \qquad (5)$$

Here $T_{\rm vW}$ and $T_{\rm TF}$ are the von Weizsäcker and Thomas-Fermi terms, and the last integral on the right hand side is the nonlocal term. The weight function, $w(|x-x'|,\rho(x))$, is usually determined by imposing the Lindhard response on the kinetic energy. Figure 2 shows the differences between KSDFT kinetic energy densities and model predictions (using ground state electron densities obtained from KSDFT calculations) for four very different systems - H_2, LiH and LiF molecules and a one-dimensional chain of eight hydrogens, H_8. The total kinetic energies obtained from one-dimensional numerical integration (using Euler method) of the kinetic energy densities for H_2, LiH, LiF and H_8 systems obtained from the model proposed by Combariza et al. are 0.17, 0.37, 0.68, 2.41 Ha, respectively, and those obtained from KSDFT are 0.10, 0.36, 0.50, 2.70 Ha, respectively. As shown in Fig. 2, the nonlocal contributions significantly improve the predictions and errors in predicted total kinetic energies are small, but disparities between KSDFT kinetic energy density profiles and the model predictions suggest that there is room for improvement.

In a way, the nonlocal term $t_{\rm nl}^{(1)}$ can be derived from a generalization of the Thomas-Fermi model.[16] On the other hand, Garcia-Aldea and Alverellos[23,24] have proposed nonlocal models based on the von-Weizsacker term that capture spatial correlations between $\rho^{1/2}(\mathbf{r})$ and its derivatives. Even though the authors focused on elemental systems, their analysis suggests that these models are able to accurately model kinetic energy densities at intermediate electron densities. Similarly, many other researchers have explore meta-GGA based kinetic energy functionals[25–30] by systematically analyzing the behavior at low and high electron densities and $|\nabla\rho|/\rho^{4/3}$ and many such models have been compared in Ref. [26, 31, and 32].[4,5]



(b)

FIG. 2. The performance of the kinetic energy functional model developed for non-interacting fermions by Combariza et al.[22] for various one dimensional systems is compared with kinetic energy densities calculated from a one-dimensional KSDFT code. Subplots a, and b correspond to H_8 and LiF systems. The distance between atoms in both of these systems is about 3.5 Bohr.

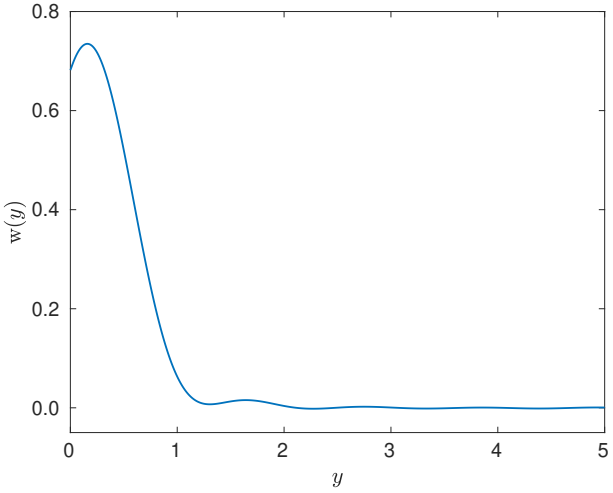


FIG. 3. nonlocal kernel of Eq. 5 proposed by Combariza et al.[22]. The abscissa x is dimensionless variable given by $y = \pi\rho|x-x'|$. The function $w(y)$ is large when density is low and/or $|x-x'|$ is small, and $w(y) \to 0$ as y becomes large.

With the development of machine learning techniques and in the absence of a proper theoretical framework to capture the nonlocal contributions, many researchers have explored the possibility of using neural networks or other nonlinear regression techniques to model the kinetic energy functional. In a seminal paper, Snyder et al.[33] were able to obtain accurate kinetic energy density models for one-dimensional systems using a kernel ridge regression based framework. However, the performance deteriorated when such models were used in conjunction with an orbital free DFT framework[34]. Recently, Meyer et al.[35] have shown that better kinetic energy models can be generated by including functional derivatives in the fitting. Ghasemi and Kühne[36] have also used a neural network model based on functional derivative of the kinetic energy to train reliable models for systems containing up to 4 electrons. However, major bottlenecks in using these methods for realistic three-dimensional systems are the requirements of large data sets for training, and that model predictions are reliable only when they are used to interpolate between data points already present in training sets. Thus, it is important to design better kinetic energy density models that need only a few structures to train. To this end, designing models by manipulating existing physics-based models can be a promising route.

Recently, Golub and Manzhos[37] have demonstrated that using linear regression to fit a model containing up to fourth order derivatives of the density results in better reproduction of the kinetic energy density[37]. This is in agreement with conclusions presented by Seino et al.[38] that the performance of kinetic energy models can be improved by incorporating higher order derivatives of the density. However, Golub and Manzhos were also able to obtain better models using neural networks containing up to 12 hidden layers and using Gaussian process regression[37,39]. The fact that a neural network model with multiple hidden layers increases the accuracy of the model suggests that spatial correlations of the ground state electron density and its derivatives as well as their nonlinear combinations play an important role in the kinetic energy functional. This is contrast to traditional models that only include correlations in the density to capture the nonlocal part of the kinetic energy density. Thus, it is natural to ask if it is possible to design nonlocal terms containing derivatives of the density in line with the models proposed by Garcia-Aldea and Alverellos.

In this contribution, we follow the data-aided learning ideology and propose a simple prescription that allows us to calculate the local and nonlocal contributions to the kinetic energy and model the total kinetic energy of any system. Our method relies on a dictionary of functional forms that have been proposed in the literature to capture the local and nonlocal contributions, and a basis function expansion to capture the seemingly difficult, and analytically unknown, nonlocal kernel. To illustrate our method, we first use a one-dimensional

KSDFT framework to screen a few kinetic energy functionals. To this end, we use chemically homogeneous and heterogeneous systems, like H_2, LiF, F_2, and a one-dimensional chain of eight hydrogens, that have been used in the literature to analyze different electronic structure theories[39–41]. Next, we use three-dimensional (3D) calculations of simple molecules to assess the predictive capabilities of the model identified from the 1D analysis. We show that by using kinetic energy densities and the corresponding electronic densities of only a few (between 2-5) structures, it is possible to accurately reconstruct nonlocal contributions to the kinetic energy. In addition, our results show that instead of sampling the whole configuration space of Kohn-Sham kinetic energy and electron densities, it is easier to develop models that can learn a small part of the configuration space pertaining to the system of interest.

II. KINETIC ENERGY FUNCTIONAL MODELS

A. Semi-local contributions

The semi-local contributions in our approximation are based on the gradient expansion of the kinetic energy. We consider all five terms in the fourth order gradient expansion[18] for one-dimensional system. The first three terms are the Thomas-Fermi, von Weizsäcker and the Laplacian terms and the last two terms are the fourth order terms in the low density region as discussed in the introduction section and shown in Fig. 1. A regularization constant, λ, is thus used in the denominator of the last two terms to avoid the aforementioned divergence issue (see Eq. 6). The Thomas-Fermi term reproduces the exact kinetic energy for a uniform, non-interacting system. On the other hand, the von Weizsäcker gradient term reproduces the exact kinetic energy for a system with a single orbital. Therefore, to model the kinetic energy of an interacting system, we consider a dictionary of terms that include the fourth-order gradient expansion terms of kinetic energy of the following form::

$$\begin{aligned} t_{\text{local}}(x) =& c_0[\rho(x)]^3 + c_1\frac{[\rho'(x)]^2}{\rho(x)} + c_2\rho''(x) + \\ & c_3\frac{\rho^{(4)}(x)}{[\rho(x)]^2+\lambda} + c_4\frac{[\rho''(x)]^2+\rho'(x)\rho^{(3)}(x)}{[\rho(x)]^3+\lambda}, \end{aligned} \tag{6}$$

where $t_{\text{local}}(\mathbf{r})$ is a KE density model, $c_j's$ are the coefficients to be determined by fitting against the data and λ is the regularization constant. Our results show that different powers of the von Weizsäcker and the Laplacian of the density are not very effective in capturing the local contributions to the kinetic energy density. This is also evident from the profiles shown in Fig. 1: Inclusion of the higher order terms leads to only minor changes in the predicted total kinetic energy densities. In addition, many researchers have neglected the Laplacian of the density, but our results show that it affects the local behavior of our models. Our model differs from other models reported in the literature[9,13–16,42–44] because the coefficients associated with each term in Eq. 6 are obtained by fitting.

B. Nonlocal contributions

It is known that nonlocal contributions play an important role in the kinetic energy functional for inhomogeneous systems[45]. Therefore, generating nonlocal functionals to describe the kinetic energy density is an active topic of research and many models - majority of which are based on models proposed by Wang and Teter,[16] Chacón et al.,[13] or García-González et al.[22] - have been proposed in the past few decades. For example, in 2018, Mi et al.[9] proposed a nonlocal kinetic energy model that includes the model proposed in Ref. [13] and a correction term obtained via functional integration. On the other hand, nonlocal kinetic energy models proposed by Emily Carter and co-workers in Ref. [14 and 15] are based on solving an ordinary differential equation in the Fourier space. In this work, we propose to learn the nonlocal kernel for an inhomogeneous system directly from KSDFT data. Thus, to model the kinetic energy functional for one-dimensional systems, we consider the following form

$$\begin{aligned} t_{\text{nl}}^{(1)}(x) &= \int [\rho(x)]^{\alpha} Q_1\left(|x-x'|, w(x,x')\right)[\rho(x')]^{\beta}dx', \\ w(x,x') &= \rho(x)+\rho(x), \end{aligned} \tag{7}$$

where $t_{\text{nl}}^{(1)}[\rho](x)$ is an approximation for nonlocal kinetic energy density and α, β are the exponents. We set both α and β to 2 so that $t_{\text{nl}}^{(1)}$ is symmetric with respect to x and x' and satisfies the uniform coordinate scaling relation[46].

C. A nonlocal kernel based on Herring's formulation

In a seminal paper on learning the kinetic energy functional, Herring[19] in 1986 suggested that for systems with spin degrees of freedom, the relative contributions of different eigenfunctions of the Hamiltonian to the ground state electron density can be captured by $u_i(\mathbf{r}) = \psi_i(\mathbf{r})/\sqrt{\rho(\mathbf{r})}$ and correspondingly the kinetic energy functional can be written as[19]

$$t(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} + \sum_i \rho(\mathbf{r})|\nabla u_i(\mathbf{r})|^2 \tag{8}$$

Therefore, Herring proposed that the total energy is given by[19]

$$E = T_{\text{KE}}[\rho(\mathbf{r})] + \int \rho(\mathbf{r})\, v(\mathbf{r})\, d\mathbf{r} \tag{9}$$

and the kinetic energy is given by the sum of the von Weizsäcker term and contributions from the phase factor (the last term in Eq. 8), i.e. $T_{\text{KE}} = T_{\text{vW}} + T_\theta$. At the ground state, $v(\mathbf{r}) = -\delta T/\delta\rho(\mathbf{r})$ and the potential $v(\mathbf{r})$ is given by $v(\mathbf{r}) = v_{vW}(\mathbf{r}) + v_\theta(\mathbf{r})$. Here $v_{vW}(\mathbf{r}) = -\delta T_{\text{vW}}/\delta\rho(\mathbf{r})$ depends on first and second derivative of the electronic density and $v_\theta(\mathbf{r}) = -\delta T_\theta/\delta\rho(\mathbf{r})$. Based on his analysis of an almost uniform electron gas, Herring suggested the following

functional form[19] for v_θ:

$$v_\theta(\mathbf{r}) = v^{(0)} + Av^{(0)}\frac{\rho(\mathbf{r})-\rho^{(0)}}{\rho^{(0)}} + v^{(0)}\int Q_2\left(\left|\mathbf{r}-\mathbf{r}'\right|,\left[\rho(\mathbf{r})+\rho(\mathbf{r}')\right]\right)\left(\rho(\mathbf{r})-\rho^{(0)}\right)d\mathbf{r}' + \cdots \tag{10}$$

Here $\rho^{(0)}$ is the uniform electronic density of the electron gas, $\rho(\mathbf{r})$ is the perturbed electronic density, similarly $v_\theta(\mathbf{r})-v^{(0)}$ is the deviation in the potential from that of an uniform Fermi gas and Q_2 is a density dependent nonlocal kernel. The nonlocal term in Eq. 10 is different from the functional form used in Eq. 3 because it convolves the perturbation in the ground state electron density with the nonlocal, density-dependent kernel Q_2. Motivated by this discussion we consider a second nonlocal kernel to model the kinetic energy density:

$$t_{\mathrm{nl}}^{(2)}(x) = [\rho(x)]^2\int Q_2\left(\left|x-x'\right|,w(x,x')\right)\left[\rho(x)-\rho\left(x'\right)\right]^2dx'. \tag{11}$$

The exponents of $\rho(x)$ and $(\rho(x)-\rho(x')$ are set to 2 to ensure that this nonlocal term satisfies the coordinate scaling criterion. Before concluding this section, we note a few salient differences between the nonlocal terms proposed in Eqs. 3 and 11. Since, the ground state electron density is a positive quantity, the kinetic energy contribution from Eq. 3 is negative only when the density-dependent kernel Q_1 is negative. On the other hand, kinetic energy contribution from Eq. 11 is negative when the difference in electronic densities at x and x' is negative or when the kernel Q_2 is negative. In addition, using a Taylor series expansion we see that

$$\rho\left(x'\right)-\rho(x) = \nabla_x\rho(x)\cdot\left[x'-x\right]+\cdots \tag{12}$$

This means that the kernel Q_2 automatically includes correlations between derivatives of the ground state electron density.

D. A kernel to capture spatial correlations between derivatives

In a series of papers, N. H. March and co-authors (including P. M. Kozlowski, A. Holas and W. H. Young)[47–52] explored the nonlocal dependence of the kinetic energy functional for various model systems and proposed a kinetic energy density model for a one-dimensional harmonic oscillator in which the nonlocal term contains the derivative of the density instead of the form shown in Eq. 3. With this discussion in mind, let us consider the expression for kinetic energy given by Liu and Parr in Ref. [53]

$$\begin{aligned} T_{\mathrm{KE}}[\rho(\mathbf{r})] = &\int\frac{1}{8}\frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})}d\mathbf{r} \\ &+\frac{1}{2}\int\rho(\mathbf{r})\nabla_{\mathbf{r}}\cdot\nabla_{\mathbf{r}'}P\left(\mathbf{r},\mathbf{r}'\right)|_{\mathbf{r}=\mathbf{r}'}d\mathbf{r} \\ &+\frac{1}{2}\int\nabla_{\mathbf{r}}\rho(\mathbf{r})\cdot\nabla_{\mathbf{r}'}P\left(\mathbf{r},\mathbf{r}'\right)|_{\mathbf{r}=\mathbf{r}'}d\mathbf{r}, \end{aligned} \tag{13}$$

where, $P^2(\mathbf{r},\mathbf{r}') = |\gamma(\mathbf{r},\mathbf{r}')|^2/\rho(\mathbf{r})\rho(\mathbf{r}')$ and $\gamma(\mathbf{r},\mathbf{r}')$ is the density matrix. Let us assume that $P(\mathbf{r},\mathbf{r}')$ is a nonlocal kernel that captures the spatial separation between two local environments (at $\mathbf{r}$ and $\mathbf{r}'$) and has a form similar to the kernel in Eq. 3 (i.e. the kernel itself depends on the local density). Then the derivative of P (with respect to $\mathbf{r}'$) in the second and third terms on the right hand side of the above equation will also contain a derivative of the density at $\mathbf{r}'$. Therefore, to capture the kinetic energy density of a general multi-component system it is instructive to consider the derivative of the two-point correlation function in terms of a nonlocal kernel. If we limit ourselves to kernels containing only first order derivatives, a tractable form is

$$\begin{aligned} &\int\nabla_{\mathbf{r}}\rho(\mathbf{r})\cdot\nabla_{\mathbf{r}'}P\left(\mathbf{r},\mathbf{r}'\right)|_{\mathbf{r}=\mathbf{r}'}d\mathbf{r} \\ &\longrightarrow\iint\nabla_{\mathbf{r}}\rho(\mathbf{r})\cdot\nabla_{\mathbf{r}'}\rho\left(\mathbf{r}'\right)Q_3\left(\left|\mathbf{r}-\mathbf{r}'\right|,u(\mathbf{r},\mathbf{r}')\right)d\mathbf{r}'d\mathbf{r} \end{aligned} \tag{14}$$

Therefore, the third nonlocal kernel we propose to model the kinetic energy density is

$$t_{\mathrm{nl}}^{(3)}(x) = \int\nabla_x\rho(x)\cdot\nabla_{x'}\rho\left(x'\right)Q_3\left(\left|x-x'\right|,w(x,x')\right)dx'. \tag{15}$$

Since the kernel in our model is a dimensionless quantity, the right hand size of Eq. 15 satisfies the coordinate scaling criterion.

E. Numerical implementation

In this section, a detailed description of the implementation of the models discussed in the previous sections is presented. Traditional nonlocal kernels discussed in the literature are obtained from slight modifications of the Lindhard's susceptibility function or by solving a differential equation in the Fourier space[13–16]. The motivation for using Lindhard's susceptibility function is that it can capture Friedel oscillations in the electronic density[54]. Since, Lindhard's susceptibility function was derived for a uniform electron gas, the characteristic features of the nonlocal kernel for an inhomogeneous system in KSDFT has remained elusive. Motivated by recent developments in data-aided learning of partial differential equations, we learn the nonlocal kernel by using a set of basis functions[55,56]. To this end, we analyzed the effectiveness of a set of orthogonal basis functions, like spherical Bessel functions and Chebyshev functions, and a set of Gaussian functions. In our analyses, a set of Gaussian functions produced stable kernels and for all subsequent analysis we approximate the nonlocal kernels by a linear combination of standard Gaussian functions as

$$\begin{aligned} &Q_i\left(\left|x-x'\right|,w(x,x')\right) = \sum_{j=1}^{N_G}d_{i,j}e^{-\beta_{i,j}\left|x-x'\right|^2\left[w(x,x')\right]^2}, \\ &w(x,x') = \rho(x)+\rho(x'),\ \ i\in\{1,2,3\}. \end{aligned} \tag{16}$$

Here $d_{i,j}$ are the coefficients, N_G is the number of Gaussian functions, and $\beta_{i,j}$ are the width of the Gaussian functions.

Systems	Parameters	Local	Q_1	Q_2	Q_3	Q_1+Q_2	Q_2+Q_3	Q_1+Q_3	$Q_1+Q_2+Q_3$
H_2	R^2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	RMSE	7.7E-05	5.6E-05	5.9E-05	4.0E-05	4.9E-05	2.6E-05	2.8E-05	5.6E-05
	T_{KSDFT}	0.1851	0.1011	0.1011	0.1011	0.1011	0.1011	0.1011	0.1011
	T_{model}	0.1841	0.1020	0.0999	0.1018	0.1003	0.1008	0.1008	0.0985
	T_{error}	0.0010	0.0009	0.0012	0.0007	0.0008	0.0003	0.0003	0.0025
H_8	R^2	0.9847	0.9975	0.9902	0.9850	0.9983	0.9929	0.9977	0.9983
	RMSE	0.0042	0.0019	0.0037	0.0041	0.0016	0.0032	0.0018	0.0015
	T_{KSDFT}	0.3550	0.3558	0.3560	0.3550	0.3558	0.3560	0.3558	0.3788
	T_{model}	0.2591	0.3768	0.3418	0.2475	0.3686	0.3451	0.3663	0.3891
	T_{error}	0.0959	0.0210	0.0141	0.1075	0.0128	0.0108	0.0105	0.0103
LiH	R^2	0.9846	0.9974	0.9955	0.9914	0.9994	0.9968	0.9989	0.9968
	RMSE	0.0074	0.0031	0.0040	0.0055	0.0014	0.0034	0.0020	0.0030
	T_{KSDFT}	0.5403	0.5403	0.5403	0.5403	0.5020	0.5403	0.5020	0.4739
	T_{model}	0.4817	0.5714	0.5574	0.4436	0.4926	0.5261	0.5144	0.5340
	T_{error}	0.0587	0.0310	0.0170	0.0967	0.0094	0.0142	0.0125	0.0602
LiF	R^2	0.9988	0.9996	0.9994	0.9990	0.9996	0.9994	0.9997	0.9996
	RMSE	0.0050	0.0027	0.0035	0.0045	0.0029	0.0034	0.0024	0.0026
	T_{KSDFT}	2.7035	2.7035	2.4304	2.4304	2.7055	2.4304	2.4304	2.4709
	T_{model}	2.7145	2.6849	2.4079	2.4220	2.6982	2.4185	2.4084	2.4440
	T_{error}	0.0111	0.0186	0.0225	0.0083	0.0073	0.0119	0.0220	0.0269

TABLE I. A comparison of kinetic energy densities and total kinetic energies between model predictions and KSDFT results for structures in test data sets. For each system, 3 structures are used for training and 27 structures are used for testing. Here R^2 and RMSE are the coefficient of determination and the root mean square error of kinetic energy density (Ha/Bohr), respectively, T_{KSDFT} and T_{Model} are the total kinetic energies (Ha) obtained from KSDFT calculations and the various models, respectively, and T_{error} is the error in total kinetic energy (Ha) between the KSDFT value and the model. For each model type, total kinetic energies (T_{KE}, T_{model} and T_{error}) only for the structure with the maximum absolute error (out of the 27 structures used for testing) are shown. Q_1, Q_2 and Q_3 correspond to the nonlocal models in Eqs. 7, 11 and 15, respectively.

The coefficients $\beta_{i,j}$ are further written as a geometric progression $\beta_{i,j} = \beta_{i,0}\beta_{i,1}^{j}$ where $\beta_{i,0}$ and $\beta_{i,1}$ are parameters of the model[57–59]. With this approximation, we obtain the following expression to model the kinetic energy of a system in KSDFT:

$$\begin{aligned} t(x) =& c_0[\rho(x)]^3 + c_1\frac{[\rho'(x)]^2}{\rho(x)} + c_2\rho''(x) + c_{v3}\frac{\rho^{(4)}(x)}{[\rho(x)]^2+\lambda} \\ &+ c_4\frac{[\rho''(x)]^2+\rho'(x)\rho^{(3)}(x)}{[\rho(x)]^3+\lambda} + \\ &\sum_{j=1}^{N_G} d_{1,j}[\rho(x)]^2\int e^{-\beta_{1,j}|x-x'|^2[w(x,x')]^2}[\rho(x')]^2dx' + \\ &\sum_{j=1}^{N_G} d_{2,j}[\rho(x)]^2\int e^{-\beta_{2,j}|x-x'|^2[w(x,x')]^2}\left[\rho(x)-\rho(x')\right]^2dx' \\ &+ \sum_{j=1}^{N_G} d_{3,j}\nabla_x\rho(x)\int e^{-\beta_{3,j}|x-x'|^2[w(x,x')]^2}\nabla_{x'}\rho(x')dx', \\ & w(x,x') = \rho(x)+\rho(x'). \end{aligned} \tag{17}$$

Here λ is a regularization parameter. For numerical implementation, let us assume that the kinetic energy density data (for structures in the training data set) at different grid points are stored in a column vector $\mathbf{t}_s$, the coefficients c_j and $d_{i,j}$ are stored in another column vector $\mathbf{c}$, and $\mathbb{A}$ is a matrix that stores the local and nonlocal terms, whose number of rows is equal to the number of grid points used for training and the number of columns is equal to the total number of coefficients. To obtain the coefficients $\mathbf{c}$, we use a linear regression framework[60,61] and minimize the following cost function with L_2-regularization:

$$\Omega_1 = \sum_{S_u} ||\mathbf{t}_s - \mathbb{A}\mathbf{c}||^2 + \lambda_1||\mathbf{c}||^2_{L_2}. \tag{18}$$

Here, the sum S_u is over all the structure in the training data set. The parameters for the widths of Gaussian functions, $\beta_{i,0}$ and $\beta_{i,1}$, are obtained by minimizing the root mean square error in kinetic energy density.

III. RESULTS

A. DFT Data

We used four different systems, namely, H_2 molecule, H_8 chain, LiH and LiF molecules to assess the performance of our models. For each of these four systems, 30 different structures are generated and ground state change densities and kinetic energy densities are saved for training and testing. For the molecules, the bond length is varied uniformly from 3.5 to 9 Bohr to obtain a total of thirty structures. For the H_8 chain, structures are obtained by randomly perturbing the bond lengths by 5% of the original bond length (4 Bohr). For each of the four systems, we use three structures for training the model and the remaining structures are used for testing the model. The three structures used in training models

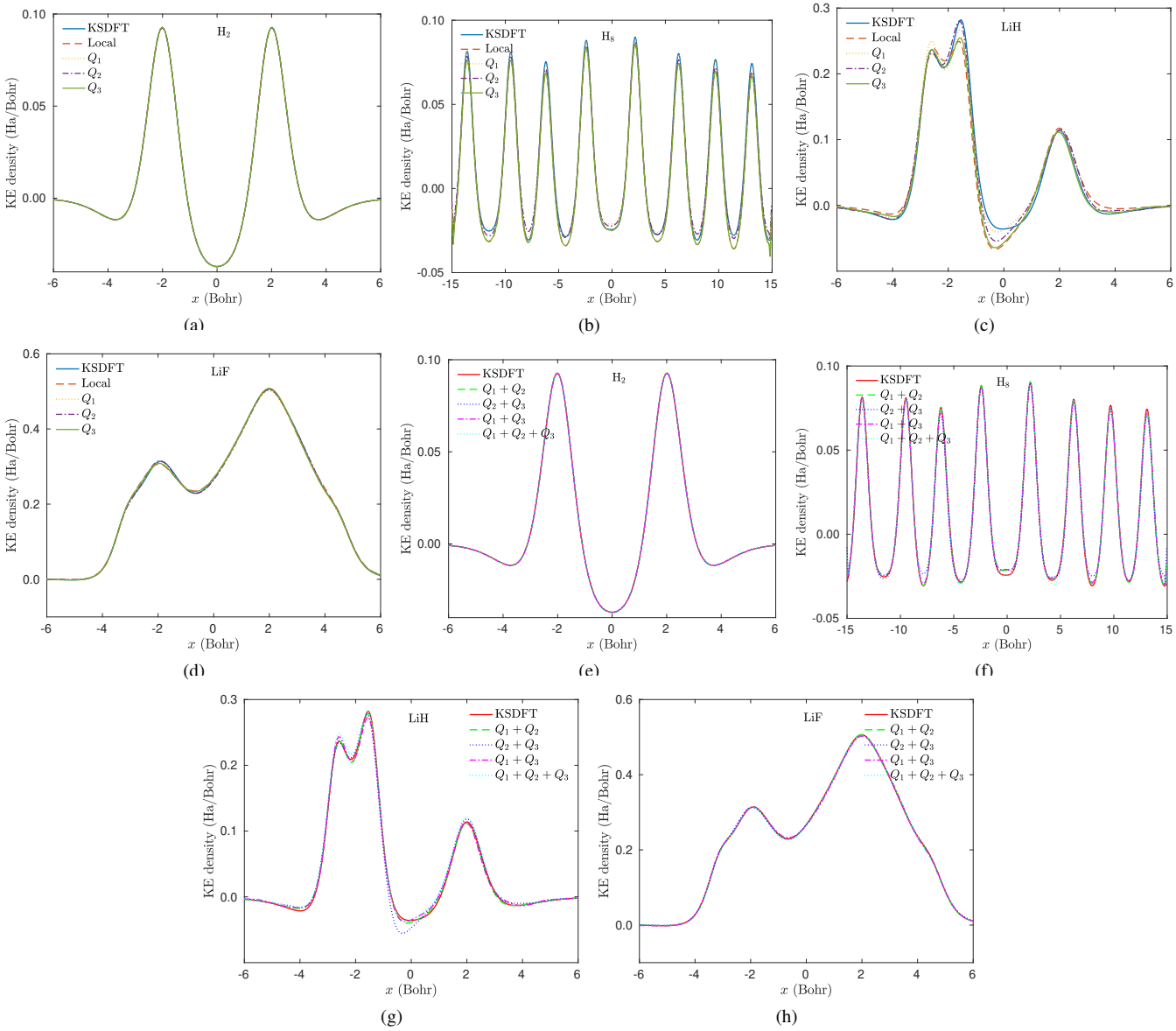


FIG. 4. Comparison of kinetic energy densities between local, nonlocal models and KSDFT results for H_2, H_8, LiH and LiF systems for representative structures not used for training. Local and nonlocal kinetic energy models for each system are trained using only 3 structures - a structure in which atoms are very close (3.5 Bohr bond length), a structure in which atoms are far apart (9 Bohr bond length) and an intermediate case. Figures *a*, *b*, *c* and *d* correspond to local, Q_1, Q_2, and Q_3 models, respectively, while *e*, *f*, *g*, and *h* correspond to Q_1+Q_2, Q_2+Q_3, Q_1+Q_3, and $Q_1+Q_2+Q_3$ models, respectively.

for the molecular systems have bond lengths of 3.5, 6 and 9 Bohr, respectively, and for H_8, three structures for training are randomly selected from the set of 30 structures. Kohn-Sham DFT data is generated using a Matlab code based on methods outlined in Section IV. To solve the one- dimensional KSDFT eigen-value problem, we used a mesh side of 0.03 Bohr, a supercell with 30 Bohr edge length and 10 k-points in the first Brillouin zone. The self-consistent iterations are considered to have converged when the change in energy per atom is less than 10^{-6} Ha.

B. Model training

To learn the kinetic energy density we consider 8 different models consisting of different combinations of semi-local and nonlocal terms shown in Eq. (17). They are (i) only the semi-local terms show in Eq. (6), (ii) semi-local terms shown in Eq. (6) and the nonlocal term, $t_{\rm nl}^{(1)}(x)$, shown in Eq. (7), (iii) semi-local terms shown in Eq. (6) and the nonlocal term, $t_{\rm nl}^{(2)}(x)$, shown in Eq. (11), (iv) semi-local terms shown in Eq. (6) and nonlocal term, $t_{\rm nl}^{(3)}(x)$, shown in Eq. (15), (v) semi-local terms shown in Eq. (6) and nonlocal terms $t_{\rm nl}^{(1)}(x)$ and $t_{\rm nl}^{(2)}(x)$, (vi)

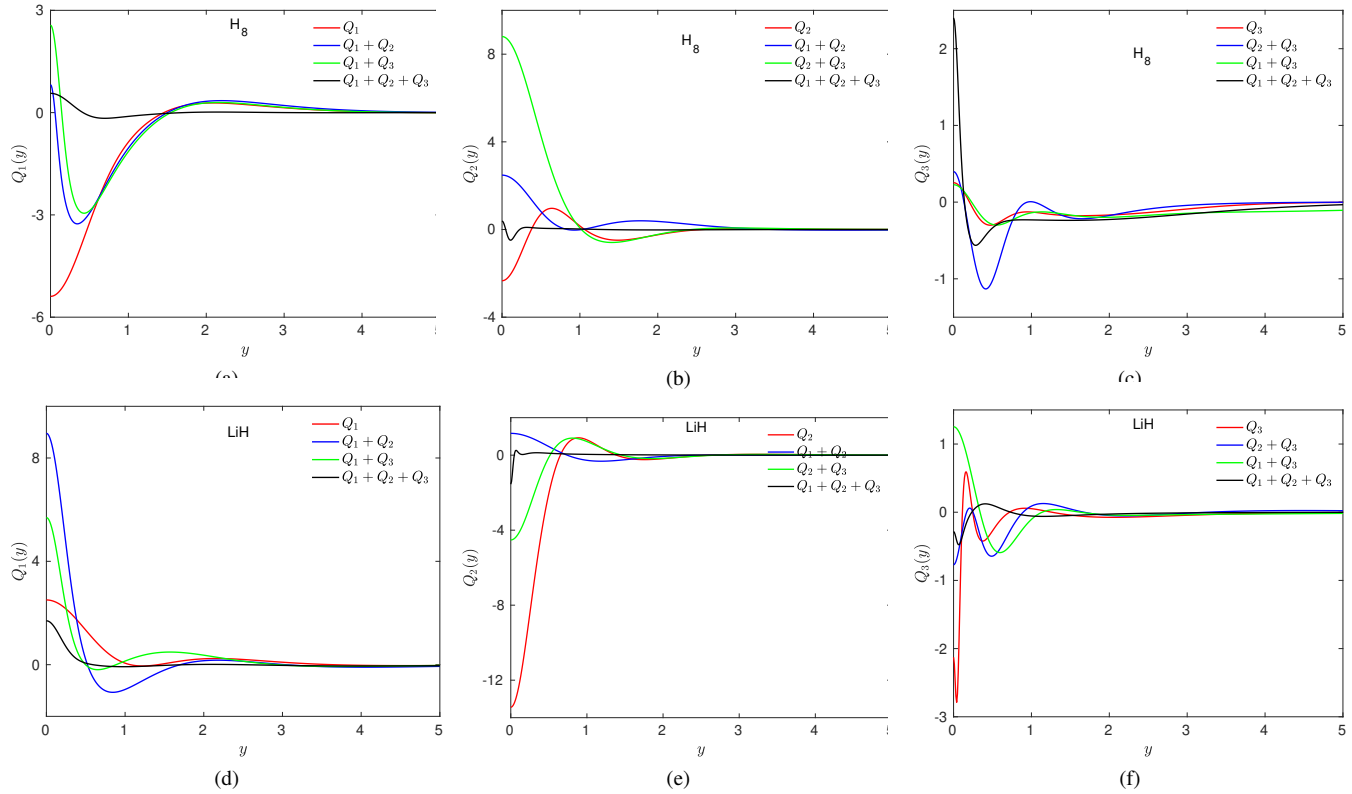


FIG. 5. Shown here are the kernels learned for each of the seven nonlocal models for two representative systems: LiH (*d*, *e*, *f*) and a one-dimensional chain of eight hydrogens, i.e. H_8 (*a*, *b*, *c*). Figures *a* and *d* show the kernel Q_1 for different models, *b* and *e* show the kernel Q_2, and *c* and *f* show the kernel Q_3. The abscissa y is a dimensionless variable given by $y = |x - x'|\,(\rho(x) + \rho(x'))$.

semi-local terms shown in Eq. (6) and nonlocal terms $t_{\rm nl}^{(2)}(x)$ and $t_{\rm nl}^{(3)}(x)$, (vii) semi-local terms shown in Eq. (6) and nonlocal terms $t_{\rm nl}^{(1)}(x)$ and $t_{\rm nl}^{(3)}(x)$, (viii) semi-local terms shown in Eq. (6) and all the three nonlocal terms $t_{\rm nl}^{(1)}(x)$, $t_{\rm nl}^{(2)}(x)$ and $t_{\rm nl}^{(3)}(x)$.

To model the nonlocal kernels in $t_{\rm nl}^{(1)}(x)$, $t_{\rm nl}^{(2)}(x)$ and $t_{\rm nl}^{(3)}(x)$ we used 6 Gaussian functions - this number of basis functions corresponds to a minimum in the mean squared error. For models containing the local terms and one nonlocal term, the total number of coefficients is 11. Similarly, for models containing the local terms and two (or three) nonlocal terms, the total number of coefficients is 17 (or 23). The widths of these Gaussian functions are optimized by using the Nelder-Mead simplex method and the coefficients of the local terms and the Gaussian functions (used to model the nonlocal kernels) are obtained by optimizing the cost function in Eq. (18) using linear regression. For these optimizations, the regularization parameter is set to 10^{-6}.

For all systems considered in this study, we use 3 structures for training and 27 structures for testing and validation. For molecules, like H_2, LiF and LiH, structures with 3.5 and 9 Bohr bond lengths (i.e. atoms are too close and far apart) and a structure with intermediate bond length are used. Structures in the testing set include systems with bond lengths in between the two extremities, i.e. 3.5 and 9 Bohr. For H_8, training and testing sets contain structures in which atomic positions are perturbed using random numbers between -0.50 and 0.50.

C. Kinetic energy density prediction

Table I shows the performance of the proposed models - measured in terms of metrics, like coefficient of determination (i.e. R^2 score), root mean square error (RMSE), and the maximum error in the total kinetic energy - for a few representative systems. The coefficient of determination and RMSE values are calculated by using all the 27 structures in the training data sets, while only the maximum errors in the predicted total kinetic energies are shown in Table I. For all the systems considered in this study, fitting the coefficients of the local terms results in models that exhibit superior performance than the model shown in Eq. (2). For example, the root mean squared errors are less than 0.008 Ha/Bohr and the maximum error in the total kinetic energy is less than 0.09 Ha. The fitted coefficients, shown in Table 1 in the supplementary material, contain a mixture of positive and negative terms and are different from the coefficients of the local terms in Eq. (2). We observe a decrease in RMSE values for all 4 systems when nonlocal terms, like $t_{\rm nl}^{(1)}$, $t_{\rm nl}^{(2)}$ or $t_{\rm nl}^{(3)}$, are included in the model along the with local terms. For majority of the test systems (except for LiF with $t_{\rm nl}^{(1)}$, LiH with $t_{\rm nl}^{(2)}$, etc.), errors in total kinetic en-

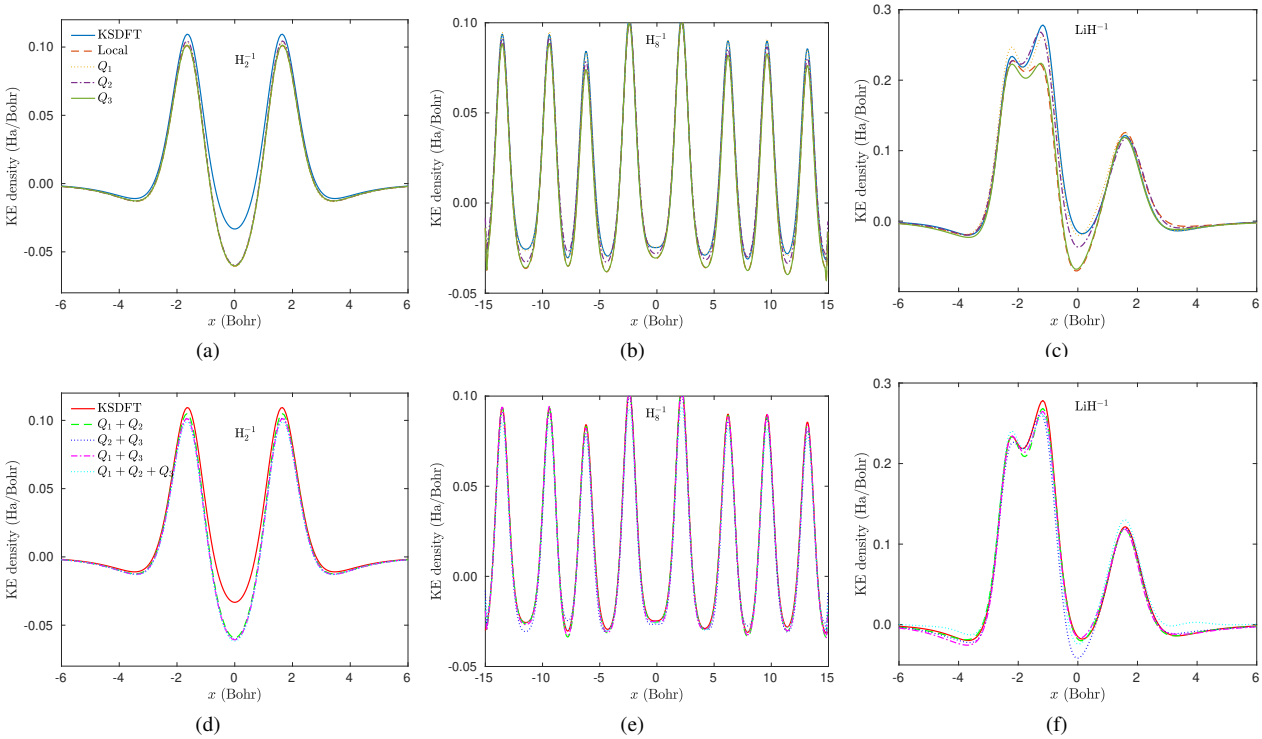


FIG. 6. Performance of all 8 local and nonlocal models on negatively charged systems: H_2^- (*a* and *d*), H_8^- (*b* and *e*), and LiH^- (*c* and *f*) . Models for each of the four systems are trained using 3 neutral structures. Figures *a*, *b* and *c* show kinetic energy densities obtained by using local, Q_1, Q_2, and Q_3 models, respectively. Figures *d*, *e*, and *f* show kinetic energy densities obtained by using Q_1+Q_2, Q_2+Q_3, Q_1+Q_3, and $Q_1+Q_2+Q_3$ models, respectively.

ergies also decreased for models containing a nonlocal term. This suggests that capturing spatial correlations in gradient of the electron density is important to learn the kinetic energy density.

From Table I, we also see that RMSE values decrease when two or three nonlocal terms are present in the models. Since the three nonlocal terms incorporate spatial correlations in density, fluctuations in density and gradient of density, intuitively we expect them to be mostly orthogonal to each other. However, the increase in RMSE values for some systems with 2 or 3 nonlocal terms are present suggests that (1) these nonlocal terms have a large overlap (i.e. they have large collinear components), or (2) the increase in number of fitting parameters leads to over-fitting.

As shown in Fig. 3, kinetic energy densities of hydrogen molecules predicted by all of the eight models lie nearly on top of the KSDFT curve. This is also evident from the fact that the RMSE values shown in Table I are on the order of 10^{-5} Ha/Bohr. In case of LiF too, the local and nonlocal models are able to model the kinetic energy densities accurately. However, for H_8 and LiH, local models and models containing only one nonlocal term are unable to accurately model the kinetic energies of all the structures in the test data set. For example, in case of LiH, the model containing $t_{\rm nl}^{(3)}(x)$ results in oscillations in the kinetic energy (at $x \sim \pm 0$). Similarly, kinetic energies in regions between two atoms are also not correctly captured in case of H_8 using models contain one nonlocal term. However, as shown in Figs. 4 (c), (g) majority of these issues are resolved when two nonlocal terms are included in the model.

Since we used a least-square procedure for training, we see a nearly uniform distribution of grid points where model predictions show either positive or negative deviations from the KSDFT calculated kinetic energy density values. This means that the predicted total kinetic energy can exhibit positive or negative deviations from the KSDFT value. At first glance, this may seem to violate the $N-$representability constraint,[62] of the kinetic energy, but this is simply a result of the fitting procedure of minimizing the magnitude of model errors during training. In principle, it is possible to impose a penalty function to make sure that the model predictions are always higher than KSDFT values. But this leads to larger errors and worse models without having any guarantee that the final model is N-representable as the condition $\int \mathbf{t} d^3\mathbf{r} < \int \mathbf{t}_{\rm DFT} d^3\mathbf{r}$ is not a sufficient one. Since, no pathological issues have been reported for functionals that do not satisfy $\int \mathbf{t} d^3\mathbf{r} < \int \mathbf{t}_{\rm DFT} d^3\mathbf{r}$,[63,64] and the fact that many data-driven models[37,39,65] reported in the literature also do not obey this condition, we focus only on developing models that minimize error magnitudes.

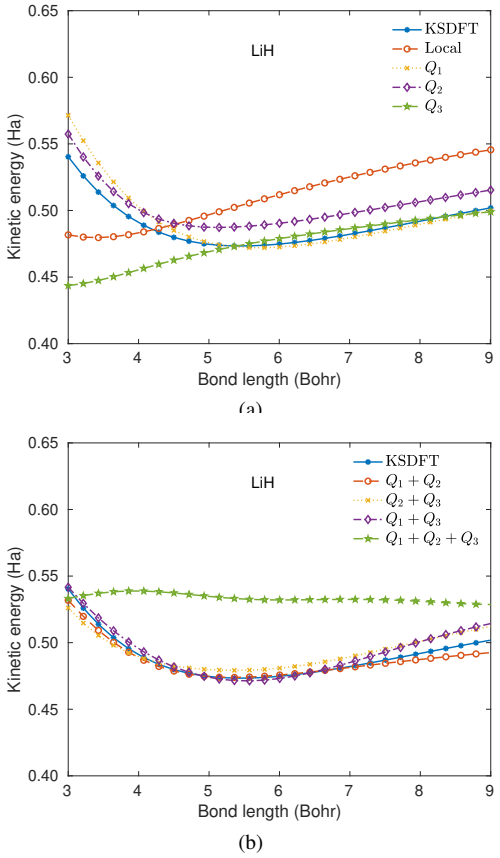


FIG. 7. Shown here are the variations in total kinetic energies (for KSDFT, local and nonlocal models) with respect to the bond length of LiH molecule.

D. Characteristic features of nonlocal kernels

The kernel shown in Fig. 3 is representative of many other nonlocal kernels that have been proposed in the literature. Since such a kernel is unable to accurately capture the kinetic energy densities of simple systems, like H_2 and H_8 (e.g. Figs. 2(a)-(b)), it is intuitive to ask how different are the kernels learned from KSDFT data. To this end, in Fig. 5 we show nonlocal kernels for two representative systems - H_8 chain and LiH molecule. We specifically focus on these two systems because their kinetic energy densities show features, such as splitting of the kinetic energy peak located close to the Li atom position (see Fig. 4) which are difficult to capture.

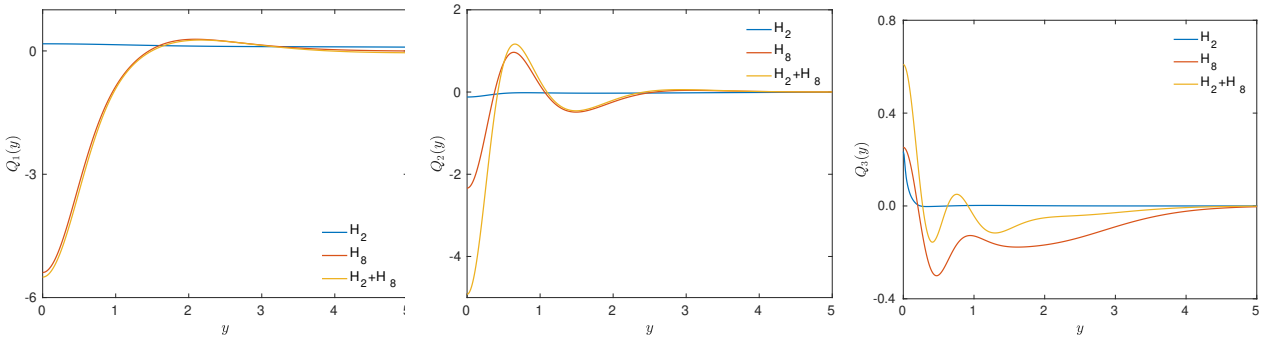
All the kernels shown in Fig. 5 exhibit similar features, i.e. they are nonzero at $y = |x-x'|\left(\rho\left(x\right)+\rho\left(x'\right)\right) = 0$, but they exhibit an oscillatory behavior as y increases from 0 and finally decay to zero at large values of y. However, unlike the kernel in Fig. 3, we do not see multiple humps in the kernel for the model containing the local terms and $t_{\rm nl}^1\left(x\right)$ for H_8 and LiH systems. In fact, for LiH, the kernel Q_1 remains positive, while for H_8 the kernel is negative only near $y \sim 1$ (see Fig. 5). We also see that a particular type of kernel often displays features that are common in all the nonlocal models. For example, the kernel Q_1 (in Fig. 5 (a), (d)) reconstructed from KSDFT data of LiH and H_8 in the model containing only Q_1, and in the model containing $Q_1+Q_2+Q_3$ have similar features. The same is true for the nonlocal kernel Q_3 (see Fig. 5 (c), (f)).

To further analyze the behavior of the nonlocal kernels, we trained models using structures from different systems. To this end, we used 3 structures from each of two systems - H_2 and H_8, and trained eight models. As evident from Fig. 8, the characteristic features of the kernels Q_1, Q_2 and Q_3 for H_8 and H_8+H_2 systems are very similar, but amplitudes of fluctuations in these kernels for H_2 are much smaller than kernels for H_8 and H_8+H_2 systems. In addition, magnitudes and signs of the kernels Q_1 and Q_2 at $y = 0$ for models trained using only H_8 structures and those trained using both H_2 and H_8 structures are different. This suggests that contributions from some of the local terms also change as the diversity of the training data set changes. Nevertheless, the oscillations in these kernels are similar to those reported in the literature obtained by using the Lindhard's susceptibility function, and we believe that our models can be used to learn kernels for other complex systems and capture effects like Friedel oscillations.

E. Behavior outside the training region: Performance for charged systems

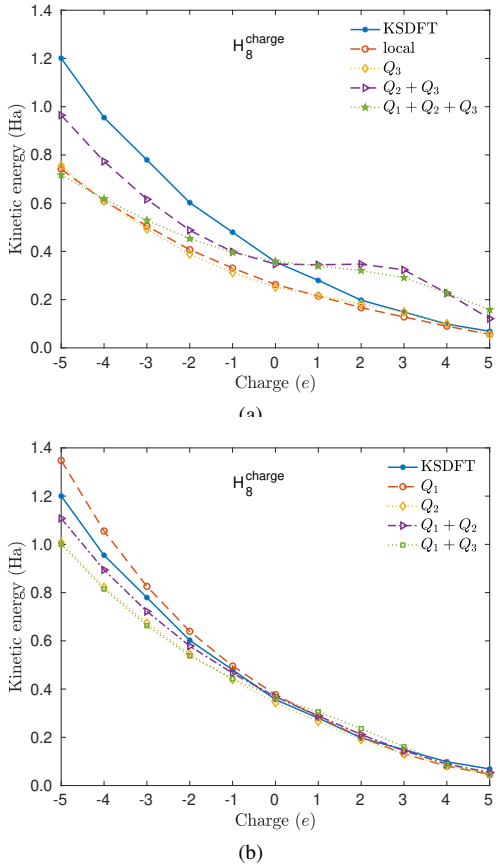
Next, to study the performance of the models outside of their training region, we analyzed the performance of our models (trained on neutral structures) by inserting one extra electron into the test structures of the 4 systems. Data (i.e. ground state electron densities and kinetic energy densities) for such charged systems is generated by adding a jellium external potential to our KSDFT formulation. We show the predicted kinetic energy densities for these systems (one representative structure for each system is shown) in Fig. 6 using models trained with neutral systems. Interestingly, all the eight models perform satisfactorily for the charged systems, though we observe that models with either Q_2 or Q_3 perform better than the rest. Here we also note that for a neutral H_2 molecule, only a single eigenfunction is used to calculate the ground state change density. Therefore, the discrepancies in the predicted kinetic energy densities for H_2^- can be attributed to the presence of additional nonlinearities in the system due to the inclusion of an additional eigenfunction in the electron density and kinetic energy calculations. In addition, it is evident from Figs. 6(b) and 6(c) that only Q_1 is not sufficient to capture all the features of the kinetic energy density. This suggests that nonlocal terms containing the gradient of the density are need to accurately model the kinetic energy density.

To further understand the suitability of the proposed models, we study the performance of our models (trained using three neutral structures) on highly charged H_8 systems. Figure 9 shows total kinetic energies of systems containing different number of electrons obtained from all of the eight local and nonlocal models. Interestingly, we observe that the local model as well as the nonlocal models containing Q_3, Q_2+Q_3 and $Q_1+Q_2+Q_3$ perform very poorly with increase in the charge on the system. On the other hand, models containing



(a) (b) (c)

FIG. 8. Comparison between kernels reconstructed for three different systems - one containing only H_2 structures, another containing only H_8 structures and a third containing 3 structures each from H_2 and H_8. The blue, orange and yellow lines in the plots correspond to the models trained with three H_8 structure, three H_2 structures and a combination of three H_2 and three H_8 structures, respectively.



(a)

(b)

FIG. 9. Shown here are the variations in total kinetic energies (for local and nonlocal models) with respect to the net charge on systems containing a one-dimensional chain of eight hydrogens. Models containing the nonlocal kernel Q_2 are shown in (a), while the remaining local and nonlocal models are shown in (b).

Q_1 along with Q_2 or Q_3 are able to capture the trend. These observations highlight the importance of incorporating spatial correlations of derivatives of the density and local fluctuations in the electron density, and models containing these quantities exhibit better transferability. The poor performance of Q_3 also suggests poor transferability of models just based on the gradient information.

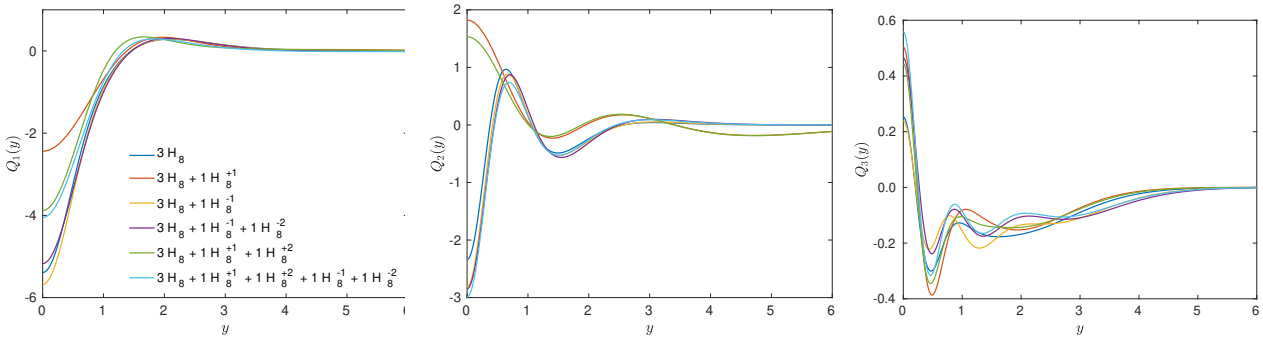
Next, we study changes in the three nonlocal kernels Q_1, Q_2 and Q_3 by including charged systems during the training process. nonlocal kernels obtained from these updated training sets are shown in Fig. 10. Interestingly, shapes of these kernels remain essentially the same, however values of the kernels near $y = 0$ changes. A similar behavior is also evident in the kernels shown in Fig. 5 and this also suggests that the semi-local part of our model is perhaps not complete and the missing contributions in the semi-local part are being captured by the nonlocal kernels.

F. Behavior outside the training region: Kinetic energy density during a self-consistent iteration

Finally, we verify the stability and performance of proposed models during a self-consistent field (SCF) iteration of an OFDFT calculation. To this end, we mimic an actual SCF iteration during an OFDFT simulation by storing electron densities and kinetic energy densities for each SCF step for the H_8 system during a KSDFT calculation. Then, the stored electron densities are taken as an inputs for the nonlocal models with Q_1, Q_2, Q_3 and Q_1+Q_3 and the predicted kinetic energy densities are compared with the corresponding KSDFT values in Fig. 11. The model containing the nonlocal kernel Q_2 performs best among the three models considered in this analysis and the maximum error in the total kinetic energy (after first SCF iteration) is about 0.017 Ha (i.e. 0.0021 Ha/atom). The error in total kinetic energy is about 0.012 Ha (i.e. 0.0015 Ha/atom) for the Q_1+Q_3 model. Motivated by the performance of our models, we conjecture that such models can be used to perform OFDFT calculations.

G. Results from 3D analysis

It is clear from the analysis of one-dimensional systems that models containing two nonlocal terms are optimal in



(a) (b) (c)

FIG. 10. Shown here are the kernels Q_1 ,Q_2 and Q_3 reconstructed from a diverse set of structures. The blue, orange and yellow lines in the plots correspond to the models trained with three H_8 structure, three H_2 structures and a combination of 3 H_2 and three H_8 structures, respectively. The models are trained with a combination of three neutral H_8 structure and some combination of charged H_8 ions as shown in the legends. For example, $3H_8$+$1H_8^{+1}$ in the legend corresponds to a model trained with three neutral H_8 and one positively charged H_8 ion.

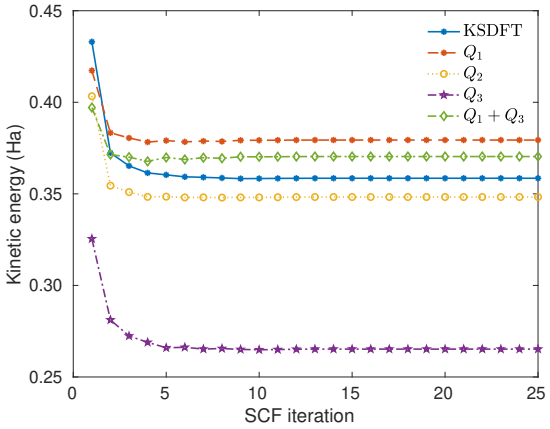


FIG. 11. Shown here are the predicted total kinetic energies of a H_8 system (predicted using models containing Q_1, Q_2, Q_3 and Q_1+Q_3) during a Kohn-Sham self-consistent field (SCF) iteration.

terms of RMSE scores as well as the number of parameters. In addition, the nonlocal term with kernel Q_2 is closely related the kernel Q_3, i.e., $t_{\rm nl}^{(2)}(x) = [\rho(x)]^4 \int Q_2 dx' + [\rho(x)]^2 \int Q_2 [\rho(x')]^2 dx' - 2[\rho(x)]^2 \int Q_2 \rho(x')\, dx'$. Thus, we analyze the performance of the following model containing Q_1 and Q_3 for three dimensional systems:

$$
\begin{aligned}
t(\mathbf{r}) &= c_0 \rho^{5/3}(\mathbf{r}) + c_1 \frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} + c_2 \Delta\rho(\mathbf{r}) \\
&+ c_3 \frac{(\Delta\rho(\mathbf{r}))^2}{\rho^{5/3}(\mathbf{r})} + c_4 \frac{|\nabla\rho(\mathbf{r})|^4}{\rho^{11/3}(\mathbf{r})} + c_5 \frac{\Delta\rho(\mathbf{r})\,|\nabla\rho(\mathbf{r})|^2}{\rho^{8/3}(\mathbf{r})} \\
&+ \rho^m(\mathbf{r}) \int Q_1\left(w\left(|\mathbf{r}-\mathbf{r}'|^2, \mathbf{r}, \mathbf{r}'\right)\right) \rho^m(\mathbf{r}')\, d\mathbf{r}' \\
&+ \int \nabla\rho(\mathbf{r}) \cdot \nabla\rho(\mathbf{r}')\, Q_3\left(w\left(|\mathbf{r}-\mathbf{r}'|^2, \mathbf{r}, \mathbf{r}'\right)\right) d\mathbf{r}' \\
w\left(|\mathbf{r}-\mathbf{r}'|^2, \rho(\mathbf{r}), \rho(\mathbf{r}')\right) &= |\mathbf{r}-\mathbf{r}'|^2 \left[\rho^{2/3}(\mathbf{r}) + \rho^{2/3}(\mathbf{r}')\right].
\end{aligned}
\tag{19}
$$

As in case of 1D, we use a set of Gaussian functions to model the nonlocal kernels and our kernel depends on the density at $\mathbf{r}$ and $\mathbf{r}'$. The exponent of the density in the nonlocal term containing Q_1 is set to $m = 5/6$. The functional form of w in Eq. 19 and the value of exponent m are motivated by models proposed in Ref. [16,44].

For general three-dimensional systems pseudo-potentials are used in routine calculations, however, in absence of orbitals, the nonlocal part in the pseudo-potential cannot be handled straightforwardly within the framework of OFDFT. For this purpose, we focus here on light elements and perform all electron calculations to generate data for training and testing (see Appendix D). We consider two systems: helium dimer and lithium dimer. These systems allow us to demonstrate our capability to model kinetic energy densities when atoms are very close and when they are far apart. For each system, we use 3 structures with bond lengths 2,4 and 6 a.u. for training and four structures (bond lengths of 2.5, 3, 5, 5.5 a.u.) for testing the model. Since we have a small molecule in a large box, we only consider grid points with electron densities greater than 0.05 e/a.u.^{3} for training (about 5×10^3-1.3×10^4 grid points per structure) and testing purposes.

In case of He, we obtain a RMSE score of 0.003 Ha/a.u.^{3} for structures in the test set while for lithium dimer we obtain a RMSE score of 0.017 Ha/a.u.^{3}. The R^2 scores for the predicted kinetic energy densities for helium and lithium dimers are better than 0.9999, which suggests that scatter in the data (due to basis set errors, see Appendix D) is responsible for the disparity in the RMSE scores of Li and He systems. In addition, the mean error in the predicted total kinetic energies of test structures for both the systems is about 0.02 Ha, which is close to the values obtained for one-dimensional systems.

H. Comparison with other models proposed in the literature

Now, we compare our results to the models that have been proposed in the literature. Snyder et al.[39] used kernel ridge regression to directly learn the kinetic energy functional from

the Kohn-Sham electron densities for one dimensional systems. We observe that the performance of our model is comparable in terms of the error in kinetic energy. They obtained a maximum error in kinetic energy of around 2×10^{-4} Ha for H_2 while we obtain a value of around 3×10^{-4}. Golub and Manzhos[37,65] , Seino et al.[38] have proposed different models for three dimensional systems by using neural networks and Gaussian process regression. We see that the performance of our models in terms of the root mean square errors in the predicted kinetic energy densities are similar to those proposed by these authors. In principle, a neural network or a deep learning based model can learn very complex functions with a large dataset.[66–69] In this regard, for simple molecules, it seems that these a few physics-inspired terms are sufficient to model the kinetic energy density. However, we note that the complexity of the framework proposed here can be increased by expanding the dictionary of physics-inspired terms by including higher order nonlocal terms[16,31,45] and we plan to expand our study to include crystalline solids in future. Our model has the advantage of requiring a much smaller training data set (electron densities and kinetic energy densities from only 3 structures in the training set, 10^4 grid points in total) than $\sim 10^6$ grid points required to train neural network models. In addition, the unknown coefficients in our framework are simply obtained by solving a linear system which makes it easy to implement and computationally efficient.

IV. CONCLUSION

In this contribution, a data-driven, physics-informed prescription is proposed to model the kinetic energy density of an inhomogeneous system using data from KSDFT calculations. To model the nonlocal contributions, we have considered two new nonlocal terms - one that includes local fluctuations in the density and another that includes gradient information. The proposed models are used to predict kinetic energy densities and total energies of molecules, such as H_2, LiH and LiF, and one dimensional chain of eight hydrogens. The two new nonlocal terms are shown to help in improving the predictive capability of the models for systems with complicated kinetic energy density profiles (like H_8) and for systems not present in the training set (e.g. systems containing extra electrons and kinetic energy densities from a self-consistent field iteration). In addition, we have used a simple procedure to learn the nonlocal kernels by using a set of Gaussian basis functions. Using only a few training structures our models are able to achieve high accuracy (i.e. root mean square error of 0.003 Ha/atom in the total kinetic energy).

From the analysis of the models with one, two or three nonlocal terms (Section IV.C) we conclude that models containing two kernels most suitable because they contain an optimum number of parameters (thereby minimize over-fitting) and have lower errors in the predicted kinetic energy densities and total energies. In addition, it is shown in Section IV.E that the nonlocal models with Q_1 lacks transferability. However, the performance is significantly improved when Q_1 is used along with Q_2 or Q_3. Similarly, in Section IV.F, it is clear that the nonlocal with Q_1 is needed to correctly reproduce the kinetic energy density. Since there is some collinearity between Q_2 and Q_3, therefore we propose the following model to learn kinetic energy density from Q_1 and Q_3:

$$
\begin{aligned}
t(x) =& c_0[\rho(x)]^3 + c_1\frac{[\rho'(x)]^2}{\rho(x)} + c_2\rho''(x) + c_{v3}\frac{\rho^{(4)}(x)}{[\rho(x)]^2+\lambda} \\
&+ c_4\frac{[\rho''(x)]^2+\rho'(x)\rho^{(3)}(x)}{[\rho(x)]^3+\lambda} + \\
&\sum_{j=1}^{N_G} d_{1,j}[\rho(x)]^2\int e^{-\beta_{1,j}|x-x'|^2[w(x,x')]^2}[\rho(x')]^2dx' \\
&+ \sum_{j=1}^{N_G} d_{3,j}\nabla_x\rho(x)\int e^{-\beta_{3,j}|x-x'|^2[w(x,x')]^2}\nabla_{x'}\rho(x')dx', \\
&w(x,x') = \rho(x)+\rho(x').
\end{aligned}
\tag{20}
$$

Here, λ is a regularization parameter. The ability of the models discussed here to model the kinetic energy density and the total kinetic energy suggests that this is a promising route to develop reliable models for OFDFT calculations. Given the need to study larger system sizes beyond those accessible in KSDFT calculations and the fact the developing a universal density-dependent, kinetic energy functional is challenging, our results show that there is a more amenable path: Instead of sampling the whole configuration space of Kohn-Sham kinetic energy and electron densities, it is easier to develop models that can learn a small part of the configuration space using a set of physics-informed descriptors.

SUPPLEMENTARY MATERIAL

See the supplementary material for the coefficients of the local terms for the proposed models, distribution of error in the kinetic energy for the testing data-set, and the Matlab code to fit the models.

IM release #: LLNL-JRNL-828938, LDRD: 21-ERD-005

ACKNOWLEDGMENT

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Computing support for this work came from the Lawrence Livermore National Laboratory (LLNL) institutional computing facility.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon request.

REFERENCES

[1] J. Chen, X.-Z. Li, Q. Zhang, M. I. Probert, C. J. Pickard, R. J. Needs, A. Michaelides, and E. Wang, Nature communications **4**, 1 (2013).

[2] T. A. Pham, K. E. Kweon, A. Samanta, V. Lordi, and J. E. Pask, Journal of Physical Chemistry C **121**, 21913 (2017), https://doi.org/10.1021/acs.jpcc.7b06457.

[3] M. Alemany, X. Huang, M. L. Tiago, L. Gallego, and J. R. Chelikowsky, Solid State Communications **146**, 245 (2008), special section on Negative Refraction and Metamaterials for Optical Science and Engineering.

[4] E. A. C. Yan Alexander Wang, *Theoretical Methods in Condensed Phase Chemistry*, 1st ed., edited by S. D. Schwartz (Springer, Dordrecht, 2002).

[5] V. V. Karasiev, D. Chakraborty, and S. B. Trickey, *Many-Electron Approaches in Physics, Chemistry and Mathematics*, 1st ed., edited by V. Bach and L. D. Site (Springer, Cham, 2014).

[6] M. Chen, X.-W. Jiang, H. Zhuang, L.-W. Wang, and E. A. Carter, Journal of Chemical Theory and Computation **12**, 2950 (2016).

[7] B. G. del Rio, J. M. Dieterich, and E. A. Carter, Journal of Chemical Theory and Computation **13**, 3684 (2017).

[8] W. C. Witt, B. G. del Rio, J. M. Dieterich, and E. A. Carter, Journal of Materials Research **33**, 777 (2018).

[9] W. Mi and M. Pavanello, Physical Review B **100**, 041105 (2019).

[10] P. Golub and S. Manzhos, Computer Physics Communications **256**, 107365 (2020).

[11] K. Luo, V. V. Karasiev, and S. B. Trickey, Phys. Rev. B **101**, 075116 (2020).

[12] P. García-González, J. Alvarellos, and E. Chacón, Physical Review B **53**, 9509 (1996).

[13] E. Chacón, J. Alvarellos, and P. Tarazona, Physical Review B **32**, 7868 (1985).

[14] C. Huang and E. A. Carter, Physical Review B **81**, 045206 (2010).

[15] Y. A. Wang, N. Govind, and E. A. Carter, Physical Review B **60**, 16350 (1999).

[16] L.-W. Wang and M. P. Teter, Physical Review B **45**, 13196 (1992).

[17] S. Ghosh and P. Suryanarayana, Journal of Computational Physics **307**, 634 (2016).

[18] L. Šamaj and J. Percus, Journal of chemical physics **111**, 1809 (1999).

[19] C. Herring, Physical Review A **34**, 2614 (1986).

[20] P. Hohenberg and W. Kohn, Physical review **136**, B864 (1964).

[21] M. Plumer and D. Geldart, Journal of Physics C: Solid State Physics **16**, 677 (1983).

[22] E. Combariza, E. Chacón, and P. Tarazona, Physica A: Statistical Mechanics and its Applications **180**, 225 (1992).

[23] D. García-Aldea and J. E. Alvarellos, Journal of Chemical Physics **129**, 074103 (2008).

[24] D. García-Aldea and J. E. Alvarellos, Phys. Chem. Chem. Phys. **14**, 1756 (2012).

[25] J. P. Perdew and L. A. Constantin, Phys. Rev. B **75**, 155109 (2007).

[26] S. Laricchia, L. A. Constantin, E. Fabiano, and F. Della Sala, Journal of chemical theory and computation **10**, 164 (2014).

[27] V. V. Karasiev, D. Chakraborty, O. A. Shukruto, and S. B. Trickey, Phys. Rev. B **88**, 161108 (2013).

[28] L. A. Constantin, E. Fabiano, and F. Della Sala, Journal of Physical Chemistry Letters **9**, 4385 (2018).

[29] K. Luo, V. V. Karasiev, and S. B. Trickey, Phys. Rev. B **98**, 041111 (2018).

[30] L. A. Constantin, E. Fabiano, and F. Della Sala, Journal of Chemical Theory and Computation **15**, 3044 (2019).

[31] D. García-Aldea and J. E. Alvarellos, The Journal of Chemical Physics **127**, 144109 (2007).

[32] S. Liu, D. Zhao, C. Rong, T. Lu, and S. Liu, Journal of Chemical Physics **150**, 204106 (2019).

[33] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, Physical review letters **108**, 253002 (2012).

[34] L. Li, T. E. Baker, S. R. White, and K. Burke, Phys. Rev. B **94**, 245129 (2016).

[35] R. Meyer, M. Weichselbaum, and A. W. Hauser, Journal of chemical theory and computation **16**, 5685 (2020).

[36] S. A. Ghasemi and T. D. Kühne, Journal of Chemical Physics **154**, 074107 (2021).

[37] P. Golub and S. Manzhos, Physical Chemistry Chemical Physics **21**, 378 (2019).

[38] J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, Journal of chemical physics **148**, 241705 (2018).

[39] J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K.-R. Müller, and K. Burke, Journal of Chemical Physics **139**, 224104 (2013).

[40] E. M. Stoudenmire, L. O. Wagner, S. R. White, and K. Burke, Phys. Rev. Lett. **109**, 056402 (2012).

[41] L. O. Wagner, T. E. Baker, E. M. Stoudenmire, K. Burke, and S. R. White, Phys. Rev. B **90**, 045109 (2014).

[42] W. C. Witt and E. A. Carter, Physical Review B **100**, 125107 (2019).

[43] V. L. Lignères and E. A. Carter, in *Handbook of materials modeling* (Springer, 2005) pp. 137–148.

[44] Y. A. Wang, N. Govind, and E. A. Carter, Physical Review B **58**, 13465 (1998).

[45] A. A. Kugler, Physical Review A **41**, 3489 (1990).

[46] M. Levy and J. P. Perdew, Phys. Rev. A **32**, 2010 (1985).

[47] A. Holas, P. Kozlowski, and N. March, Journal of Physics A: Mathematical and General **24**, 4249 (1991).

[48] P. Kozlowski and N. March, Physical Review A **39**, 4270 (1989).

[49] P. Kozlowski and N. March, EPL (Europhysics Letters) **11**, 613 (1990).

[50] A. Holas and N. March, International Journal of Quantum Chemistry **56**, 371 (1995).

[51] A. Holas and N. March, Physical Review A **64**, 016501 (2001).

[52] W. Jones and W. Young, Journal of Physics C: Solid State Physics **4**, 1322 (1971).

[53] S. Liu and R. G. Parr, Journal of chemical physics **106**, 5578 (1997).

[54] N. W. Ashcroft, N. D. Mermin, *et al.*, *Solid state physics*, Vol. 2005 (holt, rinehart and winston, new york London, 1976).

[55] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Proceedings of the National Academy of Sciences **116**, 22445 (2019).

[56] H. Schaeffer, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **473**, 20160446 (2017).

[57] D. P. Chong, Canadian journal of chemistry **73**, 79 (1995).

[58] S. Lehtola, Journal of chemical physics **152**, 134108 (2020).

[59] E. S. Kryachko and S. Wilson, International journal of quantum chemistry **93**, 112 (2003).

[60] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction* (Springer Science & Business Media, 2009).

[61] C. M. Bishop, *Pattern recognition and machine learning* (springer, 2006).

[62] P. W. Ayers and S. Liu, Phys. Rev. A **75**, 022514 (2007).

[63] E. S. Kryachko and E. V. Ludeńa, *Energy Density Functional Theory of Many-Electron Systems*, 1st ed. (Springer, Dordrecht, 1990).

[64] M. Piris, *Many-body Approaches at Different Scales A Tribute to Norman H. March on the Occasion of his 90th Birthday*, 1st ed., edited by G. G. N. Angilella and C. Amovilli (Springer, Cham, 2018).

[65] S. Manzhos and P. Golub, Journal of Chemical Physics **153**, 074104 (2020).

[66] K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359 (1989).

[67] G. Cybenko, Mathematics of Control, Signals and Systems **2**, 303 (1989).

[68] J. Park and I. W. Sandberg, Neural Computation **3**, 246 (1991).

[69] A. Barron, IEEE Transactions on Information Theory **39**, 930 (1993).

[70] W. Kohn and L. J. Sham, Physical review **140**, A1133 (1965).

[71] F. Bloch, Zeitschrift für physik **52**, 555 (1929).

[72] L. O. Wagner, E. Stoudenmire, K. Burke, and S. R. White, Physical Chemistry Chemical Physics **14**, 8581 (2012).

[73] T. E. Baker, E. M. Stoudenmire, L. O. Wagner, K. Burke, and S. R. White, Physical Review B **91**, 235141 (2015).

[74] P. Pulay, Chemical Physics Letters **73**, 393 (1980).

[75] S. Ghosh and P. Suryanarayana, Computer Physics Communications **216**, 109 (2017).

[76] B. Fornberg, Mathematics of computation **51**, 699 (1988).

[77] S. Obara and A. Saika, Journal of Chemical Physics **84**, 3963 (1986).

[78] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1st ed. (Dover Publications, Inc., 1996).

[79] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. L. Windus, Journal of Chemical Information and Modeling **47**, 1045 (2007).

[80] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus, Journal of Chemical Information and Modeling **59**, 4814 (2019).

Appendix A: Kohn-Sham DFT in 1D

In this section, we briefly outline the formulation and implementation of Kohn-Sham DFT[20,70] for a periodic one-dimensional (1D) system. In what follows, x denotes the coordinate in 1D space. The ground state electron density and energy of a system in presence of an external potential $v_{\text{ext}}(x)$ can be obtained by minimizing a total energy functional, $E[\rho(x)]$, which is given by

$$E[\rho(x)] = T_{\text{s}}[\rho(x)] + E_{\text{ext}}[\rho(x)] + E_{\text{H}}[\rho(x)] + E_{\text{XC}}[\rho(x)]. \tag{A1}$$

Here $T_{\text{s}}[\rho(x)]$ is the non-interacting kinetic energy of electrons, $E_{\text{ext}}[\rho(x)]$ is the external potential energy due to $v_{\text{ext}}(x)$, $E_{\text{H}}[\rho(x)]$ is the Hartree electron-electron repulsion energy and $E_{\text{XC}}[\rho(x)]$ is the exchange correlation energy. In Kohn-Sham formulation, the kinetic energy density is obtained by solving the following eigenvalue problem for a non-interacting system:

$$\begin{aligned}
&\left\{-\frac{1}{2}\frac{\partial^2}{\partial x^2} + v_{\text{ext}}(x) + v_{\text{H}}[\rho](x) + v_{\text{XC}}[\rho](x)\right\}\psi_j(x,k) = \\
&\varepsilon_{j,k}\psi_j(x,k), \\
&\rho(x) = \sum_{j=1}^{N_s} \fint f_j |\psi_j(x,k)|^2 dk, \quad f_j \in \{1,2\}, \\
&\psi_j(x,k) = e^{ikx} u_j(x,k),
\end{aligned} \tag{A2}$$

where $v_{\text{ext}}(x)$ is external potential, $v_{\text{H}}[\rho](x)$ is the Hartree potential, $v_{\text{XC}}[\rho](x)$ is the exchange-correlation potential, ψ_j and $\varepsilon_{j,k}$ are the Kohn-Sham orbitals and energy levels, N_s is the total number of occupied orbitals, f_j is their occupation number, $\fint f(k)dk$ denotes the average integral of some function $f(k)$ over the Brillouin zone, and $u_j(x,k)$ is some periodic function. Orbitals $\psi's$ are also a function of reciprocal space coordinate k in first Brillouin zone due to Bloch's theorem[17,71] as shown in the above equation. For one-dimensional model systems, we replace the Hartree potential and external nuclear potential by a soft form:[39,72]

$$\begin{aligned}
v_{\text{H}}[\rho](x) &= \int_{\mathbf{R}} \rho(x') \frac{dx'}{\sqrt{1+|x-x'|^2}}, \\
v_{\text{ext}}(x) &= \sum_I \frac{-Z_I}{\sqrt{1+|x-X_I|^2}}.
\end{aligned} \tag{A3}$$

Here, Z_I is the nuclear charge of the I^{th} atom. For our analysis of the kinetic energy functional, we use local density approximation (LDA) functional for the exchange-correlation potential proposed in Ref. [73]. The exchange potential is

$$v_{\text{XC}} = -\frac{A}{\pi}\tan^{-1}\left(\frac{\pi\rho}{2\kappa}\right). \tag{A4}$$

The Kohn-Sham equation in Eq. A2 can be formulated as a fixed-point iteration problem[74] $\rho = F(\rho)$, which can be solved self-consistently using an appropriate mixing scheme, such as, the Pulay mixing scheme[74]. We follow similar higher order finite difference strategy outlined in Ref. [75] to numerically solve the Kohn-Sham equations. Further details of our implementation are described in Appendix I.

Appendix B: Implementation

We solve the Kohn-Sham equation numerically using a finite difference discretization method in which a supercell with an edge length of L is discretized using an uniform finite-difference grid of spacing h. Grid points are indexed from 1 to N, where $N = L/h$. The second derivative in Eq. A2 is approximated by using a central finite-difference approximation[17,76]

$$\left.\frac{\partial^2 f}{\partial x^2}\right|^{(i)} = \sum_{k=0}^{N_{FD}} c_k \left(f^{(i+k)} + f^{(i-k)}\right) + O(h^{2N_{FD}}), \tag{B1}$$

where $f^{(i)}$ denotes the value of the function at i^{th} grid point. The weights c_k are[17]:

$$\begin{aligned}
c_0 &= -\frac{1}{h^2}\sum_{m=1}^{N_{FD}}\frac{1}{m^2}, \\
c_k &= \frac{2(-1)^{k+1}}{h^2 k^2}\frac{(N_{FD}!)^2}{(N_{FD}-k)!(N_{FD}+k)!}, \quad k = 1,2,...,N_{FD}.
\end{aligned} \tag{B2}$$

Here N_{FD} is the order of finite difference approximation. The Hartree and the external potentials in Eq. D2 can be obtained by summing over contributions from all periodic images. The mean integral in Eq. A2 over the Brillouin zone is approximated by finite k-points sampled uniformly[17]

$$\fint_{\text{BZ}} f(k)dk \approx \sum_{b=1}^{N_k} w_b f(k_b) \tag{B3}$$

where, k_b and w_b are the nodes and weights for the integration. With the above discretization scheme, Kohn-Sham equation in Eq. A2 can be written in terms of matrix eigenvalue problem:

$$\begin{aligned}
&H\psi = \varepsilon\psi, \quad H \in \mathbb{R}^{N\times N}, \psi \in \mathbb{R}^{N\times 1}, \\
&H^{(i,j)} = -\frac{1}{2}\nabla_h^2|^{(i,j)} + \delta_{ij}\left(v_H^{(j)}[\rho] + v_{ext}^{(j)} + v_{XC}^{(j)}[\rho]\right), \\
&\rho^{(i)} = \sum_{j=1}^{N_s}\sum_{b=1}^{N_k} w_b f_j \bar{\psi}_{j,b}^{(i)} \psi_{j,b}^{(i)}.
\end{aligned} \tag{B4}$$

The eigenvalue problem outlined above in Eq. B4 is solved in Matlab using the inbuilt *eig* solver. With a given initial guess of electron density, these equations are solved using self consistent iterations until convergence to desired accuracy. The kinetic energy and kinetic energy density are ob-

tained using the converged orbitals as

$$T_{\rm s}[\rho] = \sum_{i=1}^{N}\sum_{j=1}^{N_s}\sum_{b=1}^{N_k} h w_b \bar{\psi}_{j,b}^{(i)} \nabla_{h,b}^2 \psi_{j,b}^{(i)},$$
$$t_{\rm s}^{(i)}[\rho] = \sum_{j=1}^{N_s}\sum_{b=1}^{N_k} w_b \bar{\psi}_{j,b}^{(i)} \nabla_{h,b}^2 \psi_{j,b}^{(i)}. \tag{B5}$$

Here $T_{\rm s}[\rho] = \int t_{\rm s}[\rho](x)dx$ and $t_{\rm s}[\rho](x)$ are total kinetic energy and kinetic energy density respectively, $\bar{\psi}$ is the complex conjugate of ψ, ∇_h^2 is the finite-difference discrete second derivative matrix.

Appendix C: Results for modified Q_1 model

Here, we modify the Q_1 model in Eq. 7 such that it captures the correct physics in the uniform gas limit. The model proposed in section II B follows the uniform coordinate scaling relation. Hence, we modify the Q_1 model proposed before by changing the exponents α and β to $4/3$ so that it has correct asymptotic behavior in the uniform gas limit. We call the modified model as $\bar{Q}_1$ which can be written as:

$$\bar{t}_{\rm nl}^{(1)}(x) = \int [\rho(x)]^{3/2} \bar{Q}_1\left(|x-x'|, w(x,x')\right)[\rho(x')]^{3/2} dx',$$
$$w(x,x') = \rho(x) + \rho(x), \tag{C1}$$

where $\bar{Q}_1$ is the modified nonlocal kernel Q_1. We note that the modified model does not obey the uniform coordinate scaling law. We also show the performance of the modified model in terms of root mean square error in kinetic energy density, R^2 score and error in total kinetic energy in table II. We observe that the performance of the models with modified Q_1 kernel has comparable performance as the coordinate scaling compliant model.

Appendix D: Kohn-Sham DFT in 3D

In this section, we briefly outline the formulation and implementation of Kohn-Sham DFT[20,70] for a three-dimensional (3D) system. To generate data for 3D systems, we solve the Kohn-Sham equations using a set of even-tempered Gaussian basis functions and all terms in the standard Kohn-Sham Hamiltonian are calculated analytically using expressions given in Ref. [77,78]. Therefore, to obtain the ground state electron density, we solve the following eigenvalue problem

$$\mathbf{Hc} = \lambda \mathbf{Sc}, \quad \mathbf{H} = \int \psi(\mathbf{r}) H_{\rm KS} \psi(\mathbf{r}) d\mathbf{r},$$
$$\mathbf{S} = \int \psi(\mathbf{r}) \psi(\mathbf{r}) d\mathbf{r}, \quad \psi(\mathbf{r}) = \sum_p c_p \phi_p(\mathbf{r}, \bar{\mathbf{r}}_p) \tag{D1}$$

Here $\phi_1(\mathbf{r})$, $\phi_2(\mathbf{r})$, ... are a set of even-tempered Gaussian basis functions, $\bar{\mathbf{r}}_p$ are the centers and $\mathbf{c}$ is a vector that contains all the coefficients. We model the Hartree and external nuclear potentials of the Kohn-Sham Hamiltonian (denoted by $H_{\rm KS}$) by using the Coulomb kernel

$$v_{\rm H}[\rho](\mathbf{r}) = \int_{\mathbf{R}} \rho(\mathbf{r}') \frac{d\mathbf{r}'}{|\mathbf{r}-\mathbf{R}_I|}, \; v_{\rm ext}(\mathbf{r}) = \sum_I \frac{-Z_I}{|\mathbf{r}-\mathbf{R}_I|}. \tag{D2}$$

Here, Z_I is the nuclear charge of an atom located at $\mathbf{R}_I$. We use the local density approximation (LDA) functional for the exchange potential. The eigenvalue problem in Eq. D1 is solved iteratively and the electron density is updated using a simple mixing scheme with mixing coefficient set to 0.10. The widths of the Gaussian basis functions for helium dimer are set to those in STO-4G basis set (Ref. [79,80]). However, in case of lithium, kinetic energy densities obtained from DFT calculations performed using this basis set shows significant scatter at high electron densities. This scatter in the kinetic energy densities close to atom centers decreases significantly when we use an even-tempered basis set (containing 12 Gaussian functions at each atom position with widths given by 0.05×3^i, i = 0,1,2, ..,11.) for lithium dimer.

Electron densities and kinetic energy densities are obtained using a supercell of size $25\times25\times25$ Bohr^3 and 200 grid points along x, y and z-directions. The coefficients obtained from self-consistent field iterations are used to calculate the kinetic energy and electron densities for a given structure by using the analytical form of the wavefunction given in Eq. $D1$:

$$\rho(r_{ijk}) = \sum_n^{\rm occupied} 2|\psi_n(\mathbf{r})|^2 = \sum_n^{\rm occupied} 2\left[\sum_p c_p^{(n)} \phi_p(r_{ijk}, \bar{\mathbf{r}}_p)\right]^2,$$
$$t(r_{ijk}) = \sum_n^{\rm occupied} \left[\sum_p c_p^{(n)} \phi_p(r_{ijk}, \bar{\mathbf{r}}_p)\right] \Delta \left[\sum_p c_p^{(n)} \phi_p(r_{ijk}, \bar{\mathbf{r}}_p)\right]^2. \tag{D3}$$

Here $c_p^{(n)}$ is the coefficient for the n-th eigenfunction of $\mathbf{H}$, r_{ijk} is a grid point and Δ denotes the Laplacian.

Systems	Parameters	Local	$\bar{Q}_1$	Q_2	Q_3	$\bar{Q}_1+Q_2$	Q_2+Q_3	$\bar{Q}_1+Q_3$	$\bar{Q}_1+Q_2+Q_3$
H_2	R^2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	RMSE	7.7E-05	4.5E-05	5.9E-05	4.0E-05	2.9E-05	2.6E-05	3.8E-05	1.9E-04
	T_{KSDFT}	0.1851	0.1011	0.1011	0.1011	0.1011	0.1011	0.1011	0.1011
	T_{model}	0.1841	0.1023	0.0999	0.1018	0.1019	0.1008	0.1015	0.0992
	T_{error}	0.0010	0.0012	0.0012	0.0007	0.0008	0.0003	0.0005	0.0018
H_8	R^2	0.9847	0.9970	0.9902	0.9850	0.9982	0.9929	0.9984	0.9874
	RMSE	0.0042	0.0020	0.0037	0.0041	0.0015	0.0031	0.0014	0.0040
	T_{KSDFT}	0.3550	0.3550	0.3560	0.3550	0.4315	0.3560	0.3558	0.3550
	T_{model}	0.2591	0.3819	0.3418	0.2475	0.4145	0.3451	0.3684	0.2771
	T_{error}	0.0959	0.0269	0.0141	0.1075	0.0170	0.0108	0.0126	0.0778
LiH	R^2	0.9846	0.9973	0.9955	0.9914	0.9996	0.9968	0.9989	0.9915
	RMSE	0.0073	0.0030	0.0039	0.0055	0.0012	0.0033	0.0019	0.0054
	T_{KSDFT}	0.5403	0.5403	0.5403	0.5403	0.5403	0.5403	0.5403	0.5403
	T_{model}	0.4817	0.5180	0.5574	0.4436	0.4940	0.5261	0.4885	0.4785
	T_{error}	0.0587	0.0224	0.0170	0.0967	0.0463	0.0142	0.0518	0.0618
LiF	R^2	0.9988	0.9995	0.9994	0.9990	0.9995	0.9994	0.9997	0.9990
	RMSE	0.0050	0.0032	0.0035	0.0045	0.0031	0.0034	0.0024	0.0045
	T_{KSDFT}	2.7035	2.7035	2.4304	2.4304	2.4304	2.4304	2.7055	2.5385
	T_{model}	2.7145	2.6747	2.4079	2.4220	2.4437	2.4185	2.6979	2.5573
	T_{error}	0.0111	0.0288	0.0225	0.0083	0.0133	0.0119	0.0076	0.0188

TABLE II. A comparison of kinetic energy densities and total kinetic energies between model predictions and KSDFT results for structures in test data sets. For each system, 3 structures are used for training and 27 structures are used for testing. Here R^2 and RMSE are the coefficient of determination and the root mean square error of kinetic energy density (Ha/Bohr), respectively, T_{KSDFT} and T_{Model} are the total kinetic energies (Ha) obtained from KSDFT calculations and the various models, respectively, and T_{error} is the error in total kinetic energy (Ha) between the KSDFT value and the model. For each model type, total kinetic energies (T_{KE}, T_{model} and T_{error}) only for the structure with the maximum absolute error (out of the 27 structures used for testing) are shown. $\bar{Q}_1$, Q_2 and Q_3 correspond to the nonlocal models in Eqs. C1, 11 and 15, respectively.

Supplementary Materials: Accurate parameterization of the kinetic energy functional

Shashikant Kumar,[1,2] Edgar Landinez Borda,[3] Babak Sadigh,[1] Siya Zhu,[1,4] Sebastian Hamel,[1] Brian Gallagher,[5] Vasily Bulatov,[6] John Klepeis,[1] and Amit Samanta[1, a)]

[1)] *Physics Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

[2)] *College of Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

[3)] *Chemistry Department, Brown University, Providence, Rhode Island 02912, USA*

[4)] *School of Engineering, Brown University, Providence, RI 02912, USA*

[5)] *Applications, Simulations and Quality Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

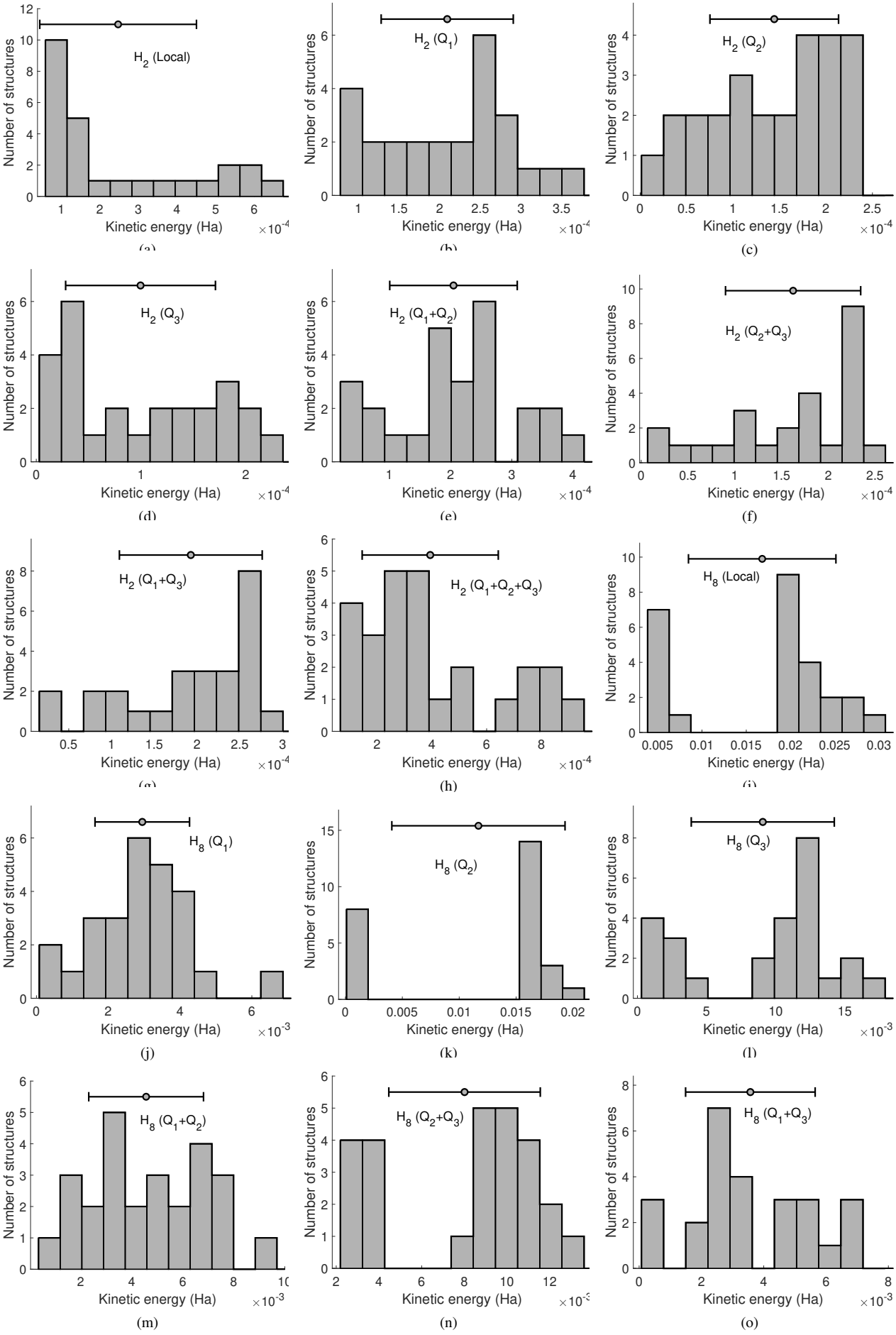
[6)] *Materials Science Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

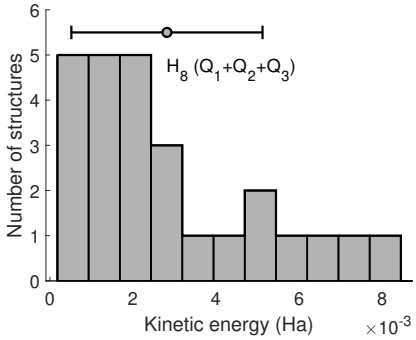
(Dated: 3 November 2021)

[a)]Electronic mail: samanta1@llnl.gov

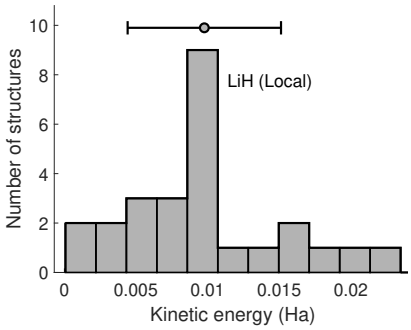
Systems	Parameters	Local	Q_1	Q_2	Q_3	Q_1+Q_2	Q_2+Q_3	Q_1+Q_3	$Q_1+Q_2+Q_3$
H_2	c_1	0.0054	-0.0018	0.0095	0.0026	0.0034	0.0029	0.0048	0.0362
	c_2	-0.0417	-0.0579	-0.0712	-0.0049	-0.0233	-0.0109	-0.0080	-0.0219
	c_3	0.0729	0.0435	0.0529	10.0271	0.0053	-0.0141	-0.0112	0.0120
	c_4	0.0088	0.0142	0.0597	0.0378	0.0273	0.0258	0.0157	0.0006
	c_5	-0.0158	0.0076	0.0387	0.0401	0.0191	0.0306	0.0198	0.0027
	c_6	-0.0165	0.0047	0.0203	0.0244	0.0107	0.0209	0.0135	0.0020
	c_7	0.1250	0.1309	0.1256	0.1286	0.1281	0.1268	0.1298	0.1325
	c_8	-0.2524	-0.2579	-0.2531	-0.2526	-0.2561	-0.2524	-0.2552	-0.2610
H_8	c_1	0.1155	0.6306	-0.0190	0.0834	0.8079	-0.0506	0.4184	-0.0919
	c_2	-0.3779	-1.2120	0.6357	0.0364	-1.0756	0.8832	-0.2055	-0.7537
	c_3	0.1063	1.0184	-2.1099	-1.0408	0.8180	-2.8271	-0.16401	0.8353
	c_4	0.9551	0.6316	1.1807	1.1273	0.3865	1.7117	0.5557	0.6457
	c_5	-0.4107	-0.5368	1.6168	0.5912	-0.3263	2.2320	0.1841	-0.2110
	c_6	-1.312	-0.9994	0.9905	-0.2452	-0.5860	1.3226	-0.1989	-0.6192
	c_7	0.0876	0.0152	0.1208	0.4493	-0.0152	0.16601	0.0834	0.1587
	c_8	-0.2413	-0.1966	-0.2236	-0.2422	-0.1940	-0.1804	-0.1912	-0.1624
LiH	c_1	-0.0973	-0.5668	-0.0996	-0.1805	-3.0154	-0.1912	0.0524	-2.0184
	c_2	0.6679	1.4608	1.0914	1.4051	-2.0306	1.6433	-0.1786	-1.0939
	c_3	-1.2626	-2.4629	-2.6248	-3.1637	2.9282	-4.4005	1.1139	1.9321
	c_4	0.4823	1.3787	1.9897	1.8827	0.0832	4.3226	-0.5579	0.3806
	c_5	3.7629	2.0615	2.7009	4.4846	-1.5788	2.3514	-0.2158	-1.0608
	c_6	-3.7048	-2.0919	-3.0603	-4.6344	0.2793	-3.8335	-0.0470	-0.1188
	c_7	0.0613	0.0718	0.0492	0.1705	0.0571	0.0804	0.1014	0.2074
	c_8	-0.1407	-0.1940	-0.1544	-0.1501	-0.1835	-0.1238	-0.2354	-0.1824
LiF	c_1	0.0690	-1.3735	0.0100	-0.0214	0.0721	-0.0175	-0.2466	0.4540
	c_2	-0.8073	-0.4971	-0.1785	0.3125	0.0094	0.2454	-0.0428	-0.1961
	c_3	3.3079	2.8944	1.4291	-0.7444	0.9409	-0.8186	0.0113	0.6546
	c_4	-4.9538	-4.5818	-2.6406	1.0195	-1.5722	2.0805	1.4844	-0.1402
	c_5	4.1186	4.1206	2.9774	0.3557	1.8270	-1.3394	-1.4675	0.0889
	c_6	-1.3232	-1.4200	-1.1875	-0.5080	-0.7418	0.2617	0.4554	-0.0568
	c_7	0.0388	0.0121	0.0267	1.1036	-0.0083	0.2629	0.7265	1.4591
	c_8	-0.0845	-0.1124	-0.1163	-0.0665	-0.1099	-0.0905	-0.0768	-0.0947

TABLE I. Coefficients of the local terms which includes the gradient expansion term and polynomial terms in all 8 models for the 4 chosen structures. The coefficients c_j correspond to the following equation: $t(x) = \sum_{j=1}^{6} c_j \rho^j(x) + c_7 \frac{(\rho'(x))^2}{\rho(x)} + c_8 \rho''(x) + t^{nl}[\rho](x)$

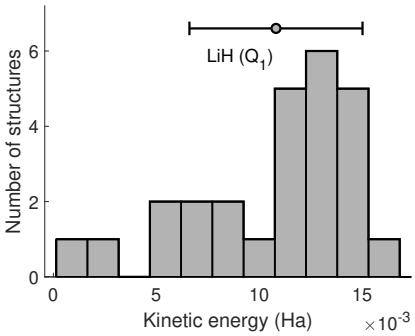

H2 (Local)
H2 (Q1)
H2 (Q2)
H2 (Q3)
H2 (Q1+Q2)
H2 (Q2+Q3)
H2 (Q1+Q3)
H2 (Q1+Q2+Q3)
H8 (Local)
H8 (Q1)
H8 (Q2)
H8 (Q3)
H8 (Q1+Q2)
H8 (Q2+Q3)
H8 (Q1+Q3)
Number of structures
Kinetic energy (Ha)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)



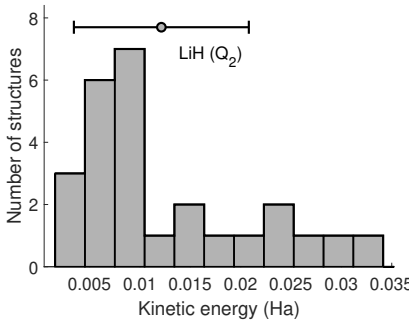
(p)



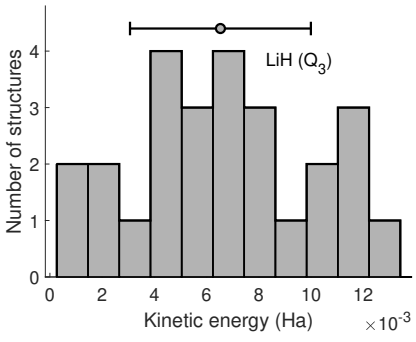
(q)



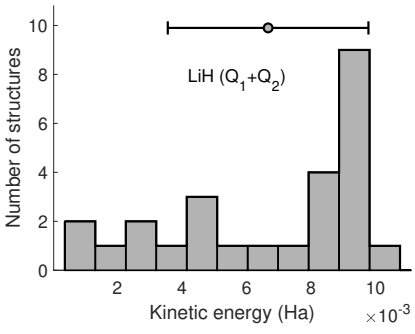
(r)



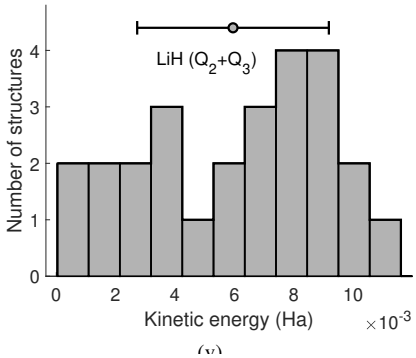
(s)



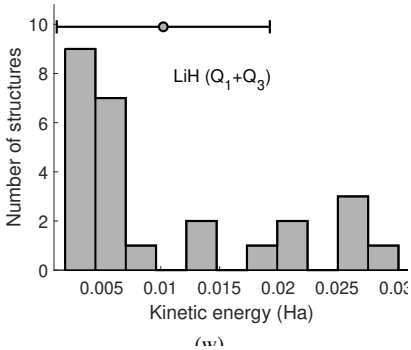
(t)



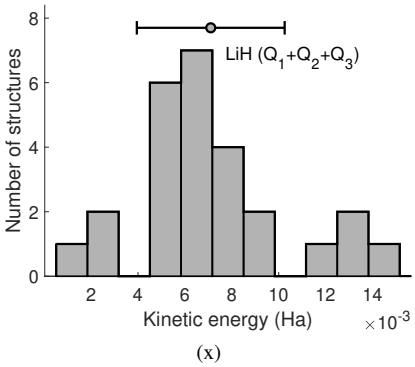
(u)



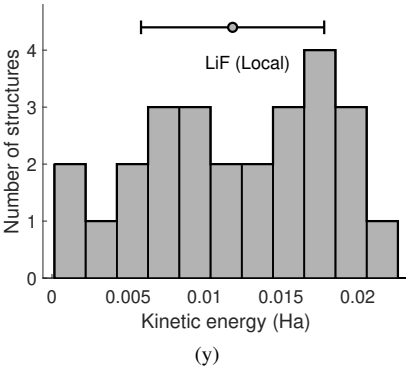
(v)



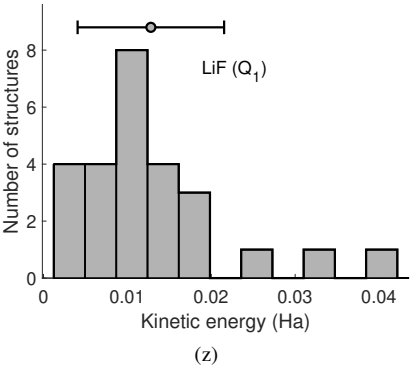
(w)



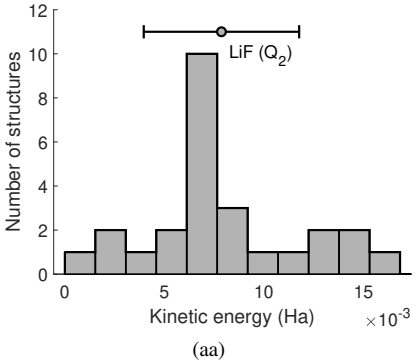
(x)



(y)



(z)



(aa)

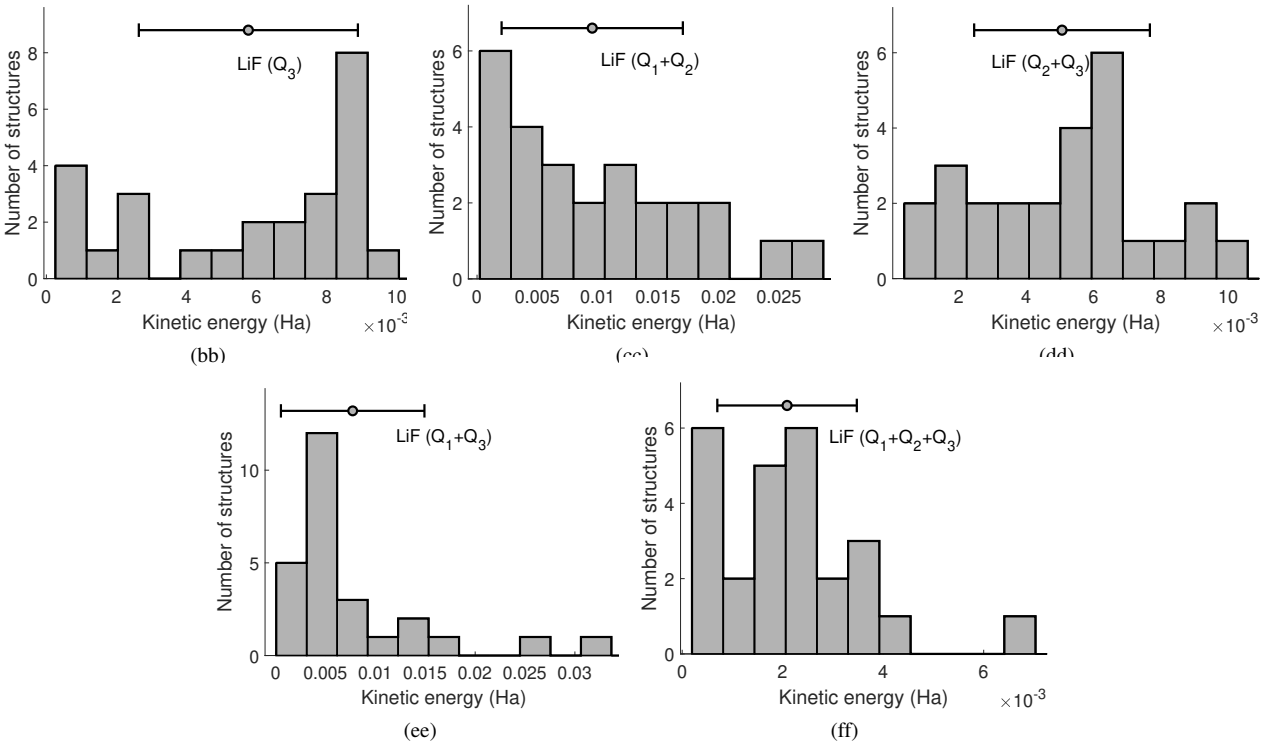


FIG. 1. Shown here are the distributions of errors in kinetic energies (Ha) predicted using local and non-local models for all structures in the test datasets of H_2, LiH, LiF and H_8. The horizontal bar on the top of each plot shows the mean (marked with a circle) and standard deviation of the errors in predicted kinetic energies.