Approaching the precision energy surface in the local density approximation level density functional theory

XX

(Dated: April 30, 2024)

TODO

I. INTRODUCTION

In this article, we combine the machine learning (ML) method and density functional theory (DFT) to predict the properties of the molecules. By predicting both the potential and the energy from the electron density of the molecule using the ML method, we can calculate the properties of the molecule, such as the energy, the bond length, and the dipole moment. The ML method is based on the well-developed UNET architecture, which is a convolutional neural network (CNN) for fast and precise segmentation of images [?]. The target potential and the target energy are obtained from the inverse DFT calculation, in this article, we use the modified Ryabinkin-Kohut-Staroverov method.

II. METHODS

The density functional theory has dominated quantum chemistry calculation in the last few decades, especially in the large molecular system. One of the standard methods is the Kohn-Sham (KS) DFT, which is based on the Hohenberg-Kohn theorem and the Kohn-Sham equation. The KS DFT can be written as

$$E[\rho(\mathbf{r})] = T_{\mathrm{S}}[\rho(\mathbf{r})] + E_{\mathrm{J}}[\rho(\mathbf{r})] + E_{\mathrm{XC}}[\rho(\mathbf{r})] + E_{\mathrm{ext}}[\rho(\mathbf{r})],$$
(1a)

$$\int d\mathbf{r}\rho(\mathbf{r}) = N.$$
(1b)

where The N is the total number of electrons in the system, $E_{\mathrm{XC}}[\rho(\mathbf{r})]$ is the exchange-correlation energy functional which has many forms and

$$T_{\mathrm{S}}[\rho(\mathbf{r})] = -\frac{1}{2}\nabla^2\rho(\mathbf{r}),$$
(2a)

$$E_{\mathrm{J}}[\rho(\mathbf{r})] = \frac{1}{2}\int d\mathbf{r}_i d\mathbf{r}_j \frac{\rho(\mathbf{r}_i)\rho(\mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|},$$
(2b)

$$E_{\mathrm{ext}}[\rho(\mathbf{r})] = \int d\mathbf{r}\rho(\mathbf{r})V_{\mathrm{ext}}(\mathbf{r}).$$
(2c)

Here, the $V_{\mathrm{ext}}(\mathbf{r})$ is the external potential. For example, the nuclear potential caused by ions with Z number of charge at position R, $V_{\mathrm{ext}}(\mathbf{r}) = \int d\mathbf{r}\frac{Z\rho(\mathbf{r})}{|\mathbf{r}-R|}$.

Solve Equation (1) directly caused the high computational cost, one of the common methods is to use the finite basis set. The finite basis set starts with expanding the molecular orbital as the linear combination of the atom orbital,

$$\phi_i^\sigma(\mathbf{r}) = \sum_\mu C_{\mu i}^\sigma \chi_\mu(\mathbf{r}),$$
(3)

here we use i to index the molecular orbital σ to index the spin and μ to index the basis set. The density $\rho(\mathbf{r}_i)$ can be rewritten as

$$\rho^\sigma(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu}^\sigma \chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}),$$
(4)

where $P_{\mu\nu}^\sigma = \sum_i^{N_\sigma} C_{\mu i} C_{\nu i}$.

The energy functional can be written as

$$
\begin{aligned}
E[\rho^\sigma(\mathbf{r})] &= E[P_{\mu\nu}^\sigma] \\
&= \sum_{\mu\nu} P_{\mu\nu}^\sigma H_{\mu\nu}^\sigma + \frac{1}{2}\sum_{ijkl} P_{ij}^\sigma J_{ijkl}^\sigma P_{kl}^\sigma \\
&\quad + \sum_{\mu\nu} E_{\mathrm{XC}}[P_{\mu\nu}^\sigma] P_{\mu\nu}^\sigma.
\end{aligned}
$$
(5)

III. RESULTS

We use carbon-hydrogen molecules in the g2/97 dataset as the training set, and the test set is the Pentane, Isopentane, Neopentane, and Cyclopentane. The performance of the ML model is shown below,

Molecule	Energy (mH)	density (me)
training set	185.89	445.34
test set	178.76	662.45

TABLE I. The performance of the ML model

Then the self-consistent field (SCF) can be written as

$$F = H^\sigma + J + V = \epsilon_i^\sigma C S^\sigma C^{-1},$$
(6)