



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Accurate parameterization of the kinetic energy functional for calculations using exact exchange

S. Kumar, B. Sadigh, S. Zhu, P. Suryanarayana, S. Hamel, B. Gallagher, V. Bulatov, J. Klepeis, A. Samanta

September 30, 2021

Journal of Chemical Physics

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Accurate parameterization of the kinetic energy functional for calculations using exact exchange

Shashikant Kumar,^{1,2} Babak Sadigh,¹ Siya Zhu,^{1,3} Phanish Suryanarayana,² Sebastian Hamel,¹ Brian Gallagher,⁴ Vasily Bulatov,⁵ 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⁾*School of Engineering, Brown University, Providence, RI 02912, USA*

⁴⁾*Applications, Simulations and Quality Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

⁵⁾*Materials Science Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA*

(Dated: 1 October 2021)

Electronic structure calculations based on Kohn-Sham density functional theory (KSDFT) that incorporate exact exchange or hybrid functionals are associated with large computational expense, a consequence of the inherent cubic scaling bottleneck and large associated prefactor, which limits the length and time scales that can be accessed. Though orbital-free density functional theory (OFDFT) calculations scale linearly with system size and are associated with significantly smaller prefactor, they are limited by the absence of accurate density-dependent kinetic energy functionals. Therefore, the development of accurate density-dependent kinetic energy functionals is important for OFDFT calculations of large realistic systems. To this end, we propose a method to train kinetic energy functional models at the exact-exchange level of theory by using a dictionary of physically relevant terms that have been proposed in the literature in conjunction with linear or nonlinear regression methods to obtain the fitting coefficients. For our dictionary, we use gradient expansion of the kinetic energy, nonlocal models proposed in the literature and their nonlinear combinations, such as a model that incorporates spatial correlations between higher order derivatives of electron density at two points. The predictive capabilities of these models are assessed by using a variety of model one-dimensional systems that exhibit diverse bonding characteristics, such as a chain of eight hydrogens, LiF, LiH, C₄H₂, C₄N₂ and C₃O₂. We show that by using data from model one-dimensional KSDFT calculations performed using the exact-exchange functional for only a few neutral structures, it is possible to generate models with high accuracy for charged systems as well as electron and kinetic energy densities during self-consistent field iterations. In addition, we show that it possible to learn both the orbital dependent terms, i.e., the kinetic energy and the exact-exchange energy, and models that incorporate additional nonlinearities in spatial correlations, such as a quadratic model, are needed to capture subtle features of the kinetic energy density that are present in exact-exchange-based KSDFT calculations.

I. INTRODUCTION

Kohn-Sham density functional theory (KSDFT)^{1,2} is one of the most popular and powerful techniques to perform electronic structure calculations³⁻⁶, with recent advances allowing for simulations of systems containing thousands of atoms⁷⁻⁹. The main limitations of KSDFT are its cubic scaling and a large prefactor associated with the underlying electronic-eigenvalue problem due to which it cannot be applied to many large systems that are encountered in various materials science and physics applications. In contrast, orbital-free density functional theory scales linearly with the number of atoms and has a significantly smaller prefactor, which opens the possibility for its application to systems containing as many as a million atoms^{10,11}, thereby enabling us to handle realistic systems that are currently not accessible in KSDFT simulations¹²⁻¹⁴. However, as discussed in literature (e.g. [15-19]), one of the main challenges in OFDFT calculations is that existing kinetic energy functionals are not accurate enough to capture subtle

features of the kinetic energy density obtained from a KSDFT calculation.

Early efforts to derive such a functional focused primarily on using gradient expansions of the kinetic energy²⁰⁻²⁵, but such expansions are typically good only under limiting conditions (for example, homogeneous systems). Therefore, for many years, researchers have focused on approximating the kinetic energy functional by using a nonlocal kernel, to capture spatial correlations in the density, along with the terms present in gradient expansion based models^{26-31,33-36}. However, these nonlocal kernels are also valid under limiting conditions (for example, homogeneous periodic systems) meaning that models containing these nonlocal correlations may not be accurate for general systems. Many researchers have used a different route to design kinetic energy functionals by using the generalized gradient approximation (GGA)^{37,38}. For example, Constantin, et al., have shown that a properly designed Laplacian-level kinetic energy functional can have similar accuracy as that of a nonlocal model^{40,41}. Systematic comparison of many such models have been performed in Refs [32, 38, and 39], and these results suggest that the true kinetic energy functional is perhaps very complex and must include GGA-type functional forms, nonlocal terms as well

^{a)}Electronic mail: samanta1@llnl.gov

as their higher order generalizations.^{27,42}

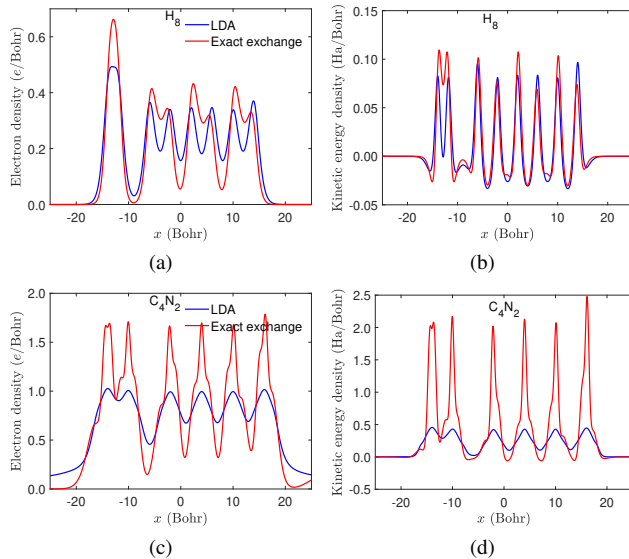


FIG. 1. Plots showing the difference in ground state electron densities and kinetic energy densities of two representative one-dimensional systems, H_8 (a, b) and C_4N_2 (c, d), obtained from model KSDFT calculations using local density approximation (LDA) exchange and exact-exchange functionals.

In the last few years, many researchers have used machine learning techniques to learn kinetic energy functionals directly from KSDFT data^{43–50}. These efforts have resulted in good approximations, but one of the fundamental limitations of data-driven methods is the transferability to systems outside of the training domain. A hybrid model which combines the data-driven learning and physics-based model parameters can be a promising path forward towards developing an accurate and transferable kinetic energy functional model. In line with this philosophy, in Ref. [15] we developed a simple methodology to learn kinetic energy functional by using a dictionary of physically motivated terms and the coefficients were obtained by using linear regression. Among the multiple models explored in that study, the model containing spatial correlations between electron densities and between electron densities and its derivatives (using two different kernels and hereafter denoted as $Q_1 + Q_3$ model) was optimal both in terms of number of parameters and in terms of performance.

In this contribution, we expand the scope of the work proposed in Ref. [15] and assess the performance of models of kinetic energy functionals generated by using data from one-dimensional model KSDFT calculations performed with the exact-exchange functional. Figure 1 shows electron densities and kinetic energy densities of two representative one-dimensional systems, H_8 and C_4N_2 , obtained from one-dimensional model KSDFT calculations performed by using the local density approximation (LDA) exchange and exact-exchange functionals. It is clear that the electron densities and kinetic energy densities obtained from exact-exchange based calculations exhibit features that are absent in LDA exchange-based calculations. Here we note that these subtle features are

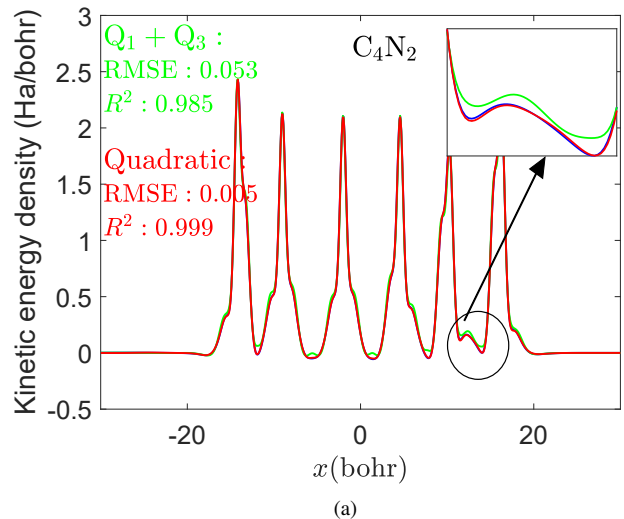


FIG. 2. Kinetic energy densities predicted for a representative one-dimensional system, C_4N_2 , using a model that incorporates spatial correlations of the electron density and its derivative (denoted by $Q_1 + Q_3$ in Ref. [15]) are compared with exact exchange based calculations. The insets show the deviations of the model from direct calculations. Predictions made by using a quadratic model (see details in Section II) proposed in this work are in better agreement with KSDFT results.

related to the fact that the exchange operator is full-rank because it incorporates nonlocal correlations between different wavefunctions of the Hamiltonian. Thus, it is natural to ask if the models proposed in Ref. [15] can reliably capture subtle features present in the kinetic energy density profiles. In addition, a density functional model for kinetic energy alone is not sufficient to run OFDFT calculations with exact exchange level of accuracy. This is because the exact exchange energy itself depends on the wavefunctions. Hence, to run an OFDFT calculation with the exact-exchange level of accuracy, we need a density functional of both the kinetic energy and the exact exchange energy. In that regard, the second objective of this study is to explore the effectiveness of the proposed models to describe a modified kinetic energy functional which is obtain by the sum of non-interacting kinetic energy and the exact exchange energy minus the LDA exchange energy.

Another goal of this study is to explore the possibility of using these empirical models to train generic or universal kinetic energy models for a large number of heterogeneous systems. Such a generic model will allow us to perform large-scale OFDFT calculations. For example, Fig. 2 shows that a generic model trained by using structures from a variety of systems (further details are given in Section III A) using the model ($Q_1 + Q_3$) proposed in our previous work (Ref 15) is able to capture the important features of kinetic energy density but the R^2 score (i.e. the coefficient of determination) is only 0.985 and deviations from the ground truth are clearly evident. To this end, it is instructive to explore different ways to enhance the predictive capabilities of such generic models.

With these objectives in mind, we propose physics-inspired

empirical models of the kinetic energy functional by including spatial correlations between higher order derivatives of the electron density. As discussed in Section II, we start with a dictionary of functional forms that have been proposed in the literature to capture the local and nonlocal contributions, and use a basis function expansion to model the nonlocal kernel. To illustrate the effectiveness of our models, in Section III, we use a variety of heterogeneous one-dimensional systems, such as LiF, LiH, a one-dimensional chain of eight hydrogens, C₄H₂, C₄N₂ and C₃O₂. These systems exhibit a variety of bonding characteristics, like the presence of single, double and triple bonds, and form a good test set to explore the capabilities of the models. We show that reliable density-dependent models for the kinetic energy can be trained by using data from only a few structures.

II. KINETIC ENERGY FUNCTIONAL MODELS

To model the kinetic energy obtained from solving the model one-dimensional Kohn-Sham equations with the exact-exchange functional we consider the nonlocal models proposed in Ref. [15] and also generalize these models to incorporate spatial correlations and higher order nonlinearities. Details of the model one-dimensional Kohn-Sham equations for exact-exchange are given in Appendix I.

Equation 1 combines all the models proposed in Ref. [15]. It contains a variety of local terms and a few nonlocal terms: The local terms in the model are the first three terms in the gradient expansion of the kinetic energy functional for a one-dimensional system derived by Samaj et al²⁰. To model the nonlocal contributions, three different nonlocal kernels are used and these nonlocal kernels are modeled by using a set of basis functions.

$$\begin{aligned}
t(x) &= t_{\text{local}}(x, \mathbf{c}) \\
&+ \sum_{j=1}^{N_G} d_{1,j} [\rho(x)]^2 \int e^{-\beta_{1,j}|x-x'|^2} [u(x,x')]^2 [\rho(x')]^2 dx' + \\
&\sum_{j=1}^{N_G} d_{2,j} [\rho(x)]^2 \int e^{-\beta_{2,j}|x-x'|^2} [u(x,x')]^2 (\rho(x) - \rho(x'))^2 dx' \\
&+ \sum_{j=1}^{N_G} d_{3,j} \nabla_x \rho(x) \int e^{-\beta_{3,j}|x-x'|^2} [u(x,x')]^2 \nabla_{x'} \rho(x') dx',
\end{aligned} \tag{1}$$

$$t_{\text{local}}(x, \mathbf{c}) = c_1 \rho^3(x) + c_2 \frac{[\nabla \rho(x)]^2}{\rho(x)} + c_3 \nabla^2 \rho(x)$$

$$u(x, x') = \rho(x) + \rho(x').$$

Here, c_j and $d_{i,j}$ are the coefficients of the model, $\beta_{i,j}$ are the widths of the Gaussian functions, and N_G is the number of Gaussian basis functions used to model the nonlocal kernel.

The three terms present in the local kinetic energy expression, t_{local} , are based on gradient-expansion and satisfy the uniform coordinate scaling constraint⁵¹⁻⁵⁴. The kernels in the nonlocal terms are dimensionless quantities and the exponents of density and gradient of density have been selected to satisfy coordinate scaling requirements. The optimal set of coefficients is

obtained using linear regression by fitting the model to kinetic energy density data obtained by solving the eigenvalue problem of the Kohn-Sham Hamiltonian using the exact-exchange functional. A detailed description of the numerical implementation of model KSDFT with exact-exchange functional is given in Appendix II. To calculate the coefficients, we optimize the following objective function

$$\Omega = |\mathbf{t}_{\text{KSDFT}} - \mathbf{A}\mathbf{c}|^2, \tag{2}$$

where, $\mathbf{t}_{\text{KSDFT}}$ is a vector that contains information about kinetic energy densities at different grid points, $\mathbf{c} = \{c_1, c_2, \dots, d_{1,1}, d_{1,2}, \dots\}$ is a vector that contains the unknown coefficients and \mathbf{A} is the design matrix such that the number of rows is equal to the number of grid points used for fitting (i.e. length of $\mathbf{t}_{\text{KSDFT}}$), and the number of columns is equal to the number of terms used to model the kinetic energy functional. Therefore, each row of \mathbf{A} contains values of the terms shown in Eq. 1 at a grid point and for a given set of values of the widths of the Gaussian functions, the vector \mathbf{c} is obtained by solving a linear system. The widths of the Gaussian kernels are optimized by using the Nelder-Mead method.

In the following sub-sections, we propose a few other nonlocal models to approximate the kinetic energy density. These models are different from the models discussed above and are also inspired by physic-based models proposed in the literature. However, parameters present in these new nonlocal models cannot be obtained by using a linear regression framework and a nonlinear minimization method has to be employed to learn the coefficients.

A. A nonlocal model inspired by a model proposed by Chacón et al

Chacón et al. [26] introduced the notion of a locally averaged density, $\bar{\rho}(\mathbf{r})$, to incorporate spatial correlations into the kinetic energy functional:

$$\bar{\rho}(\mathbf{r}) = \int \rho(\mathbf{r}') W(\mathbf{r} - \mathbf{r}', \rho(\mathbf{r})) d\mathbf{r}', \text{ and } \int W(\mathbf{r}, \rho(\mathbf{r})) d\mathbf{r} = 1. \tag{3}$$

Using this averaged electron density, the authors proposed the following form to model the kinetic energy

$$T[\rho] = T_{\text{vW}}[\rho] - \frac{1}{3} T_{\text{LDA}}[\rho] + \frac{4}{3} \int \rho(\mathbf{r}) t(\bar{\rho}(\mathbf{r})) d\mathbf{r}. \tag{4}$$

Here, $T_{\text{vW}}[\rho]$ is the von Weizsäcker term, $T_{\text{LDA}}[\rho]$ is the Thomas-Fermi term of kinetic energy functional and t is a function of density such that $T_{\text{LDA}}[\rho] = \int \rho(\mathbf{r}) t(\rho(\mathbf{r})) d\mathbf{r}$. Intuitively, this model can be generalized by using models proposed in Ref. [15] for one-dimensional systems. For example, if we focus on the local terms (Eq. 1), then a probable model

for the kinetic energy is

$$T[\rho] = \int (t_{\text{local}}(\rho(x), \mathbf{c}) + \bar{t}_{\text{local}}(\bar{\rho}(x), \rho(x), \bar{\mathbf{c}})) dx,$$

$$\bar{t}_{\text{local}}(\bar{\rho}(x), \rho(x), \bar{\mathbf{c}}) = \bar{c}_2 \bar{\rho}^3(x) + \bar{c}_2 \frac{[\nabla \bar{\rho}(x) \cdot \nabla \rho(x)]}{\rho(x)} + \bar{c}_3 \nabla^2 \bar{\rho}(x),$$

$$\bar{\rho}(x) = \frac{\int \rho(x') \left[\sum_{p=1}^{N_g} d_p e^{-|x-x'|^2 (\rho(x) + \rho(x'))^2} \right] dx'}{\int \left[\sum_{p=1}^{N_g} d_p e^{-|x-x'|^2 (\rho(x) + \rho(x'))^2} \right] dx'}, \quad (5)$$

where d_p for $p = 1, 2, 3, \dots$ are the coefficients used to model the nonlocal kernel. Our preliminary analysis suggests that this model is indeed very powerful, but due to the nonlinearities embedded in the higher order polynomial terms, such as $\bar{\rho}^2(x)$, $\bar{\rho}^3(x)$, etc., obtaining the best set of fitting parameters becomes challenging. To this end, we simplify this model and consider local averaging of only a few quantities - $\rho(x)$ and $\nabla \rho(x)$ and consider the following nonlocal model:

$$t_{\text{nl}}^{(4)}(x) = t_{\text{local}}(\rho(x), \mathbf{c})$$

$$+ \alpha_1 [\rho(x)]^2 \int Q_4(|x-x'|, u(x, x')) [\rho(x')]^2 dx' \quad (6)$$

$$+ \alpha_2 \nabla_x \rho(x) \int Q_4(|x-x'|, u(x, x')) \nabla_{x'} \rho(x') dx', \text{ and}$$

$$u(x, x') = \rho(x) + \rho(x').$$

Here α_1 and α_2 are two parameters introduced to make the two terms dimensionally consistent. **We note that the exponent of 2 in the first nonlocal term is required to satisfy the coordinate scaling criterion.** The model described above is similar to the model containing two nonlocal kernels (with spatial correlations in the electron density and the derivative of the electron density) as discussed in Ref. [15] except that in Eq. 6 the same kernel is used to capture spatial correlations between electron density and its derivative. We model the nonlocal kernel Q_4 by using a set of basis functions:

$$Q_4(|x-x'|, u(x, x')) = \sum_{k=1}^{N_g} d_k e^{-\beta_k |x-x'|^2 [u(x, x')]^2}. \quad (7)$$

Interestingly, the model is linear with respect to all the coefficients, however, there is a coupling between the coefficients d_k and the parameters α_1, α_2 . Thus, we adopt a two-step iterative optimization scheme to obtain the optimal set of coefficients: In the first step, we use linear regression to solve for \mathbf{c} and d_k while keeping the parameters α_1, α_2 constant. In the second step, we fix d_k and solve another linear system to obtain \mathbf{c} and α_1, α_2 . This two-step scheme is repeated until convergence.

B. A quadratic model

The model proposed in the previous section includes spatial correlations between electron densities and between derivatives of the electron densities at two points. To further understand the relevance of spatial correlations between higher

order derivatives of the electron density, we consider a model that is quadratic in terms of $t_{\text{local}}(\rho(x), \mathbf{c})$:

$$t_{\text{nl}}^{(5)}(x) = t_{\text{local}}(\rho(x), \mathbf{c})$$

$$+ t_{\text{local}}(\rho(x), \bar{\mathbf{c}}) \int Q_5(|x-x'|, u(x, x')) t_{\text{local}}(\rho(x'), \bar{\mathbf{c}}) dx',$$

$$t_{\text{local}}(\rho(x), \mathbf{c}) = c_1 \rho^3(x) + c_2 \frac{[\nabla \rho(x)]^2}{\rho(x)} + c_3 \nabla^2 \rho(x). \quad (8)$$

Here, Q_5 is a nonlocal kernel which is modeled using a set of Gaussian basis functions as discussed in Section II.A, and $t_{\text{local}}(\rho(x), \mathbf{c})$ contains terms that depend on the density and its derivatives at x . **We also note that the nonlocal part of $t_{\text{nl}}^{(5)}(x)$ does not satisfy the coordinate scaling law, however the local part of the model satisfies it. Therefore, to design a quadratic model that satisfies the coordinate scaling condition, we introduce the dimensionless variables p and q that are commonly used in GGA models of kinetic energy functional^{37,38}. The expression of the modified quadratic model in 1D is:**

$$t_{\text{nl}}^{(5, \text{scaling})}(x) = t_{\text{local}}(\rho(x), \mathbf{c})$$

$$+ \bar{t}(\rho(x), \bar{\mathbf{c}}) \int Q_5^{\text{scaling}}(|x-x'|, u(x, x')) \bar{t}(\rho(x'), \bar{\mathbf{c}}) dx',$$

$$t_{\text{local}}(\rho(x), \mathbf{c}) = c_1 \rho^3(x) + c_2 \frac{[\nabla \rho(x)]^2}{\rho(x)} + c_3 \nabla^2 \rho(x),$$

$$\bar{t}(\rho(x), \bar{\mathbf{c}}) = (1 + \bar{c}_1 p(x) + \bar{c}_2 q(x)) \rho^2(x),$$

$$p(x) = \frac{[\nabla \rho(x)]^2}{\rho^4(x)}, q(x) = \frac{\nabla^2 \rho(x)}{\rho^3(x)}. \quad (9)$$

We note that the functional form of both $t_{\text{nl}}^{(5)}(x)$ and $t_{\text{nl}}^{(5, \text{scaling})}(x)$ is quadratic $\bar{\mathbf{c}}$ and we adopt a two-step optimization scheme to obtain an optimal set of coefficients: In the first step, we initialize $\bar{\mathbf{c}}$ with random numbers and solve the linear system in \mathbf{c} and \mathbf{d} . In the second step, we optimize the objective function with respect to $\bar{\mathbf{c}}$ by using a trust-region algorithm. These two steps are repeated until convergence.

C. A nonlocal model based on a kinetic energy functional proposed by Liu and Parr

Liu and Parr [55] argued that the Kohn-Sham kinetic energy can be expressed as a functional of degree one in electron density:

$$T[\rho] = -\frac{1}{8} \int \nabla_{\mathbf{r}}^2 \gamma_1(\mathbf{r}, \mathbf{r}') d\mathbf{r}$$

$$= \frac{1}{8} \int \frac{|\nabla_{\mathbf{r}} \rho(\mathbf{r})|^2}{\rho(\mathbf{r})} d\mathbf{r} + \frac{1}{2} \int [\rho(\mathbf{r}) \nabla_{\mathbf{r}} \cdot \nabla_{\mathbf{r}'} P(\mathbf{r}, \mathbf{r}') |_{\mathbf{r}=\mathbf{r}'} + \nabla_{\mathbf{r}} \rho(\mathbf{r}) \cdot \nabla_{\mathbf{r}'} P(\mathbf{r}, \mathbf{r}') |_{\mathbf{r}=\mathbf{r}'}] d\mathbf{r},$$

$$\text{where } P^2(\mathbf{r}, \mathbf{r}') = \frac{|\gamma_1(\mathbf{r}, \mathbf{r}')|^2}{\rho(\mathbf{r}) \rho(\mathbf{r}')}, \quad (10)$$

and $\gamma_1(\mathbf{r}, \mathbf{r}')$ is the first order density matrix associated with the ground state electron density. The authors further suggested that $P(\mathbf{r}, \mathbf{r}')$ is a function of degree zero in $\rho(\mathbf{r})$, but to construct nonlocal models we relax this constraint and model the function $P(\mathbf{r}, \mathbf{r}')$ by using a set of Gaussian functions. Therefore, we use (in 1D)

$$P(x, x') = \sum_{k=1}^{N_G} d_k e^{-\beta_k |x-x'|^2} [u(x, x')]^2, \quad (11)$$

$$u(x, x') = \rho(x) + \rho(x'),$$

where d_k are the coefficients and β_k are the widths of the Gaussian functions used to model the function $P(x, x')$.

It is easy to see that after the derivatives of $P(x, x')$ are evaluated and substituted in Eq. 10, imposing the condition $x = x'$ reduces the kinetic energy to a model that depends on only electron density and its derivatives at x . Therefore, to construct a nonlocal model we do not impose this constraint and instead integrate over x' to incorporate spatial correlations in the electron density and its derivatives. This leads to the following nonlocal model

$$t_{\text{nl}}^{(6)}(x) = t_{\text{local}}(\rho(x), \mathbf{c}) + \int \frac{[\rho(x) \nabla_x \nabla_{x'} P(x, x') + \nabla_x \rho(x) \cdot \nabla_{x'} P(x, x')]}{\int P(x, x') dx'} dx' \quad (12)$$

The additional term in the denominator is introduced to preserve the overall dimensionality of the terms on the right hand side of Eq. 10 and this normalization introduces additional nonlinearity into the model. The local terms are same as in the quadratic model. The analytic expressions for the first and second order derivatives of $P(x, x')$ are given in Appendix III. To evaluate the coefficients and the widths of the Gaussian functions we use the version of the Nelder-Mead algorithm implemented in MATLAB.

III. RESULTS AND DISCUSSIONS

A. DFT data and model training

We use six different one-dimensional systems, namely, H₈, LiH, LiF, C₄N₂, C₄H₂ and C₃O₂ to assess the performance of our models. Model one-dimensional Kohn-Sham calculations (using exact-exchange) are performed for all of these six systems with open boundary conditions and kinetic energy densities and electron densities for 30 different structures for each of these six systems are generated for training and testing purposes. For each system, three structures are randomly chosen to train models and the remaining 27 structures are used for testing the trained model. We use six Gaussian functions to approximate the nonlocal kernels in $t_{\text{nl}}^{(4)}(x)$, $t_{\text{nl}}^{(5)}(x)$, $t_{\text{nl}}^{(5, \text{scaling})}(x)$ and $t_{\text{nl}}^{(6)}(x)$. Combining local and nonlocal terms, there are 9 parameters in the model $t_{\text{nl}}^{(4)}$, 12 parameters in the models $t_{\text{nl}}^{(5)}$ and $t_{\text{nl}}^{(5, \text{scaling})}(x)$, and 9 parameters in the model $t_{\text{nl}}^{(6)}$. To assess the performance of the models proposed in Ref.

[15], for exact-exchange calculations we use the same nomenclature, i.e. the model containing the local terms and the first nonlocal term in Eq. 1 (the second term on the right hand side) is denoted by Q_1 . Similarly, the model containing the local terms and the third (fourth) term in Eq. 1 is denoted by Q_2 (Q_3).

B. Kinetic energy density prediction

The performance of the eight models proposed in Ref. [15] is measured in terms of the coefficient of determination (i.e. R² score), root mean square error (RMSE), and the maximum absolute error in the predicted total kinetic energy is shown in the supplementary material. Interestingly, even for data obtained from KSDFT calculations with exact-exchange functional, kinetic energy models developed for each system have R² scores larger than 0.99 and maximum absolute errors in the predicted kinetic energies are less than 1% of the KSDFT values. For all systems considered in this study, it is clear that inclusion of nonlocal terms leads to improved performance. For example, models containing one nonlocal term have lower RMSE scores than the local models. Similarly, models containing two nonlocal terms have lower RMSE scores than models containing one nonlocal term. However, we do not see an improvement in the predictive capability when all the three nonlocal terms are present in the models. Finally, we find that the models containing Q_1+Q_3 nonlocal kernels perform better than the other models for almost all systems. For some systems, $Q_1 + Q_2 + Q_3$ performs better than $Q_1 + Q_3$, but it also contains significantly larger number of coefficients compared to $Q_1 + Q_3$ and is prone to over-fitting.

In Table I, the performance of the three nonlocal models introduced in this work is summarized. We see that for most of the systems listed in Table I, kinetic energies predicted using the quadratic models have the lowest RMSE scores and lowest maximum absolute errors. For many systems, RMSE scores of the models generated by using $t_{\text{nl}}^{(4)}$ are similar to that of the quadratic models. However, models generated by using $t_{\text{nl}}^{(6)}$ are unable to properly learn the kinetic energy density. We also observe that the quadratic model, $t_{\text{nl}}^{(5)}$, performs best in terms of the error in the total kinetic energy among the three models. **This is interesting because $t_{\text{nl}}^{(5)}$ does not satisfy the coordinate scaling condition and the quadratic model that satisfies the scaling condition has much higher RMSE values. Therefore, for the subsequent analyses we do not consider $t_{\text{nl}}^{(5, \text{scaling})}(x)$.** Here we note that the quadratic models have 12 fitting parameters compared to 9 in case of $t_{\text{nl}}^{(6)}$ which can potentially make $t_{\text{nl}}^{(6)}$ a bit inflexible. In terms of the computational cost, training the quadratic models are expensive while training $t_{\text{nl}}^{(4)}$ incurs the lowest computational cost. We also note that the performance of quadratic model, $t_{\text{nl}}^{(5)}$, is comparable to the best models in our previous paper¹⁵ and shown in the supplementary material in terms of RMSE and significantly better in terms of the maximum absolute error in the predicted total kinetic energy. These results indicate the ef-

Systems	Parameters	$t_{nl}^{(4)}$	$t_{nl}^{(5)}$ (Quadratic)	$t_{nl}^{(6)}$ (Liu-Parr)	$t_{nl}^{(5,scaling)}(x)$ (Coordinate scaling)
H ₈	R^2	0.9941	0.9996	0.9898	0.9953
	RMSE	0.0029	0.0015	0.0038	0.0025
	T_{KSDFT} (Ha)	1.1147	1.1145	1.1145	1.1147
	T_{model} (Ha)	1.1136	1.1152	1.1202	1.1651
	T_{error} (Ha)	0.0011	0.0007	0.0057	0.0504
LiH	R^2	0.9984	0.9987	0.9982	0.9981
	RMSE	0.0038	0.0035	0.0040	0.0041
	T_{KSDFT} (Ha)	0.7583	0.7428	0.8068	0.7338
	T_{model} (Ha)	0.7651	0.7642	0.8566	0.7592
	T_{error} (Ha)	0.0075	0.0061	0.0498	0.0253
LiF	R^2	0.9999	0.9999	0.9994	0.9998
	RMSE	0.0043	0.0031	0.0087	0.0065
	T_{KSDFT} (Ha)	6.7109	7.0165	6.7263	6.9105
	T_{model} (Ha)	6.7355	7.0304	6.7752	7.0209
	T_{error} (Ha)	0.0246	0.0138	0.0489	0.1103
C ₄ H ₂	R^2	0.9996	0.9996	0.9986	0.9973
	RMSE	0.0069	0.0075	0.0100	0.0185
	T_{KSDFT} (Ha)	13.1728	12.4828	13.0915	13.4613
	T_{model} (Ha)	13.2473	12.4428	13.1402	13.5490
	T_{error} (Ha)	0.0745	0.0399	0.0995	0.0878
C ₄ N ₂	R^2	0.9999	0.9999	0.9993	0.9997
	RMSE	0.0031	0.0039	0.0089	0.0065
	T_{KSDFT} (Ha)	20.7398	20.2343	20.7399	20.6913
	T_{model} (Ha)	20.6410	20.2450	20.8654	20.7679
	T_{error} (Ha)	0.0987	0.0107	0.1834	0.0765
C ₃ O ₂	R^2	0.9999	0.9998	0.9995	0.9998
	RMSE	0.0030	0.0039	0.0066	0.0061
	T_{KSDFT} (Ha)	19.3773	19.3775	19.2867	19.3338
	T_{model} (Ha)	19.3316	19.4445	19.1988	19.4088
	T_{error} (Ha)	0.0456	0.0670	0.0878	0.0750

TABLE I. A summary of kinetic energy densities and total kinetic energies obtained from nonlocal models and model KSDFT calculations for one-dimensional structures in test data sets for the three nonlocal models proposed in Section II. For each system, three 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 model one-dimensional KSDFT calculations and the various models, respectively, and T_{error} is the error in total kinetic energy (Ha) between the one-dimensional model KSDFT value and the model. For each model type, T_{KE} , T_{model} and T_{error} only for the structure with the maximum absolute error (i.e. the worst case scenario) are shown meaning that for different models the worst case scenarios can correspond to different structures.

fectiveness of including higher order terms in the gradient expansion of kinetic energy into the nonlocal models. But this also suggests that by using a relatively small dataset (which may or may not cover the whole phase space of electron and kinetic densities), it is possible to generate models that violate traditional constraints but have lower RMSE values than models that satisfy those constraints.

Now, we compare the performance of our models with the other data-driven models proposed in the literature. Recently, Golub and Manzhos^{43,45} and Seino et al.⁵⁶ proposed models to learn kinetic energy functional of 3D systems with pseudopotentials using deep neural networks and Gaussian process regression. The performance of their models in terms of RMSE scores and errors in kinetic energy is similar to the models proposed in this paper. However, we have only focused on all electron, one-dimensional systems. One important difference is that our model has much fewer parameters to fit compared to the neural network based models. Hence, our model requires significantly lesser data points to train. Snyder

et. al. also proposed a kernel ridge regression based model to fit the total kinetic energy (instead of kinetic energy density) for model one-dimensional systems. The performance of our model in terms of total error in kinetic energy is similar. However, their model requires a much larger dataset to train as they have only a scalar quantity (i.e. total kinetic energy) to fit for a given structure.

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

Next, to assess the transferability of models outside of the training domain, we analyze the performance of the three new models (proposed in Section II) on charged systems obtained by inserting or removing electrons. Figure 3 shows kinetic energy densities of two prototypical systems, C₄N₂ and H₈, obtained from KSDFT calculations using the exact-exchange functional and from the three models (proposed in Section II)

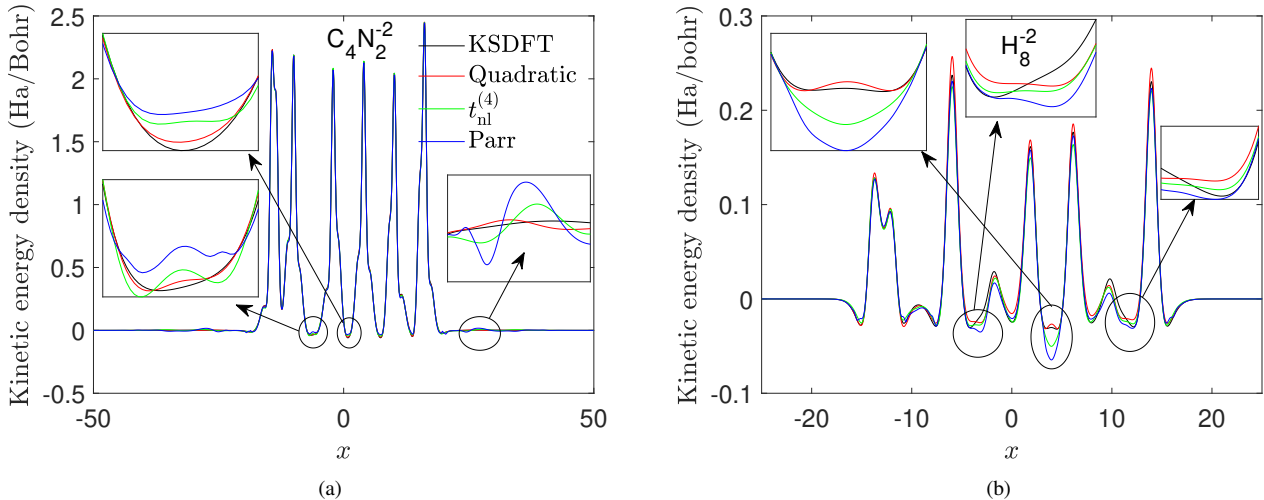


FIG. 3. Kinetic energy densities obtained from model KSDFT calculations using the exact-exchange functional are compared with model predictions for negatively charged one-dimensional systems, i.e. H_8^{-2} (a) and $\text{C}_4\text{N}_2^{-2}$ (b). Electron densities obtained from model one-dimensional KSDFT calculations (using exact-exchange functional) for charged systems are taken as inputs for model predictions. The insets highlight the behavior of the models close to the local minima of the kinetic energy density profiles.

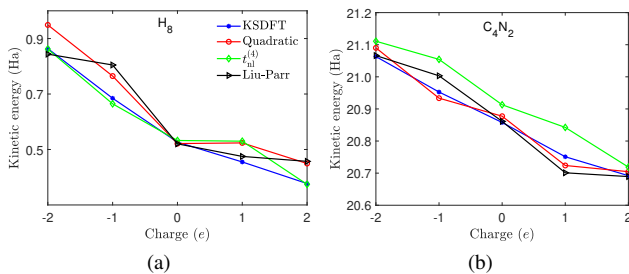


FIG. 4. Predicted total kinetic energies for charged H_8 and C_4N_2 systems from $t_{\text{nl}}^{(4)}$, $t_{\text{nl}}^{(5)}$ (quadratic model) and $t_{\text{nl}}^{(6)}$ (based on Liu-Parr formulation) models are compared to assess their performance outside of the training domain.

trained on three neutral structures of C_4N_2 and H_8 . Electron densities obtained from KSDFT calculations for these charged systems are used as model inputs. We observe that the quadratic model, $t_{\text{nl}}^{(5)}$, performs better than the two other models - the insets in Fig. 3 shows that kinetic energy densities predicted from the two other models show significant oscillations near local minima and inflection points (of the kinetic energy density). While nonlocal correlations between electron density and its derivatives are present in all three models, the oscillations are absent only in the results obtained from the quadratic model, $t_{\text{nl}}^{(5)}$. Therefore, we conjecture that this satisfactory performance of quadratic models is related to cancellation of errors caused by spatial correlations (Fig. 5) between the different derivative terms.

To further understand the behavior of these models, variations in total kinetic energies with respect to the total charge on these systems is shown in Fig. 4. For the one-dimensional chain of hydrogens, the total kinetic energies lie in the range

of 0.4 to 1 Ha which is much smaller than magnitudes of total kinetic energies of C_4N_2 . The errors in predicted total kinetic energies in case of systems with ± 1 charge are in the range of 10-15% in case of H_8 , but are about 5% in case of C_4N_2 , suggesting that all the models perform satisfactorily (in terms of errors in kinetic energy) outside of the training domain. However, if we consider both the slope of kinetic energy vs net charge on the system and kinetic energy density profiles, the quadratic model performs better than the other two models. These results suggest that a generic model for kinetic energy with good transferability (far away from the training domain) can be designed by using only a few structures for training.

D. A model for all systems

Inspired by the performance of the proposed models for charged systems, we train a *universal* model by using data from all the one-dimensional systems considered in this study. We use the quadratic model, $t_{\text{nl}}^{(5)}$, for this purpose because of its superior performance for test systems both inside and outside of the training domain. We train a model (universal model 1) by using three structures from each of the six systems, i.e. 18 structures in total. The RMSE and R^2 values for the predicted kinetic energy densities of structures not used in training (i.e. 162 structures) for this universal model 1 are 0.0035 Ha/Bohr and 0.999, respectively and the results are summarized in Table II. The good predictive capability of the quadratic model, $t_{\text{nl}}^{(5)}$, is clear from the high R^2 (more than 0.99) except for H_8 and LiH.

To further investigate this, the kinetic energy density obtained from model one-dimensional KSDFT calculations using exact-exchange functional and the *universal* model are shown in Fig. 5. It is clear that the model is able to cap-

	systems/metrics	C ₄ N ₂	C ₄ H ₂	C ₃ O ₂	H ₈	LiH	LiF
Universal model 1	R^2	0.9998	0.9999	0.9997	0.9068	0.9997	0.9933
	RMSE (Ha/Bohr)	0.0060	0.0046	0.0068	0.0113	0.0088	0.0080
	T _{KSDFT} (Ha)	20.4839	19.1199	13.1749	0.4397	6.7529	0.8068
	T _{model} (Ha)	20.4454	19.1496	13.2281	0.2873	6.9459	0.8486
	T _{error} (Ha)	0.0385	0.0297	0.0532	0.1524	0.1934	0.0418
Universal model 2	R^2	0.9999	0.9999	0.9998	-	-	-
	RMSE (Ha/Bohr)	0.0028	0.0022	0.0034	-	-	-
	T _{KSDFT} (Ha)	20.4745	19.1213	13.1876	-	-	-
	T _{model} (Ha)	20.4842	19.1289	13.1703	-	-	-
	T _{error} (Ha)	0.0098	0.0076	0.0172	-	-	-
Universal model 3	R^2	-	-	-	0.9976	0.9999	0.9999
	RMSE (Ha/Bohr)	-	-	-	0.0089	0.0056	0.0049
	T _{KSDFT} (Ha)	-	-	-	0.4410	6.8143	0.7986
	T _{model} (Ha)	-	-	-	0.4841	6.8634	0.8162
	T _{error} (Ha)	-	-	-	0.0431	0.0490	0.0176

TABLE II. The performance of the three *universal* quadratic models (described in $t_{nl}^{(5)}$) is assessed using different metrics - root mean square errors and R^2 scores corresponding to the predicted kinetic energy densities for each system, and the maximum absolute errors in predicted total kinetic energies. The universal model 2 corresponds to a model trained with 3 systems - C₄N₂, C₄H₂ and C₃O₂, and the universal model 3 corresponds to a model trained with 3 systems - LiH, LiF and H₈. The training set contains three structures from each of the six one-dimensional model systems and the remaining structures are set aside for testing purposes.

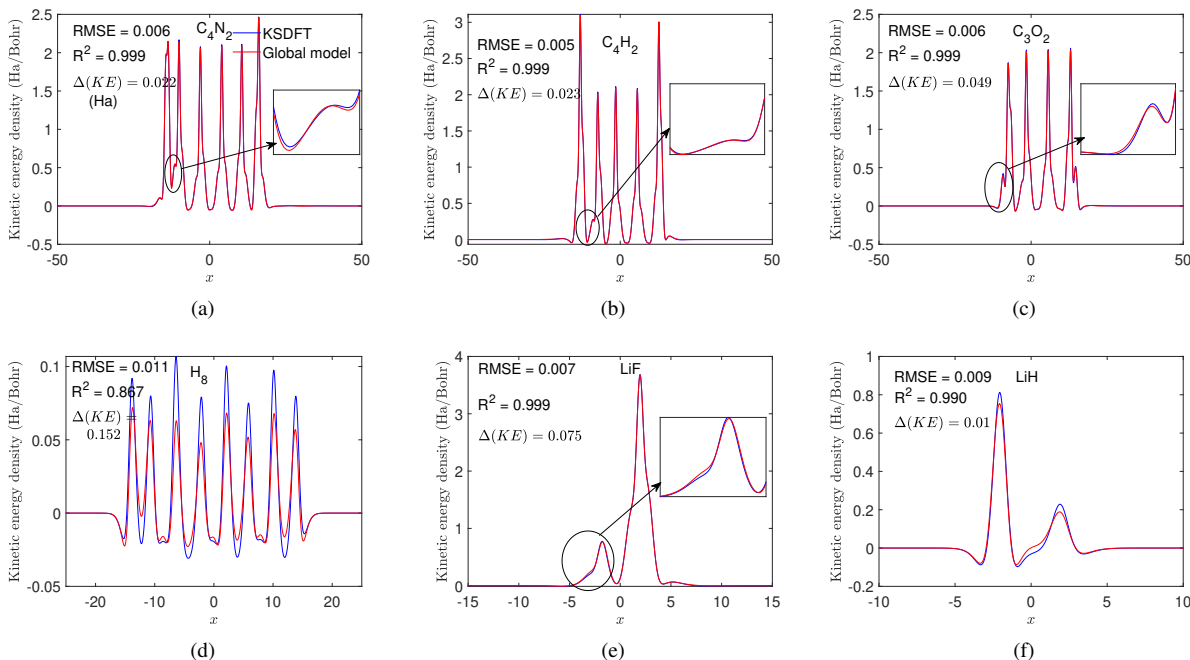


FIG. 5. Kinetic energy densities (for each of the six one-dimensional systems explored in this study) predicted using the *universal* quadratic model 1 for structures in the test set. The maximum absolute errors in predicted total kinetic energies, RMSE and high R^2 scores for individual systems are also shown. The insets show that model predictions are able to capture subtle features of the KSDFT kinetic energy densities. The training set used to train the universal model contains three structures from each of the six systems and the remaining structures are set aside for testing.

ture subtle features of the kinetic energy density for C₄N₂, C₄H₂ and C₃O₂. However, deviations from one dimensional model KSDFT results are clearly evident in H₈ and LiH. We conjecture that this discrepancy is related to the magnitude of kinetic energy densities of these systems - since the kinetic energy densities of H₈ and LiH are about an order of magnitude smaller than other systems, the *universal* model 1 is biased

towards larger values of kinetic energy densities. The kernels for quadratic models obtained for the individual systems and for the *universal* model 1 (highlighted in black) are shown in Fig. 6. It is clear that the kernels for LiH and H₈ are different from the other systems, and the kernel in the *universal* model is similar to the systems with larger kinetic energy densities. One possible remedy is to explore different objective func-

Systems	Parameters	Individual model	Universal model 1	Universal model 2	Systems	Parameters	Individual model	Universal model 1	Universal model 3
H ₈	R^2	0.9914	0.8817	0.9876	C ₄ H ₂	R^2	0.9989	0.9984	0.9992
	RMSE	0.0039	0.0140	0.0054		RMSE	0.0074	0.0086	0.0081
	\bar{T}_{KSDFT} (Ha)	-1.7549	-1.1145	-1.2430		\bar{T}_{KSDFT} (Ha)	3.1179	3.2522	3.1183
	\bar{T}_{model} (Ha)	-1.6669	-1.8698	-1.3449		\bar{T}_{model} (Ha)	3.0108	3.3805	3.2315
	\bar{T}_{error} (Ha)	0.0879	0.2696	0.1029		\bar{T}_{error} (Ha)	0.1071	0.1282	0.1132
LiH	R^2	0.9969	0.9711	0.9903	C ₄ N ₂	R^2	0.9997	0.9992	0.9993
	RMSE	0.0030	0.0093	0.0065		RMSE	0.0044	0.0082	0.0051
	\bar{T}_{KSDFT} (Ha)	-0.6192	-0.6193	-0.6192		\bar{T}_{KSDFT} (Ha)	7.1003	6.8673	6.8673
	\bar{T}_{model} (Ha)	-0.6063	-0.5541	-0.5905		\bar{T}_{model} (Ha)	7.1551	6.7514	6.9454
	\bar{T}_{error} (Ha)	0.0129	0.0652	0.0287		\bar{T}_{error} (Ha)	0.0548	0.1158	0.0781
LiF	R^2	0.9999	0.9988	0.9994	C ₃ O ₂	R^2	0.9996	0.9995	0.9992
	RMSE	0.0035	0.0112	0.0045		RMSE	0.0064	0.0069	0.0073
	\bar{T}_{KSDFT} (Ha)	2.4865	2.4040	2.4576		\bar{T}_{KSDFT} (Ha)	6.7275	7.1814	6.8234
	\bar{T}_{model} (Ha)	2.5129	2.6726	2.4963		\bar{T}_{model} (Ha)	6.6423	7.0795	6.8357
	\bar{T}_{error} (Ha)	0.0265	0.2686	0.0387		\bar{T}_{error} (Ha)	0.0851	0.1018	0.0123

TABLE III. A summary of modified kinetic energy densities ($\bar{T}_{\text{KE}} = T_s + E_X - E_{\text{LDA}}$) and total kinetic energies obtained from the quadratic model and one-dimensional model KSDFT calculations for structures in test data sets for the three nonlocal models proposed in Section II. Individual model is trained separately for each system with 3 structures, Universal model 1 is a single global model trained using a collection of 18 structures with 3 structure taken from each systems, and universal model 2 is trained with H₈, LiH and LiF with 3 structures from each and universal model 3 is trained with C₄H₂, C₄N₂ and C₃O₂ with 3 structures from each system. 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 one-dimensional model KSDFT calculations and the various models, respectively, and T_{error} is the error in total kinetic energy (Ha) between the 1D model KSDFT value and the model. For each model type, T_{KSDFT} , T_{model} and T_{error} only for the structure with the maximum absolute error are shown.

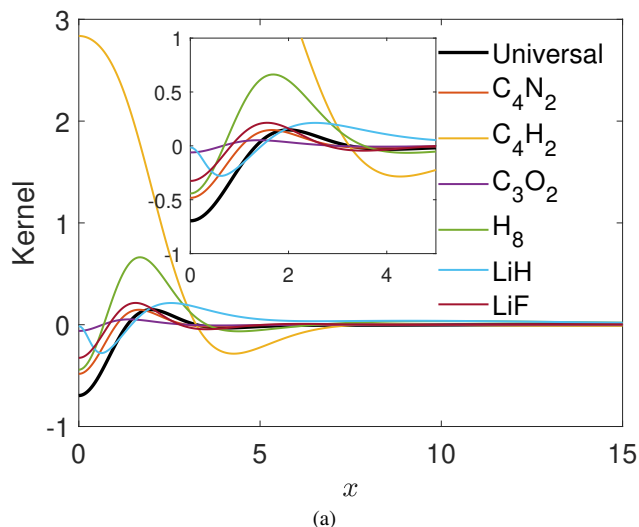


FIG. 6. Kernels obtained from quadratic models trained for the each of the six one-dimensional model systems explored in this study and the kernel obtained from the *universal* quadratic model described in Section III.D.

tions to obtain the model parameters. Another option is to use feature selection to design a better quadratic model with one or more kernels and we plan to explore this in future.

Interestingly, we also observe large errors in the total kinetic energy for LiH (0.19 Ha) and H₈ (0.15 Ha) suggesting that there are no cancellation of errors. Hence, we train two different quadratic models - universal model 2 with structures from LiH, LiF and H₈ and universal model 3 with structures

from C₄H₂, C₄N₂ and C₃O₂. We use 3 structures from each material, hence a total of 9 structures for each of the two models. We observe that errors in kinetic energies and kinetic energy densities (summarized in Table II) for all systems have reduced significantly for *universal model 2* and *universal model 3* compared to *universal model 1*. This validates our previous point that *universal models* work better when similar types of systems are used for training, but this behavior (i.e. failure of universal model 1 to handle a diverse set of systems) suggests that our model is still not complete and the true kinetic energy density functional perhaps is more complex and incorporates higher order spatial correlations. For comparison, we also train a *universal* model using the $Q_1 + Q_3$ model proposed in Ref. [15]. However, for the 162 structures in the test set, the RMSE and R^2 scores for the predicted kinetic energy densities are 0.061 Ha/Bohr and 0.983, respectively. Therefore, the quadratic model has better transferability and can be used to develop generic models.

E. Model for modified kinetic energy density

Inspired by the results of universal model based on the quadratic model, we use the quadratic model to fit a density functional which includes the kinetic energy and the exact exchange energy. The goal is to obtain a functional which in conjunction to the external potential and Hartree energy functionals can be used to perform an OFDFT calculation with exact exchange level of accuracy. Thus, we define a modified kinetic energy by adding the two orbital dependent terms (i.e., kinetic energy and exact-exchange energy

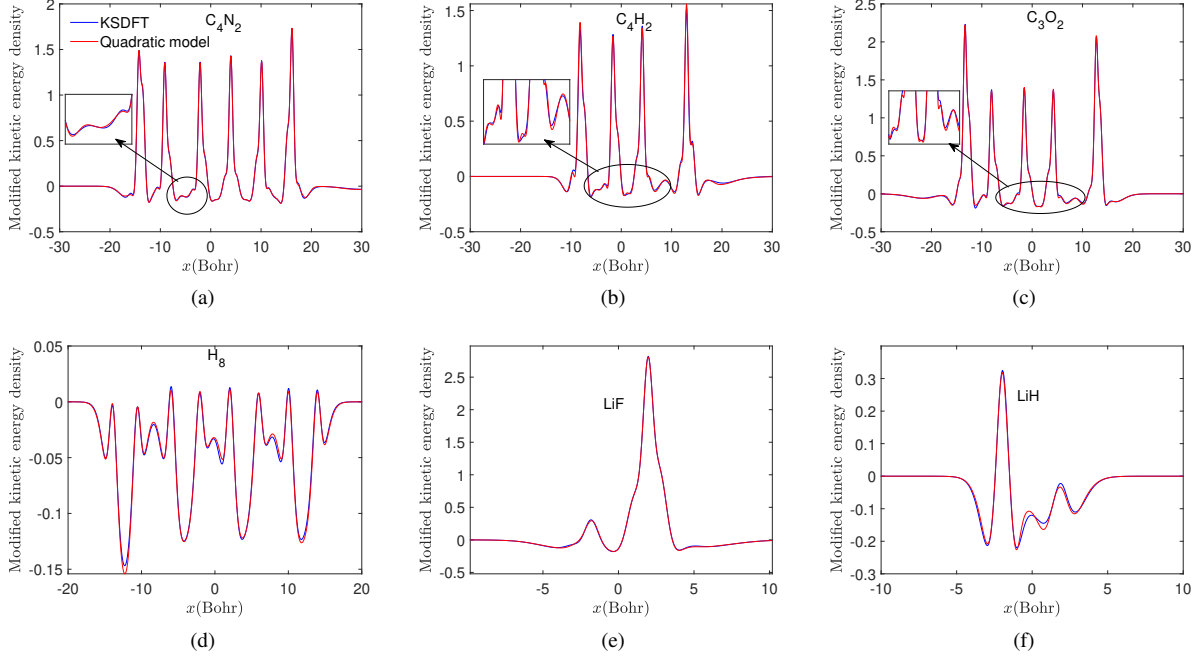


FIG. 7. Shown here are the modified kinetic energy densities (for each of the six one-dimensional systems explored in this study) predicted using the quadratic model for structures in the test set. The insets show that model predictions are able to capture subtle features of the model one-dimensional KSDFT kinetic energy densities. The training set used to train the universal model contains three structures from of given system and the remaining structures are set aside for testing.

densities). However, for consistency, we also subtract a local or semi-local exchange energy, i.e., we define $\bar{T}_{\text{KE}}[\rho] = T_s[\rho] + E_X[\rho(\{\psi_k\})] - E_{\text{LDA}}[\rho]$. Here we use the LDA exchange energy as we have already fitted the Kohn-Sham kinetic energy for LDA in our previous work. Therefore, the expression for the modified kinetic energy functional is given by

$$\begin{aligned}
 \bar{T}_{\text{KE}} &= T_s + E_X - E_{\text{LDA}}, \\
 E_X &= -\frac{1}{2}f_i f_j \\
 &\sum_i^{N_s} \sum_j^{N_s} \int \int \psi_i(x) \psi_j(x) \psi_i(x') \psi_j(x') \frac{1}{\sqrt{1+|x-x'|^2}} dx dx', \\
 E_{\text{LDA}}[\rho] &= \int v_{\text{LDA}}(\rho(x)) \rho(x) dx, \\
 v_{\text{LDA}} &= -\frac{A}{\pi} \tan^{-1} \left(\frac{\pi \rho}{2\kappa} \right),
 \end{aligned} \tag{13}$$

where \bar{T}_{KE} is the modified kinetic energy for which we want to develop a density-dependent model, T_s is the Kohn-Sham kinetic energy, E_X is the exact-exchange energy and E_{LDA} is the local density approximation (LDA) exchange energy. The modified kinetic energy essentially captures the kinetic energy of interacting electrons hence justifying the use of the kinetic energy density models developed in this paper.

We fit the quadratic model for modified kinetic energy (in Eq. 13) for the six systems. For training the model, we randomly select three structure and set aside the remaining 27

structures for testing. The parameters of the fitting, such as number of Gaussian functions, degree of polynomial terms in the local term, are same as the ones used in training the Kohn-Sham kinetic energy in previous sections. We summarize the performance of the quadratic model for modified kinetic energy in the third column of Table III. The plots showing modified kinetic energy densities obtained from KSDFT calculations and the model predictions are also shown in Fig. 7. We observe that the R^2 scores for all systems are greater than 0.99 indicating a very good performance of the quadratic model. From Fig. 7 we observe that the quadratic models are able to capture subtle features present in the modified kinetic energy density profiles. The maximum error in fit is observed near the minima points in the kinetic energy density, a pattern which has also been observed in the models developed by Golub et al⁴³. One of the reasons for this behavior may be the lack of data points in these regions as pointed out in the Ref. 43.

To further check the validity of quadratic model for modified kinetic energy density, we train three universal model with data from all systems using the modified kinetic energy density in Eq. 13. We use three structures from each systems for training the model. The results summarized in the fourth column of Table III show that R^2 scores for universal models 1 are similar to the models which were trained for the individual systems except for LiH and H₈. The discrepancy for these two materials is related to the significantly smaller kinetic energies of these two systems compared to the others as we noted in case of universal model for the Kohn-Sham ki-

netic energy. However, for the universal models 2 and 3, the model predictions agree very well with KSDFT results and the R^2 scores are very high. These results suggest that quadratic models can be used to describe kinetic energy as well as exact exchange energy with good accuracy.

IV. CONCLUSION

In this study, using data from a few model one-dimensional systems, such as H_8 , C_4H_2 , C_4N_2 , LiH , LiF and C_3O_2 , we explored a variety of empirical models of kinetic energy functional. To capture the subtle features of the kinetic energy density obtained from one-dimensional model DFT calculations with the exact-exchange functional, we used a variety of models proposed in our previous work and three nonlocal models developed in this work. These nonlocal models are inspired by physics-based models that have been proposed in the literature and incorporate spatial correlations between electron density and its higher order derivatives. The nonlocal kernels used to incorporate spatial correlations in the electron density and its derivatives are modeled by using a set of Gaussian basis functions. The predictive capabilities of the proposed models are systematically tested for structures both similar to those present in training sets and for charged systems (not present in training sets). The models are able to predict kinetic energy densities to a very high level of accuracy, i.e. with R^2 scores > 0.99 , and total kinetic energies with maximum errors of less than 0.001 Ha/atom and mean absolute error of about 0.004 Ha/atom. **We do observe large errors (maximum absolute error of 0.19 Ha) in total kinetic energy for universal models trained with all the systems, however those errors are greatly reduced (maximum absolute error of 0.04 Ha, mean absolute errors are less than 0.01 Ha) if the universal models are trained on structures with similar magnitudes of kinetic energy density values.** In addition to kinetic energy density, we show that the models are able to predict the exact exchange energy density to a high level of accuracy as well.

If we compare the R^2 scores, RMSE values for predicted kinetic energy densities and the maximum absolute errors in the total kinetic energies, the quadratic model performs much better than the two other nonlocal models proposed in this study. This implies that incorporating spatial correlations between the electron density and its higher order derivatives (up to the fourth order) helps in capturing salient features of kinetic energies obtained from one-dimensional model KSDFT with the exact-exchange functional. This implies that it is indeed possible to generate reliable models within the framework of data-informed, physics-based learning paradigm by using only a few structures. Furthermore, we show that it is possible to generate generic kinetic energy models for a broad class of heterogeneous systems. Even in this case, the models have good predictive capability and the quadratic model performs better than the other models analyzed in this study.

Although the models are validated for model one-dimensional systems, they can be generalized to three-dimensional systems. **We plan to tackle this as well as issues related to learning the nonlocal part of pseudopotential in a**

future publication. There are still a few challenges associated with these models, namely effective optimization scheme to obtain the best set of parameters for the quadratic model, data selection and avoiding bias towards systems with large kinetic energies, and we plan to tackle them in future. Overall, the models considered in this study are a step forward in the direction of generating accurate, reliable and computationally efficient data-driven orbital-free density functional theory models.

SUPPLEMENTARY MATERIAL

Table of results for models proposed in in Ref 15 and the Matlab codes to fit the models in this paper

ACKNOWLEDGMENT

This work was performed with LDRD funding (21-ERD-005) 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. The authors acknowledge helpful discussions with Xin Jing about the implementation of the exact exchange functional.

IM release #: LLNL-JRNL-827314

DATA AVAILABILITY

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

V. REFERENCES

- ¹W. Kohn and L. J. Sham, Physical review **140**, A1133 (1965).
- ²P. Hohenberg and W. Kohn, Physical review **136**, B864 (1964).
- ³F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, Nature communications **8**, 1 (2017).
- ⁴M. Brack, Reviews of modern physics **65**, 677 (1993).
- ⁵G. Kresse and J. Hafner, Physical Review B **47**, 558 (1993).
- ⁶G. Kresse and J. Hafner, Physical Review B **48**, 13115 (1993).
- ⁷S. Ghosh and P. Suryanarayana, Computer Physics Communications **212**, 189 (2017).
- ⁸Q. Xu, A. Sharma, B. Comer, H. Huang, E. Chow, A. J. Medford, J. E. Pask, and P. Suryanarayana, SoftwareX **15**, 100709 (2021).
- ⁹M. 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.
- ¹⁰W. Mi and M. Pavanello, Physical Review B **100**, 041105 (2019).
- ¹¹X. Shao, K. Jiang, W. Mi, A. Genova, and M. Pavanello, Wiley Interdisciplinary Reviews: Computational Molecular Science **11**, e1482 (2021).
- ¹²S. Carr, S. Fang, and E. Kaxiras, Nature Reviews Materials **5**, 748 (2020).
- ¹³H.-S. Do and B.-J. Lee, Scientific reports **8**, 1 (2018).
- ¹⁴A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weinreich, and D. E. Shaw, Proceedings of the National Academy of Sciences **116**, 4244 (2019), <https://www.pnas.org/content/116/10/4244.full.pdf>.
- ¹⁵S. Kumar, E. Landinez, B. Sadigh, S. Hamel, B. Gallagher, V. Bulatov, J. Klepeis, and A. Samanta, (2021).

- ¹⁶W. C. Witt, G. Beatriz, J. M. Dieterich, and E. A. Carter, *Journal of Materials Research* **33**, 777 (2018).
- ¹⁷V. Bach and L. Delle Site, *Many-Electron Approaches in Physics, Chemistry and Mathematics* (Springer, 2014).
- ¹⁸Y. Wang and E. Carter, *Progress in Theoretical Chemistry and Physics* **5**, 117 (2000).
- ¹⁹Y. A. Wang and E. A. Carter, *Theoretical methods in condensed phase chemistry*, 117 (2002).
- ²⁰L. Šamaj and J. Percus, *The Journal of chemical physics* **111**, 1809 (1999).
- ²¹D. Murphy, *Physical Review A* **24**, 1682 (1981).
- ²²W.-P. Wang, R. G. Parr, D. R. Murphy, and G. A. Henderson, *Chemical Physics Letters* **43**, 409 (1976).
- ²³C. Ma and V. Sahni, *Physical Review B* **16**, 4249 (1977).
- ²⁴C. Lee and S. K. Ghosh, *Physical Review A* **33**, 3506 (1986).
- ²⁵R. G. Parr, in *Horizons of quantum chemistry* (Springer, 1980) pp. 5–15.
- ²⁶E. Chacón, J. Alvarellos, and P. Tarazona, *Physical Review B* **32**, 7868 (1985).
- ²⁷L.-W. Wang and M. P. Teter, *Physical Review B* **45**, 13196 (1992).
- ²⁸W. C. Witt and E. A. Carter, *Physical Review B* **100**, 125107 (2019).
- ²⁹V. L. Lignères and E. A. Carter, in *Handbook of materials modeling* (Springer, 2005) pp. 137–148.
- ³⁰Y. A. Wang, N. Govind, and E. A. Carter, *Physical Review B* **60**, 16350 (1999).
- ³¹D. Garcia-Aldea and J. Alvarellos, *Physical Review A* **76**, 052504 (2007).
- ³²D. Garcia-Aldea and J. Alvarellos, *Journal of Chemical Physics* **127**, 052504 (2007).
- ³³W. Mi, A. Genova, and M. Pavanello, *The Journal of chemical physics* **148**, 184107 (2018).
- ³⁴Q. Xu, J. Lv, Y. Wang, and Y. Ma, *Physical Review B* **101**, 045110 (2020).
- ³⁵D. García-Aldea and J. E. Alvarellos, *Physical Chemistry Chemical Physics* **14**, 1756 (2012).
- ³⁶C. Huang and E. A. Carter, *Physical Review B* **81**, 045206 (2010).
- ³⁷J. P. Perdew and L. A. Constantin, *Physical Review B* **75**, 155109 (2007).
- ³⁸S. Laricchia, L. A. Constantin, E. Fabiano, and F. Della Sala, *Journal of chemical theory and computation* **10**, 164 (2014).
- ³⁹S. Liu, D. Zhao, C. Rong, T. Lu, and S. Liu, *Journal of Chemical Physics* **150**, 204106 (2019).
- ⁴⁰L. A. Constantin, E. Fabiano, and F. Della Sala, *Journal of chemical theory and computation* **15**, 3044 (2019).
- ⁴¹L. A. Constantin, E. Fabiano, and F. Della Sala, *The journal of physical chemistry letters* **9**, 4385 (2018).
- ⁴²A. A. Kugler, *Physical Review A* **41**, 3489 (1990).
- ⁴³P. Golub and S. Manzhos, *Physical Chemistry Chemical Physics* **21**, 378 (2019).
- ⁴⁴J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, *Physical review letters* **108**, 253002 (2012).
- ⁴⁵S. Manzhos and P. Golub, *The Journal of Chemical Physics* **153**, 074104 (2020).
- ⁴⁶S. Manzhos, *Machine Learning: Science and Technology* **1**, 013002 (2020).
- ⁴⁷M. M. Denner, M. H. Fischer, and T. Neupert, *Physical Review Research* **2**, 033388 (2020).
- ⁴⁸F. Imoto, M. Imada, and A. Oshiyama, *Physical Review Research* **3**, 033198 (2021).
- ⁴⁹S. A. Ghasemi and T. D. Kühne, *The Journal of Chemical Physics* **154**, 074107 (2021).
- ⁵⁰K. Yao and J. Parkhill, *Journal of chemical theory and computation* **12**, 1139 (2016).
- ⁵¹E. Fabiano and L. A. Constantin, *Physical Review A* **87**, 012511 (2013).
- ⁵²A. Borgoo and D. J. Tozer, *Journal of chemical theory and computation* **9**, 2250 (2013).
- ⁵³A. Borgoo, A. M. Teale, and D. J. Tozer, *Physical Chemistry Chemical Physics* **16**, 14578 (2014).
- ⁵⁴M. Levy and J. P. Perdew, *Physical Review A* **32**, 2010 (1985).
- ⁵⁵S. Liu and R. G. Parr, *Chemical physics letters* **278**, 341 (1997).
- ⁵⁶J. Seino, R. Kageyama, M. Fujinami, Y. Ikabata, and H. Nakai, *The Journal of chemical physics* **148**, 241705 (2018).
- ⁵⁷L. O. Wagner, E. Stoudenmire, K. Burke, and S. R. White, *Physical Chemistry Chemical Physics* **14**, 8581 (2012).
- ⁵⁸L. Lin, *Journal of Chemical Theory and Computation* **12**, 2242 (2016), pMID: 27045571, <https://doi.org/10.1021/acs.jctc.6b00092>.
- ⁵⁹X. Wu, A. Selloni, and R. Car, *Physical Review B* **79**, 085102 (2009).
- ⁶⁰P. Pulay, *Chemical Physics Letters* **73**, 393 (1980).
- ⁶¹P. P. Pratapa and P. Suryanarayana, *Chemical Physics Letters* **635**, 69 (2015).
- ⁶²A. S. Banerjee, P. Suryanarayana, and J. E. Pask, *Chemical Physics Letters* **647**, 31 (2016).
- ⁶³P. Suryanarayana and D. Phanish, *Journal of Computational Physics* **275**, 524 (2014).
- ⁶⁴B. Fornberg, *Mathematics of computation* **51**, 699 (1988).
- ⁶⁵S. Ghosh and P. Suryanarayana, *Journal of Computational Physics* **307**, 634 (2016).

Appendix I: Formulation

In this section, we briefly outline formulation and implementation of a one-dimensional model analog of Kohn-Sham DFT^{1,2} with exact-exchange potential with open boundary conditions. In what follows, x denotes the coordinate in 1D space. The ground state can be determined by solving the following nonlinear eigenvalue problem self-consistently:

$$\left\{ -\frac{1}{2} \frac{\partial^2}{\partial x^2} + v_{\text{ext}}(x) + v_{\text{H}}[\rho](x) + V_{\text{X}}[\{\psi_j\}](x) \right\} \psi_i(x) = \varepsilon_i \psi_i(x),$$

$$\int \psi_i(x) \psi_j(x) dx = \delta_{ij},$$

$$\rho(x) = \sum_{j=1}^{N_s} f_j \psi_j^2(x), \quad f_j \in \{1, 2\},$$
(14)

where ψ_i and ε_i are the model Kohn-Sham eigenfunctions and eigenvalues, respectively, $v_{\text{ext}}(x)$ is external potential, $v_{\text{H}}[\rho](x)$ is the Hartree potential, $V_{\text{X}}[\{\psi_j\}](x, x')$ is the exact-exchange operator, N_s is the total number of occupied orbitals and f_j is the occupation number of i^{th} orbital. In our model system, we replace the Coulomb form of Hartree potential and external nuclear potential by a soft form:⁵⁷

$$v_{\text{H}}[\rho](x) = \int_{\mathbf{R}} \frac{\rho(x') dx'}{\sqrt{1 + |x - x'|^2}},$$

$$v_{\text{ext}}(x) = \sum_I \frac{-Z_I}{\sqrt{1 + |x - X_I|^2}}.$$
(15)

Here, Z_I and X_I denote the nuclear charge and position of the I^{th} atom, respectively. The exact-exchange operator V_{X} is a nonlocal operator which depends on the occupied orbitals $\{\psi_j\}_{j=1}^{N_s}$ and has the following form^{58,59}

$$V_{\text{X}}[\{\psi_j\}](x, x') = - \sum_{j=1}^{N_s} f_j \frac{\psi_j(x) \psi_j(x')}{\sqrt{1 + |x - x'|^2}}.$$
(16)

The action of V_{X} on a given orbital $\psi(x)$ is given by^{58,59}

$$(V_{\text{X}}[\{\psi_j\}]\psi)(x) = - \sum_{j=1}^{N_s} f_j \psi_j(x) \int \frac{\psi(x') \psi_j(x')}{\sqrt{1 + |x - x'|^2}} dx'.$$
(17)

Therefore, the Hamiltonian in eigenvalue problem outlined in Eq. 14 depends both on the electron density, $\rho(x)$, as well as the orbitals, $\{\psi_j\}_{j=1}^{N_s}$. Hence, we solve this eigenvalue problem by using two sets of self-consistent field iterations⁵⁸ - in the outer loop exchange operator is updated and in the inner loop, the electron density is updated using Pulay mixing⁶⁰⁻⁶² while keeping the exchange operator constant.

Appendix II: Implementation

We solve the partial differential equation in Eq. 14 numerically by using a finite difference discretization method in which a supercell with edge length 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. 14 is approximated using the central finite-difference approximation⁶³⁻⁶⁵

$$\left. \frac{\partial^2 g}{\partial x^2} \right|^{(i)} = \sum_{k=0}^{N_{\text{FD}}} W_k \left(g^{(i+k)} + g^{(i-k)} \right) + O(h^{2N_{\text{FD}}}), \quad (18)$$

where $g^{(i)}$ denotes the value of the function at the i -th grid point and N_{FD} is the order of finite difference approximation. The weights, W_k , associated with the function values are:

$$W_0 = -\frac{1}{h^2} \sum_{k=1}^{N_{\text{FD}}} \frac{1}{k^2},$$

$$W_k = \frac{2(-1)^{k+1}}{h^2 k^2} \frac{(N_{\text{FD}}!)^2}{(N_{\text{FD}} - k)!(N_{\text{FD}} + k)!}, \quad k = 1, 2, \dots, N_{\text{FD}}. \quad (19)$$

With the above discretization scheme, eigenvalue equation in Eq. 14 can be written in terms of matrix eigenvalue problem:

$$H\psi = \varepsilon\psi, \quad H \in \mathbb{R}^{N \times N}, \psi \in \mathbb{R}^{N \times 1},$$

$$H\psi = -\frac{1}{2}\nabla_h^2\psi + (v_H[\rho] + v_{\text{ext}})\psi + (V_X[\{\psi_j\}])\psi, \quad (20)$$

$$\rho^{(i)} = \sum_{j=1}^{N_s} f_j \psi_j^{(i)} \psi_j^{(i)},$$

where ∇_h^2 is the finite-difference discrete second derivative matrix, $\psi_j^{(i)}$ is the value of the j -th wavefunction at the i -th grid point.

The eigenvalue problem outlined above in Eq. 20 is solved in Matlab by using the inbuilt eigensolver. With a given initial guess of electron density and the exchange operator, these equations are solved by using two sets of nested self consistent iterations until convergence to a desired accuracy is reached. The convergence for both the inner and outer loops are checked with respect to relative error in electron density and the change in energy of the system and the self-consistent iterations are considered to have converged when the change

in energy per atom is less than 10^{-6} Ha. The kinetic energy and kinetic energy density are obtained by using the converged orbitals using the following expressions:

$$T_s = \sum_{i=1}^N \sum_{j=1}^{N_s} h \psi_j^{(i)} \nabla_h^2 \psi_j^{(i)},$$

$$t_s^{(i)} = \sum_{j=1}^{N_s} \psi_j^{(i)} \nabla_h^2 \psi_j^{(i)}. \quad (21)$$

Here $T_s = \int t_s(x) dx$ and $t_s(x)$ are total kinetic energy and kinetic energy density respectively. Information about the mesh sizes used to discretize the domain are given in Table IV below.

systems	H ₈ , LiH, LiF	C ₄ N ₂	C ₄ H ₂ , C ₃ O ₂
grid size (Bohr)	0.05	0.06	0.06
edge length (Bohr)	50	100	100

TABLE IV. Parameters used for solving the one-dimensional model Kohn-Sham equation using the exact-exchange functional.

Appendix III: Analytic expression for the derivatives of $P(\mathbf{r}, \mathbf{r}')$

We obtain the analytical expressions of the first and second order derivative terms of $P(\mathbf{r}, \mathbf{r}')$ appearing in Eq. 12 for the nonlocal model based on density matrix proposed by Liu and Parr. These expression can be substituted in Eq. 12 to obtain the expression of kinetic energy density $t_{\text{nl}}^{(6)}(\mathbf{r})$ as a function of the fitting coefficients which are later obtained by minimizing the error from kinetic energy density obtained from one-dimensional model KSDFT simulations.

$$P(\mathbf{r}, \mathbf{r}') = \sum_{k=1}^{N_G} d_k e^{-\beta_k |\mathbf{r} - \mathbf{r}'|^2} (\rho(\mathbf{r}) + \rho(\mathbf{r}'))^2,$$

$$\nabla_{\mathbf{r}} P(\mathbf{r}, \mathbf{r}') = \sum_{k=1}^{N_G} d_k e^{\beta_k \alpha} (-2\beta_k |\mathbf{r} - \mathbf{r}'|^2 [\rho(\mathbf{r}) + \rho(\mathbf{r}')] \nabla_{\mathbf{r}} \rho(\mathbf{r}) - 2\beta_k (\mathbf{r} - \mathbf{r}') [\rho(\mathbf{r}) + \rho(\mathbf{r}')]^2),$$

$$\nabla_{\mathbf{r}'} P(\mathbf{r}, \mathbf{r}') = \sum_{k=1}^{N_G} d_k e^{\beta_k \alpha} (-2\beta_k |\mathbf{r} - \mathbf{r}'|^2 [\rho(\mathbf{r}) + \rho(\mathbf{r}')] \nabla_{\mathbf{r}'} \rho(\mathbf{r}') - 2\beta_k (\mathbf{r}' - \mathbf{r}) [\rho(\mathbf{r}) + \rho(\mathbf{r}')]^2),$$

$$\nabla_{\mathbf{r}} \cdot \nabla_{\mathbf{r}'} P(\mathbf{r}, \mathbf{r}') = \sum_{k=1}^{N_G} d_k e^{\beta_k \alpha} [-4\beta_k (\mathbf{r} - \mathbf{r}') [\rho(\mathbf{r}) + \rho(\mathbf{r}')] \cdot (\nabla_{\mathbf{r}'} \rho(\mathbf{r}') - \nabla_{\mathbf{r}} \rho(\mathbf{r})) - 2\beta_k |\mathbf{r} - \mathbf{r}'|^2 \nabla_{\mathbf{r}} \rho(\mathbf{r}) \cdot \nabla_{\mathbf{r}'} \rho(\mathbf{r}') + 2\beta_k [\rho(\mathbf{r}) + \rho(\mathbf{r}')]^2 + \sum_{k=1}^{N_G} d_k e^{\beta_k \alpha} (-2\beta_k |\mathbf{r} - \mathbf{r}'|^2 [\rho(\mathbf{r}) + \rho(\mathbf{r}')] \nabla_{\mathbf{r}} \rho(\mathbf{r}) - 2\beta_k (\mathbf{r} - \mathbf{r}') [\rho(\mathbf{r}) + \rho(\mathbf{r}')]^2) \cdot (-2\beta_k |\mathbf{r} - \mathbf{r}'|^2 [\rho(\mathbf{r}) + \rho(\mathbf{r}')] \nabla_{\mathbf{r}'} \rho(\mathbf{r}') + 2\beta_k (\mathbf{r}' - \mathbf{r}) [\rho(\mathbf{r}) + \rho(\mathbf{r}')]^2),$$

$$\alpha = -|\mathbf{r} - \mathbf{r}'|^2 (\rho(\mathbf{r}) + \rho(\mathbf{r}'))^2. \quad (22)$$

Here, d_k are the coefficients, β_k and N_G are the widths and the number of the Gaussian functions used to model the function $P(\mathbf{r}, \mathbf{r}')$.