



# Orbital-Free Density Functional Theory for Molecular Systems Using Deep Learning

Chang Liu  
Microsoft Research AI for Science

# Computational Methods for Molecular Science

## molecular problems

biomolecule  
understanding

drug design

material/catalyst  
discovery

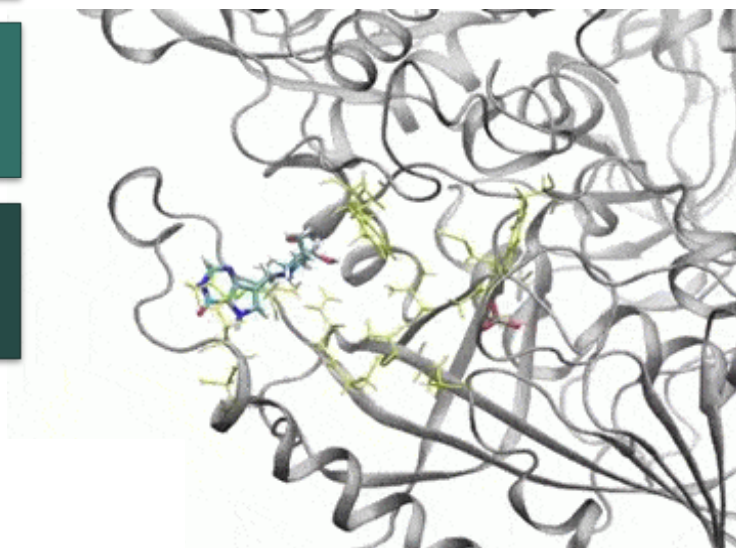
...

# Computational Methods for Molecular Science

## molecular problems



simulation



## molecular properties

$E$

$\rho$

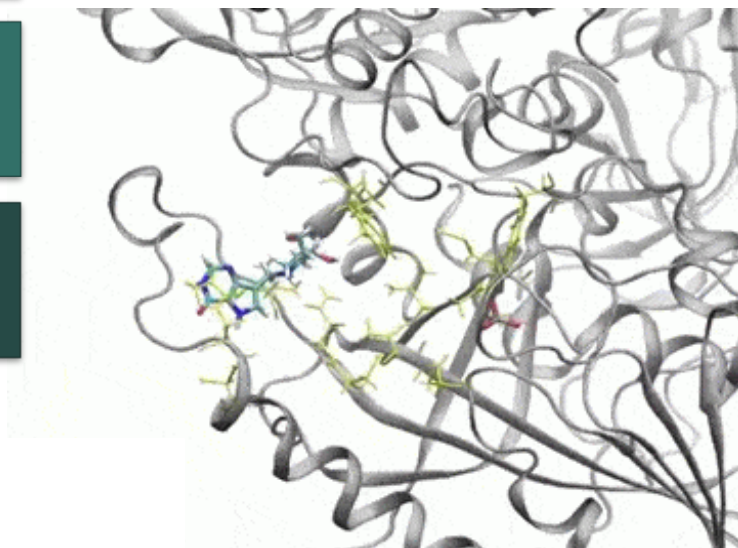
$\mathbf{f}$

$\vdots$

# Computational Methods for Molecular Science

## molecular problems

- biomolecule understanding
- drug design
- material/catalyst discovery
- ...



simulation

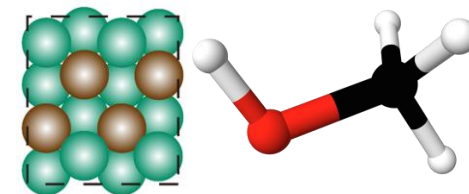
## molecular properties

$E$   
 $\rho$   
 $\mathbf{f}$   
...

Solve the electronic structure:  
Schrödinger equation

$$\hat{H}\psi = E\psi$$

## molecular structure



# Electronic Structure Methods

- To solve the  $N$ -electron Schrödinger equation  $\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ :

# Electronic Structure Methods

- To solve the  $N$ -electron Schrödinger equation  $\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ :

**Variational Principle:**

$$\min_{\psi} \langle \psi | \hat{H} | \psi \rangle.$$

↖ one function on  $\mathbb{R}^{3N}$ .

# Electronic Structure Methods

- To solve the  $N$ -electron Schrödinger equation  $\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ :

**Variational Principle:**

$$\min_{\psi} \langle \psi | \hat{H} | \psi \rangle.$$

↖ one function on  $\mathbb{R}^{3N}$ .

Complexity:

$$O(\exp(N))$$

# Electronic Structure Methods

- To solve the  $N$ -electron Schrödinger equation  $\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ :

## Variational Principle:

$$\min_{\psi} \langle \psi | \hat{H} | \psi \rangle.$$

↖ one function on  $\mathbb{R}^{3N}$ .

In hope to reduce complexity



## Density Functional Theory:

$$\min_{\rho} F[\rho] + \underbrace{E_{\text{ext}}[\rho]}_{\text{known}}.$$

↖ one function on  $\mathbb{R}^3$ .

Complexity:

$$O(\exp(N))$$



# Electronic Structure Methods

- To solve the  $N$ -electron Schrödinger equation  $\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ :

## Variational Principle:

$$\min_{\psi} \langle \psi | \hat{H} | \psi \rangle.$$

↖ one function on  $\mathbb{R}^{3N}$ .

In hope to reduce complexity

## Kohn-Sham DFT:

$$\min_{\{\phi_i\}_{i=1}^N} \underbrace{T_S[\{\phi_i\}] + E_H[\rho_{\{\phi_i\}}]}_{\text{known}} + \underbrace{E_{xc}[\rho_{\{\phi_i\}}]}_{\text{unknown but easier to approx.}} + \underbrace{E_{\text{ext}}[\rho_{\{\phi_i\}}]}_{\text{known}}.$$

## Density Functional Theory:

$$\min_{\rho} F[\rho] + \underbrace{E_{\text{ext}}[\rho]}_{\text{known}}.$$

↖ one function on  $\mathbb{R}^3$ .

$F[\rho]$  is unknown

Complexity:

$$O(\exp(N))$$

# Electronic Structure Methods

- To solve the  $N$ -electron Schrödinger equation  $\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ :

## Variational Principle:

$$\min_{\psi} \langle \psi | \hat{H} | \psi \rangle.$$

↖ one function on  $\mathbb{R}^{3N}$ .

In hope to reduce complexity

## Density Functional Theory:

$$\min_{\rho} F[\rho] + \underbrace{E_{\text{ext}}[\rho]}_{\text{known}}.$$

↖ one function on  $\mathbb{R}^3$ .

$F[\rho]$  is unknown

## Kohn-Sham DFT:

$$\min_{\{\phi_i\}_{i=1}^N} \underbrace{T_S[\{\phi_i\}] + E_H[\rho_{\{\phi_i\}}]}_{\text{known}} + \underbrace{E_{\text{xc}}[\rho_{\{\phi_i\}}]}_{\text{unknown but easier to approx.}} + \underbrace{E_{\text{ext}}[\rho_{\{\phi_i\}}]}_{\text{known}}.$$

↖  $N$  functions on  $\mathbb{R}^3$ .

Complexity:

$$O(\exp(N))$$

$$\Omega(N^3)$$

# Electronic Structure Methods

- To solve the  $N$ -electron Schrödinger equation  $\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$ :

## Variational Principle:

$$\min_{\psi} \langle \psi | \hat{H} | \psi \rangle.$$

one function on  $\mathbb{R}^{3N}$ .

In hope to reduce complexity

## Density Functional Theory:

$$\min_{\rho} F[\rho] + \underbrace{E_{\text{ext}}[\rho]}_{\text{known}}.$$

one function on  $\mathbb{R}^3$ .

$F[\rho]$  is unknown

## Kohn-Sham DFT:

$$\min_{\{\phi_i\}_{i=1}^N} \underbrace{T_S[\{\phi_i\}] + E_H[\rho_{\{\phi_i\}}]}_{\text{known}} + \underbrace{E_{\text{xc}}[\rho_{\{\phi_i\}}]}_{\text{unknown but easier to approx.}} + \underbrace{E_{\text{ext}}[\rho_{\{\phi_i\}}]}_{\text{known}}.$$

$N$  functions on  $\mathbb{R}^3$ .

$$T_S[\rho] := \min_{\{\phi_i\}_{i=1}^N : \rho_{\{\phi_i\}} = \rho} T_S[\{\phi_i\}]$$

## Orbital-Free DFT:

$$\min_{\rho} T_S[\rho] + E_H[\rho] + E_{\text{xc}}[\rho] + E_{\text{ext}}[\rho].$$

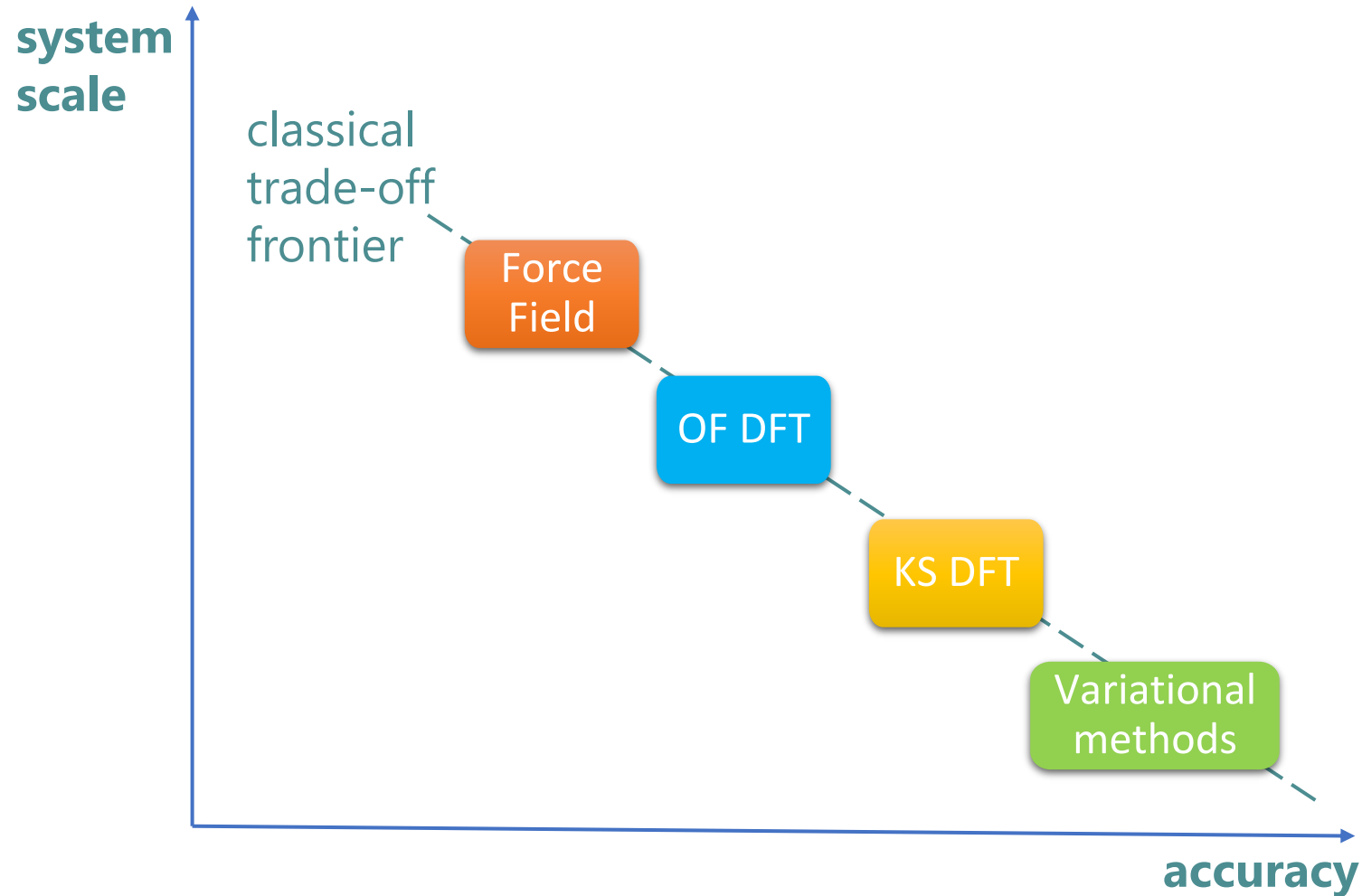
Complexity:

$$O(\exp(N))$$

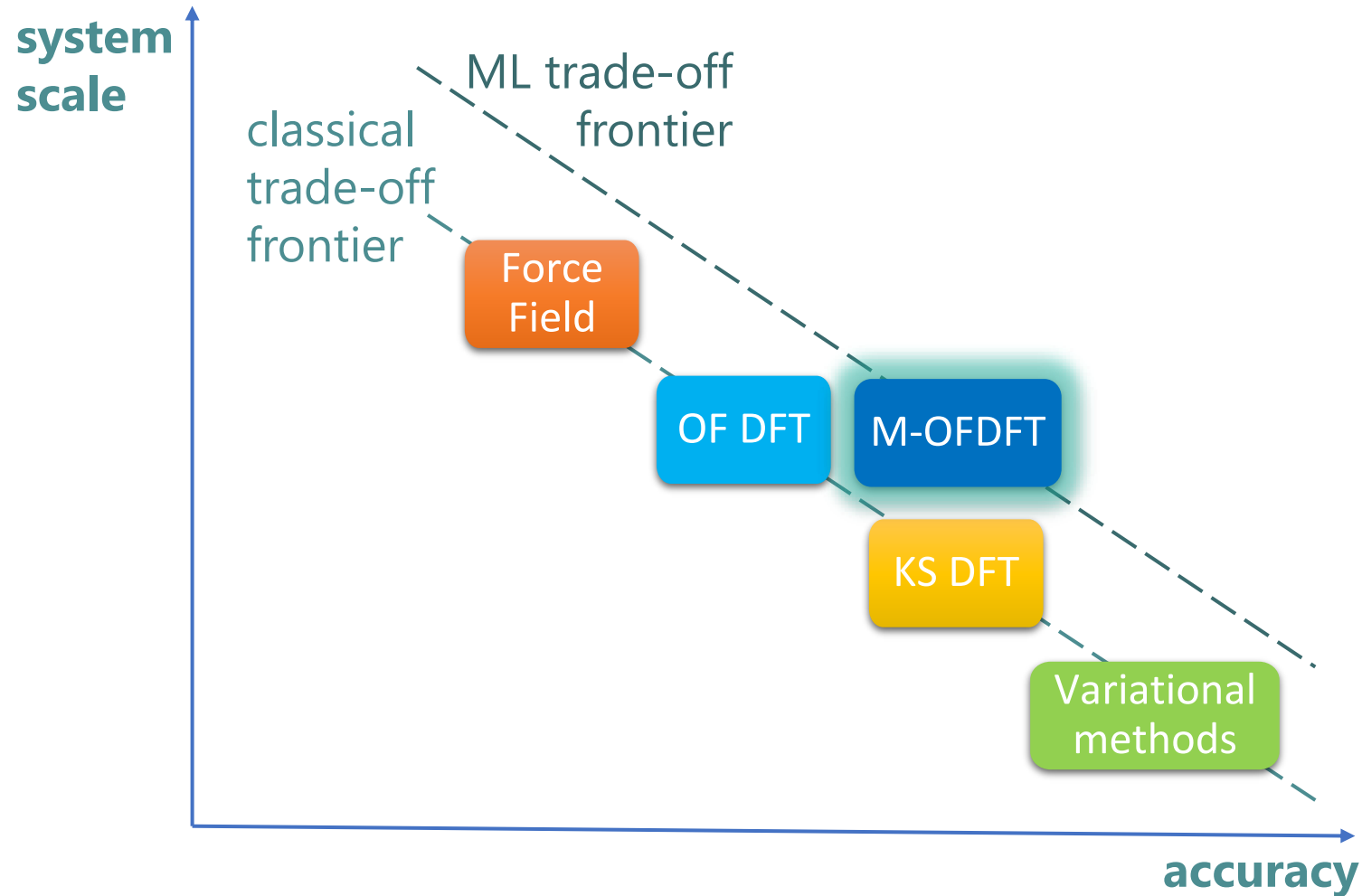
$$\Omega(N^3)$$

$$O(N^2)$$

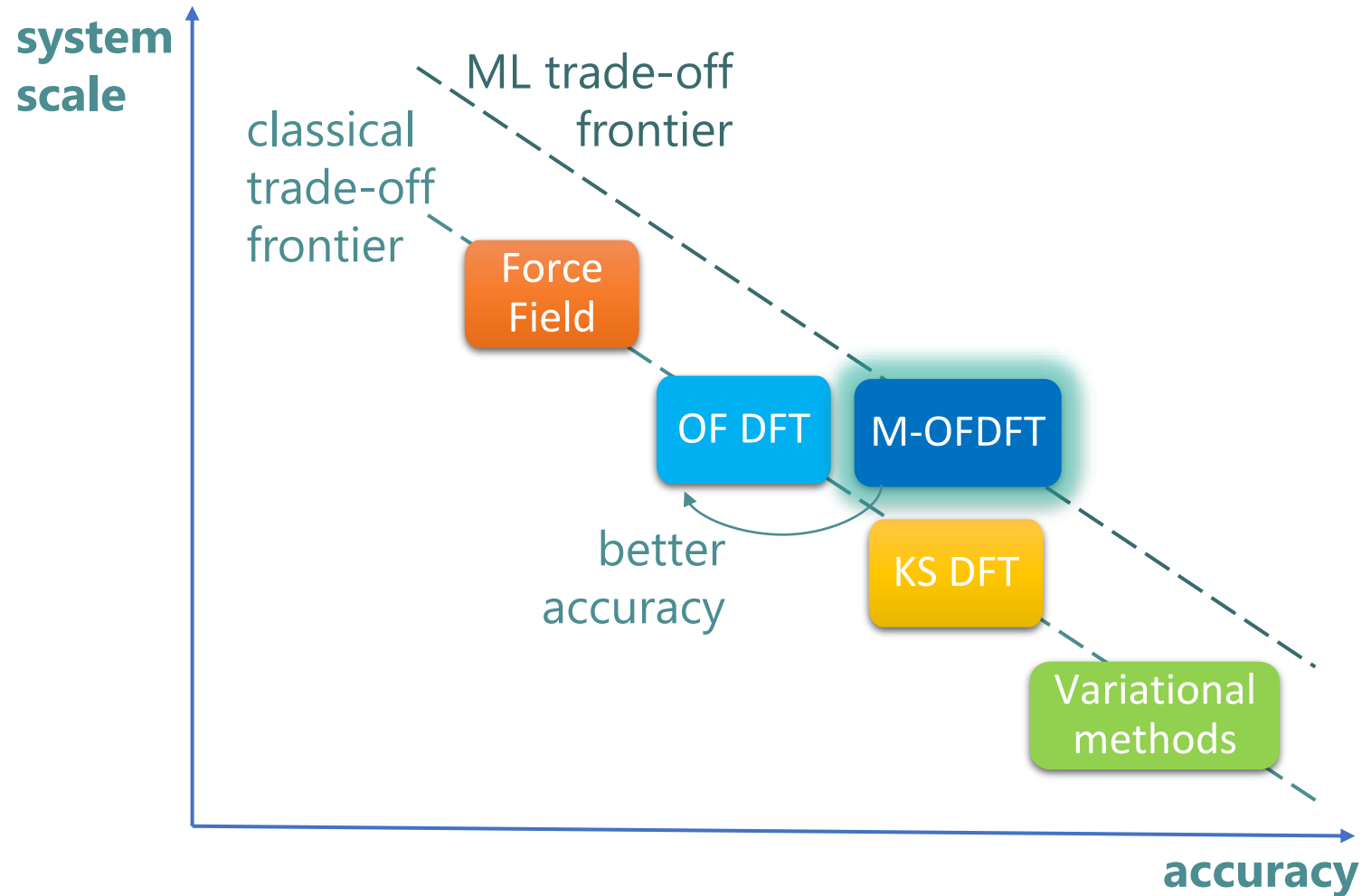
# Electronic Structure Methods



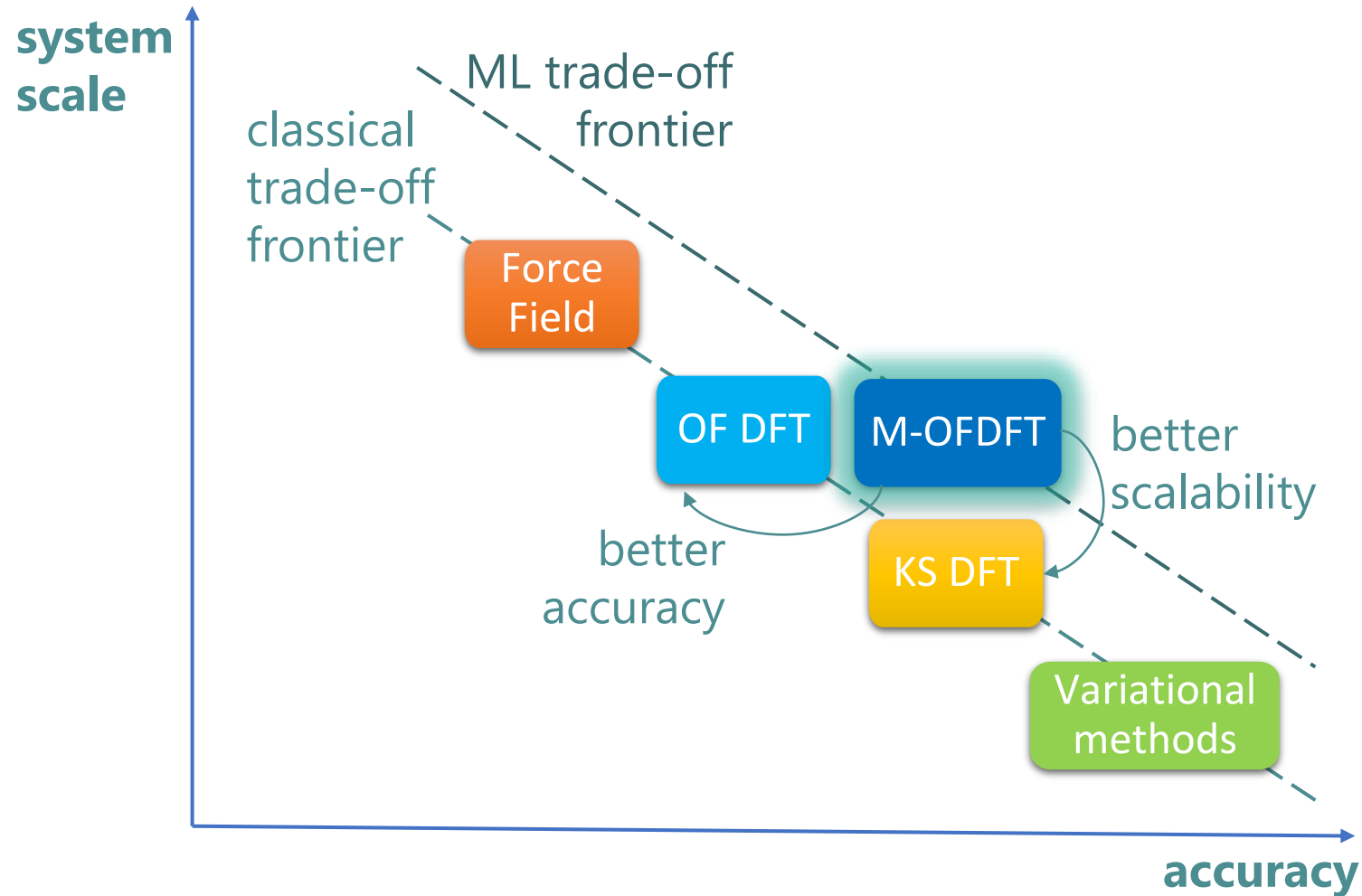
# Electronic Structure Methods



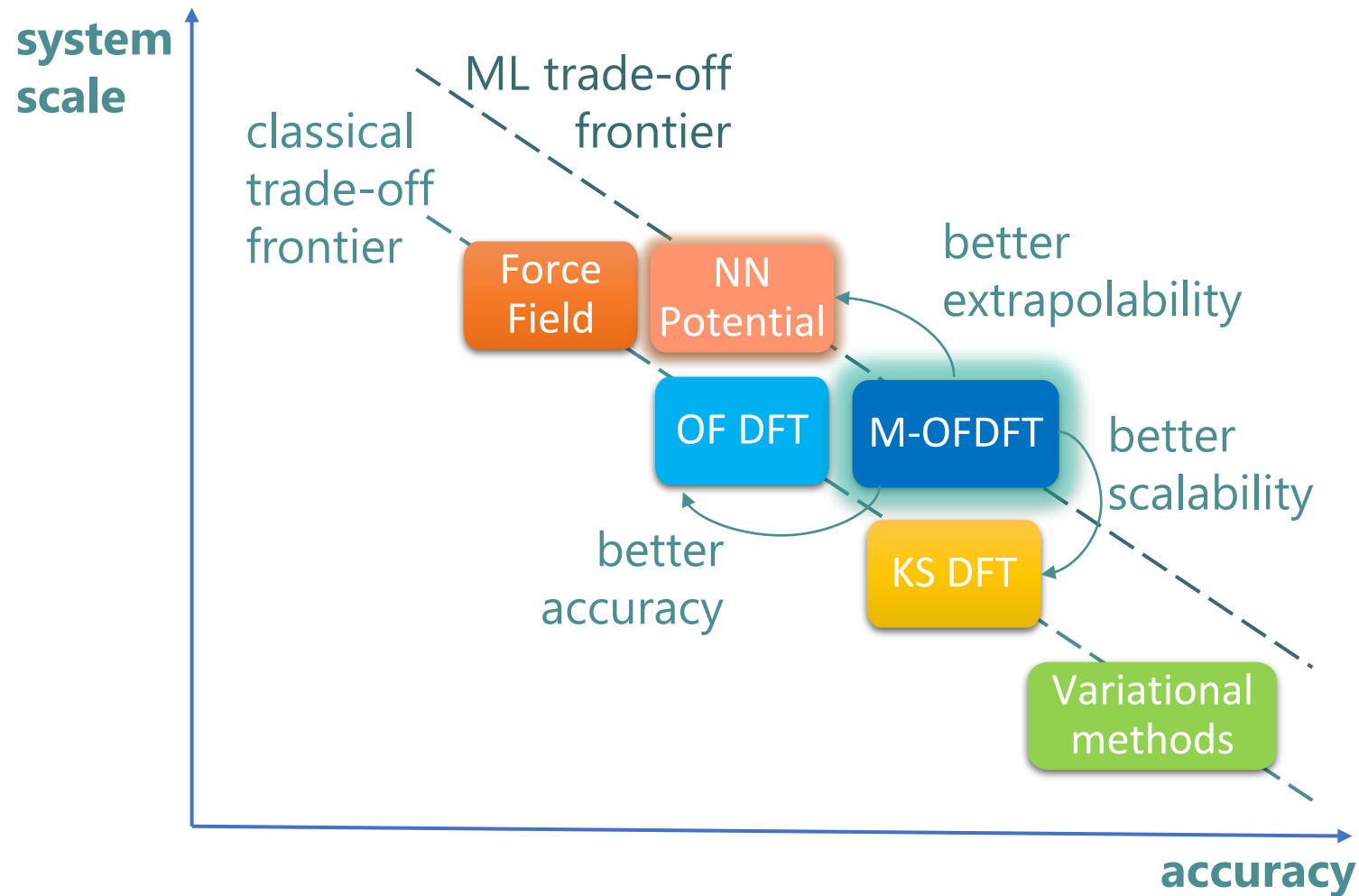
# Electronic Structure Methods



# Electronic Structure Methods



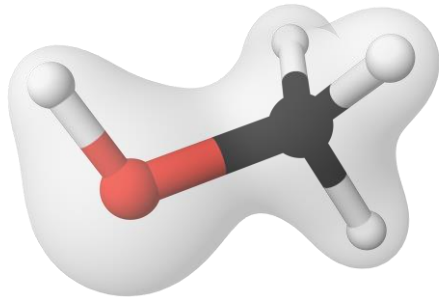
# Electronic Structure Methods





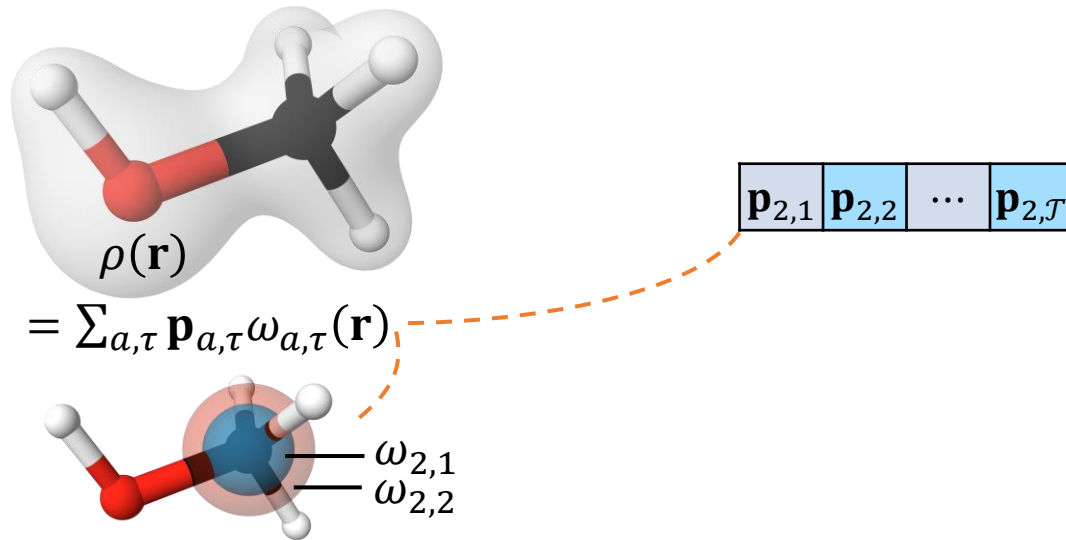
# Approximating $T_s[\rho]$

- Density representation
  - Grid-based density representation:  
nonlocal calculation is expensive on  $\sim 10^4 N$  grid points.
  - Limited to 10~20 atoms.



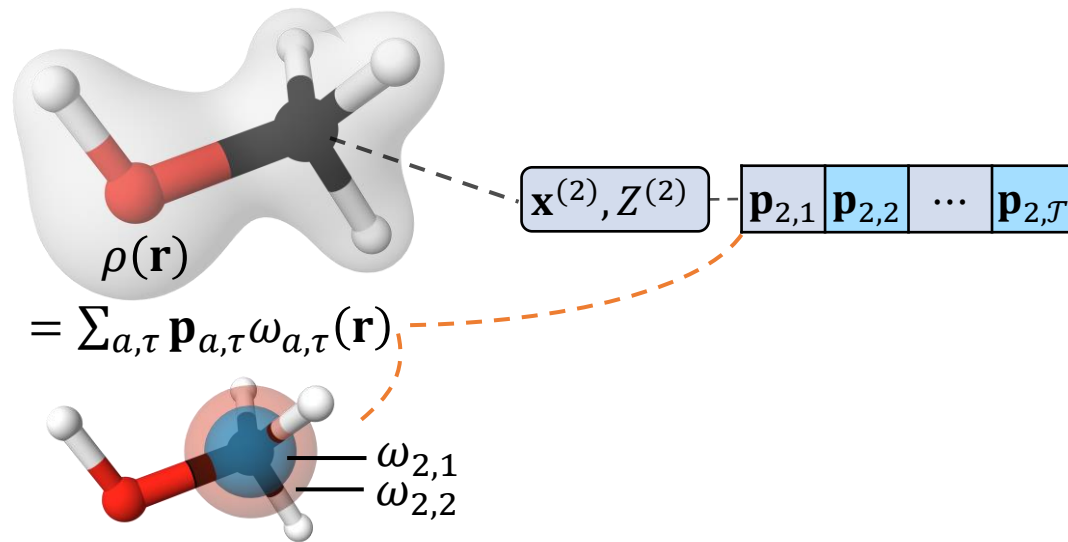
# M-OFDFT: the $T_S[\rho]$ Model

- Density representation using **atomic basis**:  
Expansion coefficients  $\mathbf{p}$  on atomic basis:  $\rho$  represented as  $(\mathbf{p}, \mathcal{M})$ .



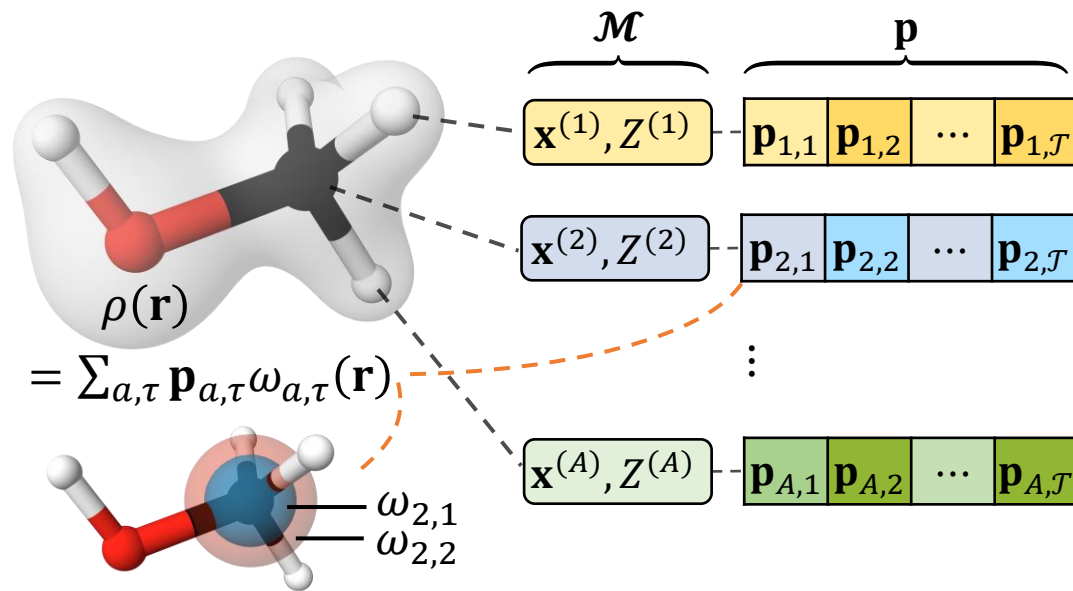
# M-OFDFT: the $T_S[\rho]$ Model

- Density representation using **atomic basis**:  
Expansion coefficients  $\mathbf{p}$  on atomic basis:  $\rho$  represented as  $(\mathbf{p}, \mathcal{M})$ .



# M-OFDFT: the $T_S[\rho]$ Model

- Density representation using **atomic basis**:  
Expansion coefficients  $\mathbf{p}$  on atomic basis:  $\rho$  represented as  $(\mathbf{p}, \mathcal{M})$ .



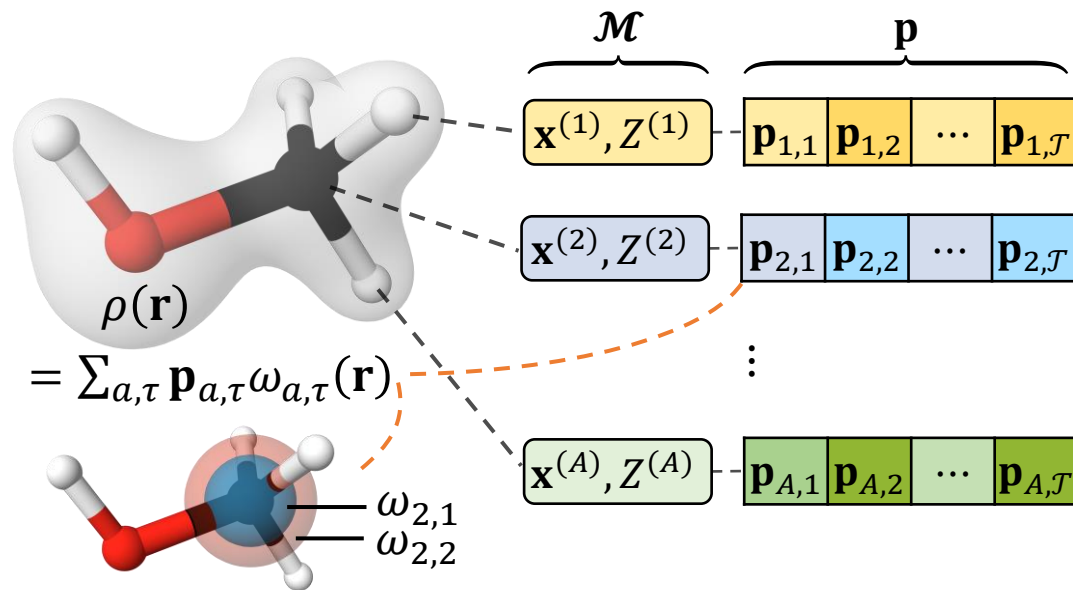
# M-OFDFT: the $T_S[\rho]$ Model

- Density representation using **atomic basis**:

Expansion coefficients  $\mathbf{p}$  on atomic basis:  $\rho$  represented as  $(\mathbf{p}, \mathcal{M})$ .

Electrons distribute around atoms, cusp condition, shell structure, ...

→ ~1000 times less dimensions than grid.



# M-OFDFT: the $T_S[\rho]$ Model

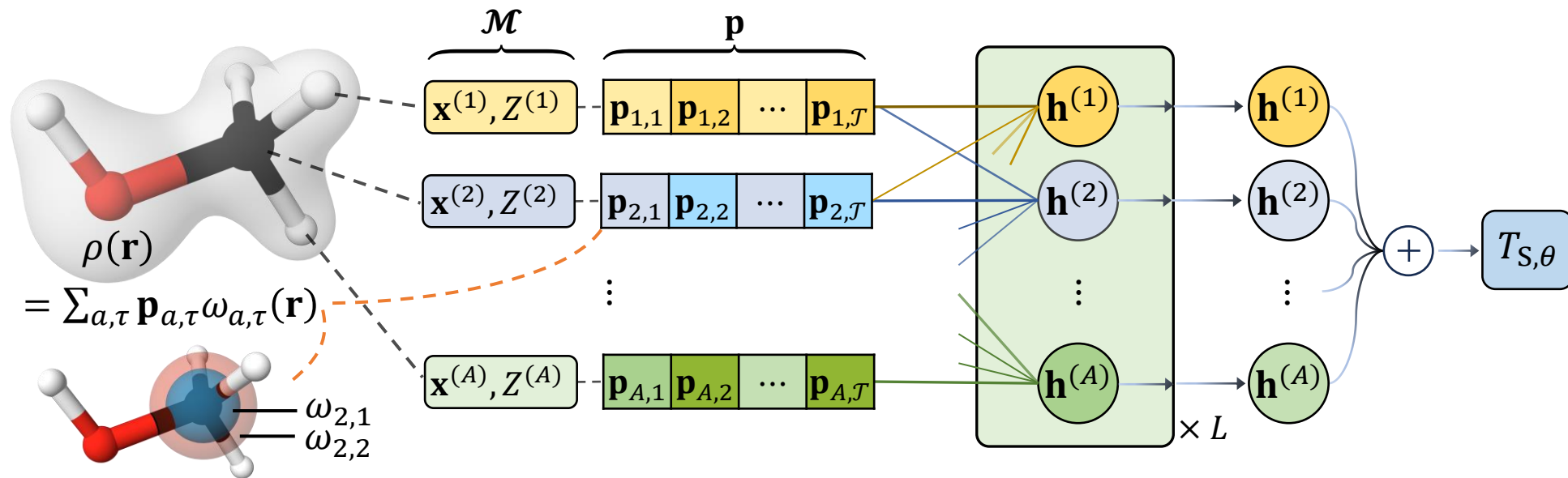
- Density representation using **atomic basis**:

Expansion coefficients  $\mathbf{p}$  on atomic basis:  $\rho$  represented as  $(\mathbf{p}, \mathcal{M})$ .

Electrons distribute around atoms, cusp condition, shell structure, ...

→ ~1000 times less dimensions than grid.

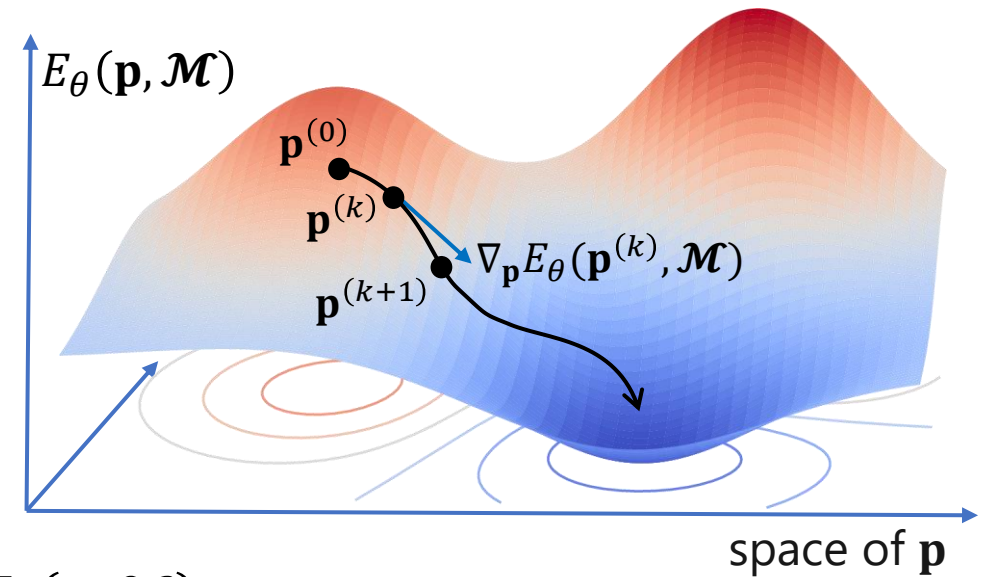
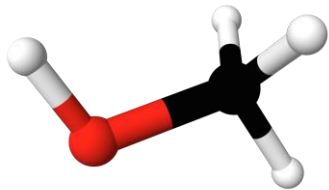
- Explicit **nonlocal** model: Graphormer (Transformer-like model).



# M-OFDFT Workflow

- Density optimization by gradient descent.

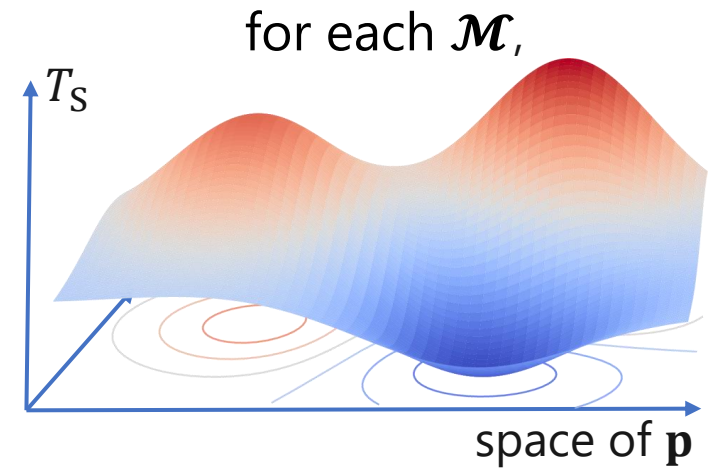
molecular structure  $\mathcal{M}$



$$E_{\theta}(\mathbf{p}, \mathcal{M}) := T_{S,\theta}(\mathbf{p}, \mathcal{M}) + E_{\text{H}}(\mathbf{p}, \mathcal{M}) + E_{\text{XC}}(\mathbf{p}, \mathcal{M}) + E_{\text{ext}}(\mathbf{p}, \mathcal{M})$$

# Challenges in Learning

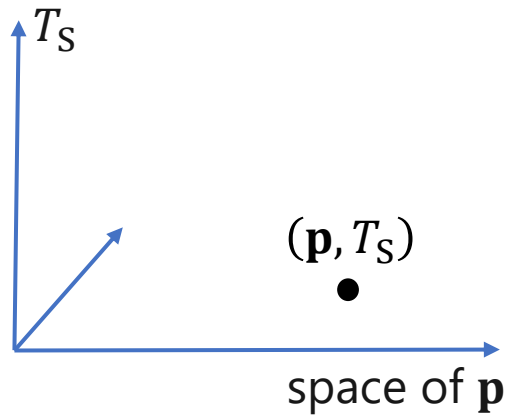
- Learning an objective: capture the whole landscape.



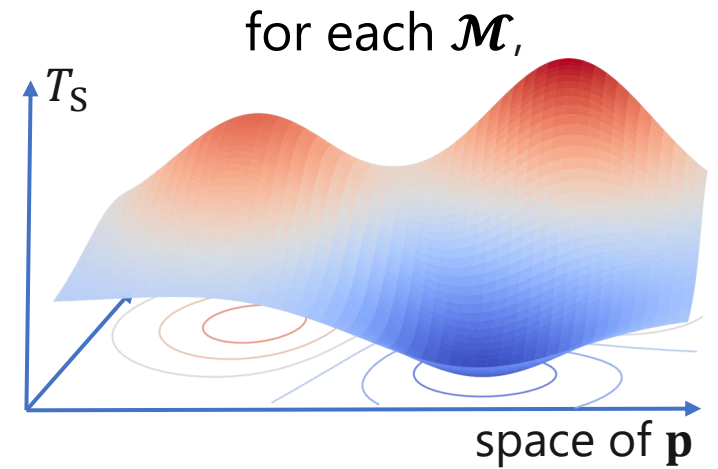


# Challenges in Learning

- Learning an objective: capture the whole landscape.
  - One datapoint is not enough.

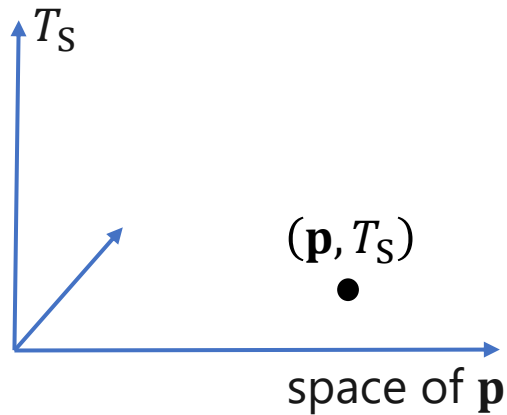


from KSDFT converged solution

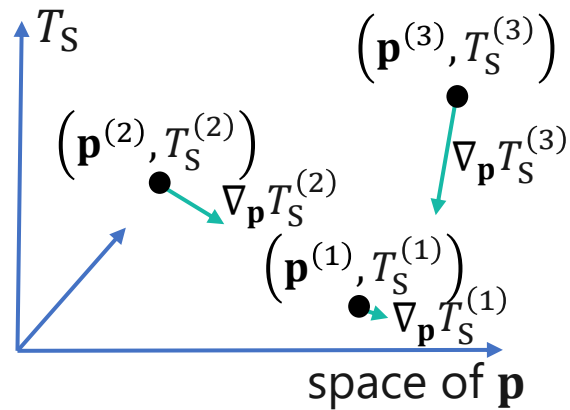


# Challenges in Learning

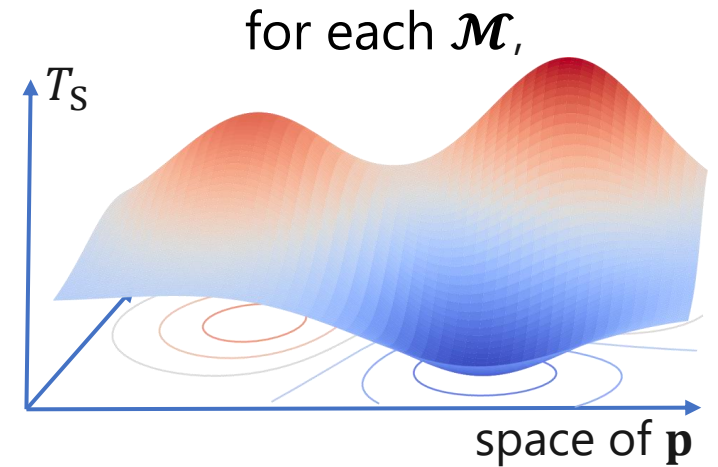
- Learning an objective: capture the whole landscape.
  - One datapoint is not enough.
  - Generate **multiple datapoints** each with a **gradient label**.



from KSDFT converged solution

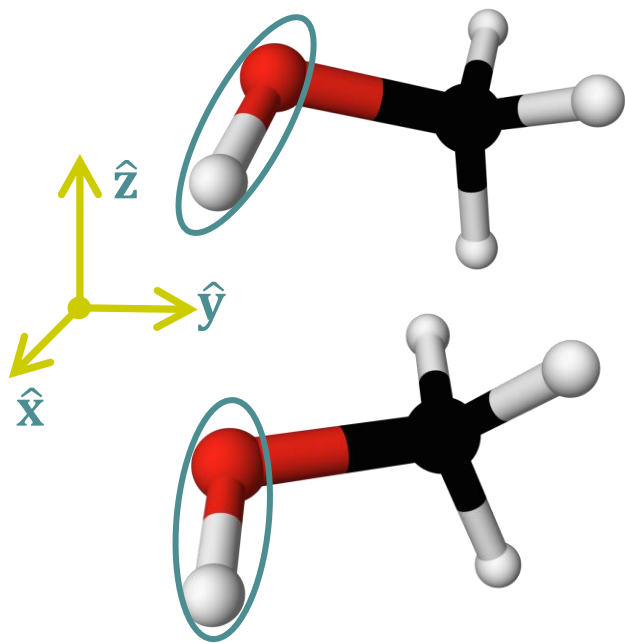


from each of KSDFT iterations



# Challenges in Learning

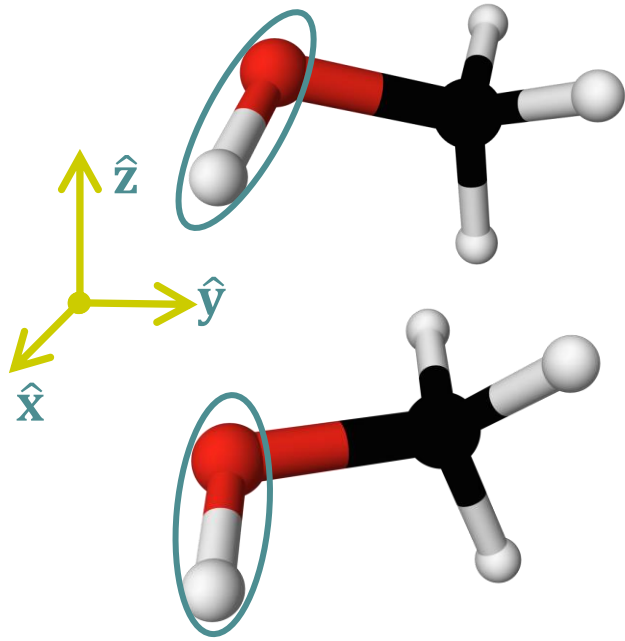
- Geometric invariance:  $T_S$  does not change with the rotation of molecule.



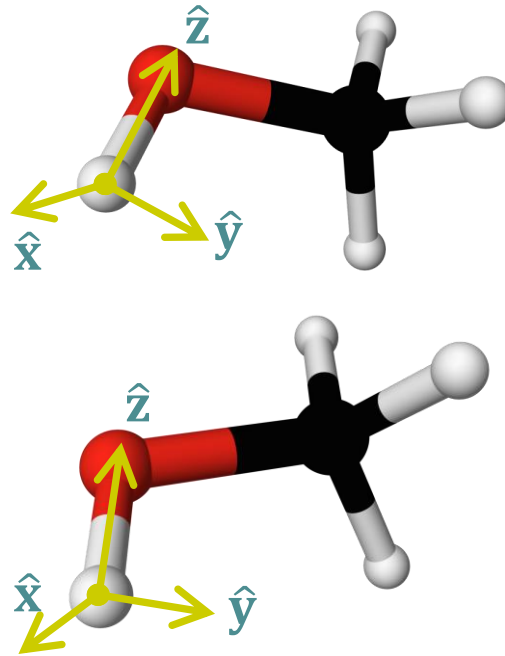
$\mathbf{p}$  rotates with molecule.

# Challenges in Learning

- Geometric invariance:  $T_S$  does not change with the rotation of molecule.



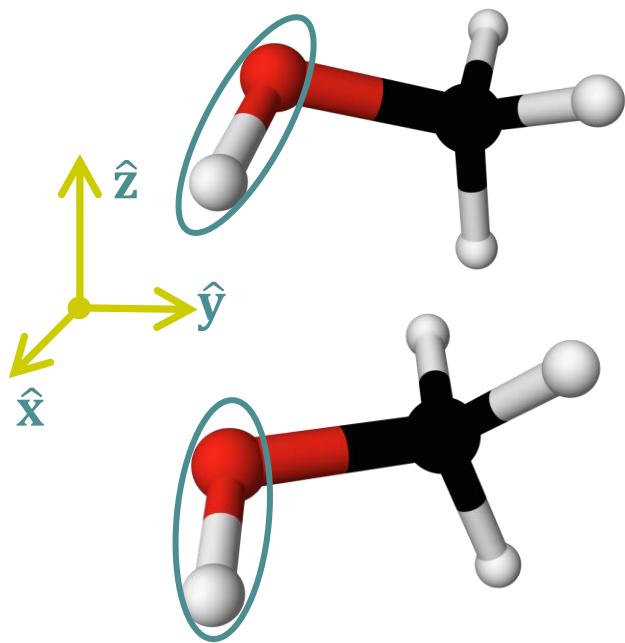
**p** rotates with molecule.



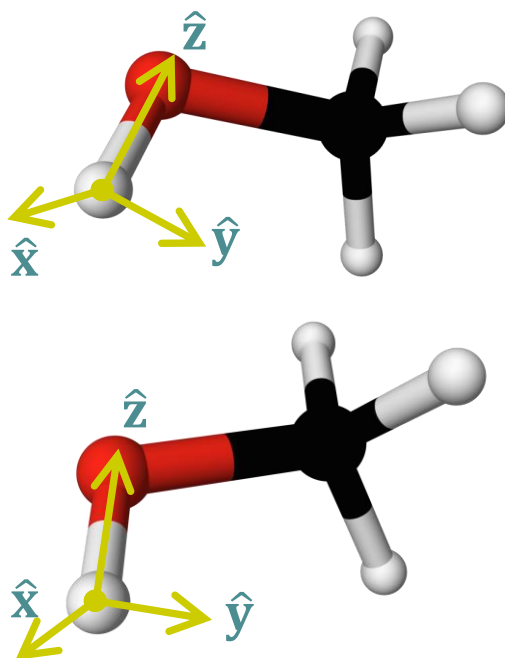
**Local frame:**  
**p** remains the same.

# Challenges in Learning

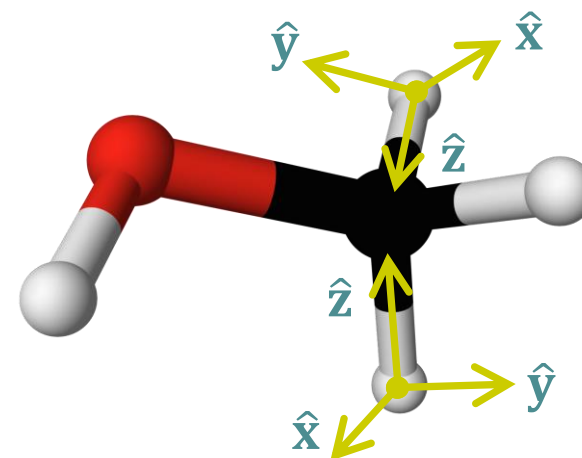
- Geometric invariance:  $T_S$  does not change with the rotation of molecule.



$\mathbf{p}$  rotates with molecule.



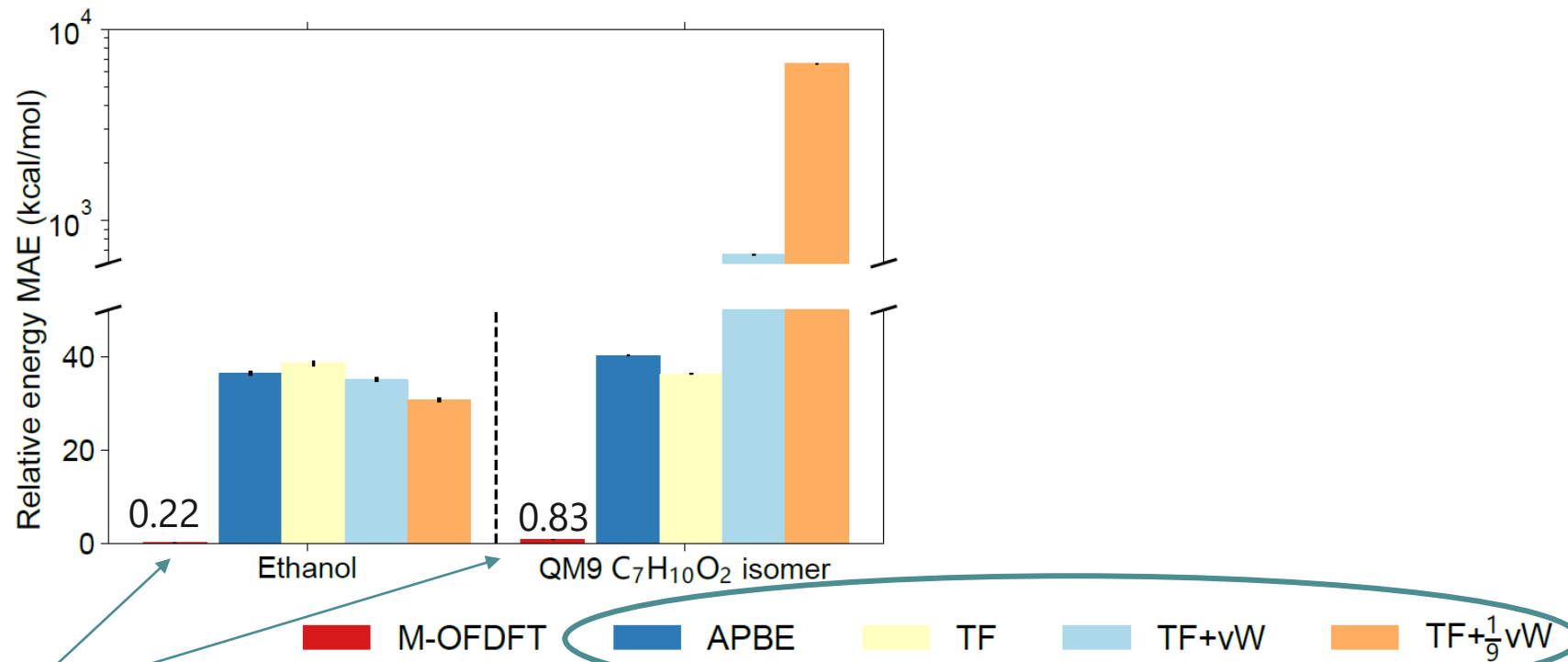
**Local frame:**  
 $\mathbf{p}$  remains the same.



**Additional benefit:**  
similar local structures  
have similar  $\mathbf{p}$

# Better Accuracy than Classical OFDFT

- Energy and force:

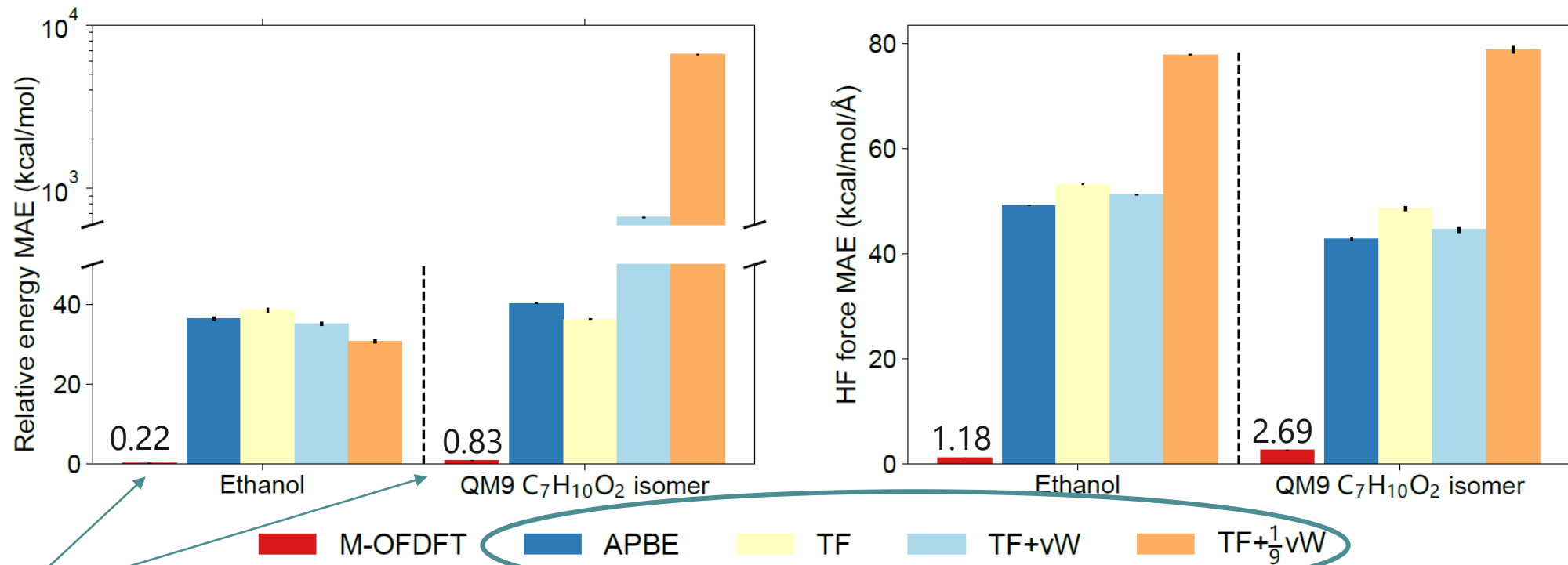


Within chemical accuracy  
(1 kcal/mol) to KSDFT

Classical OFDFT

# Better Accuracy than Classical OFDFT

- Energy and force:

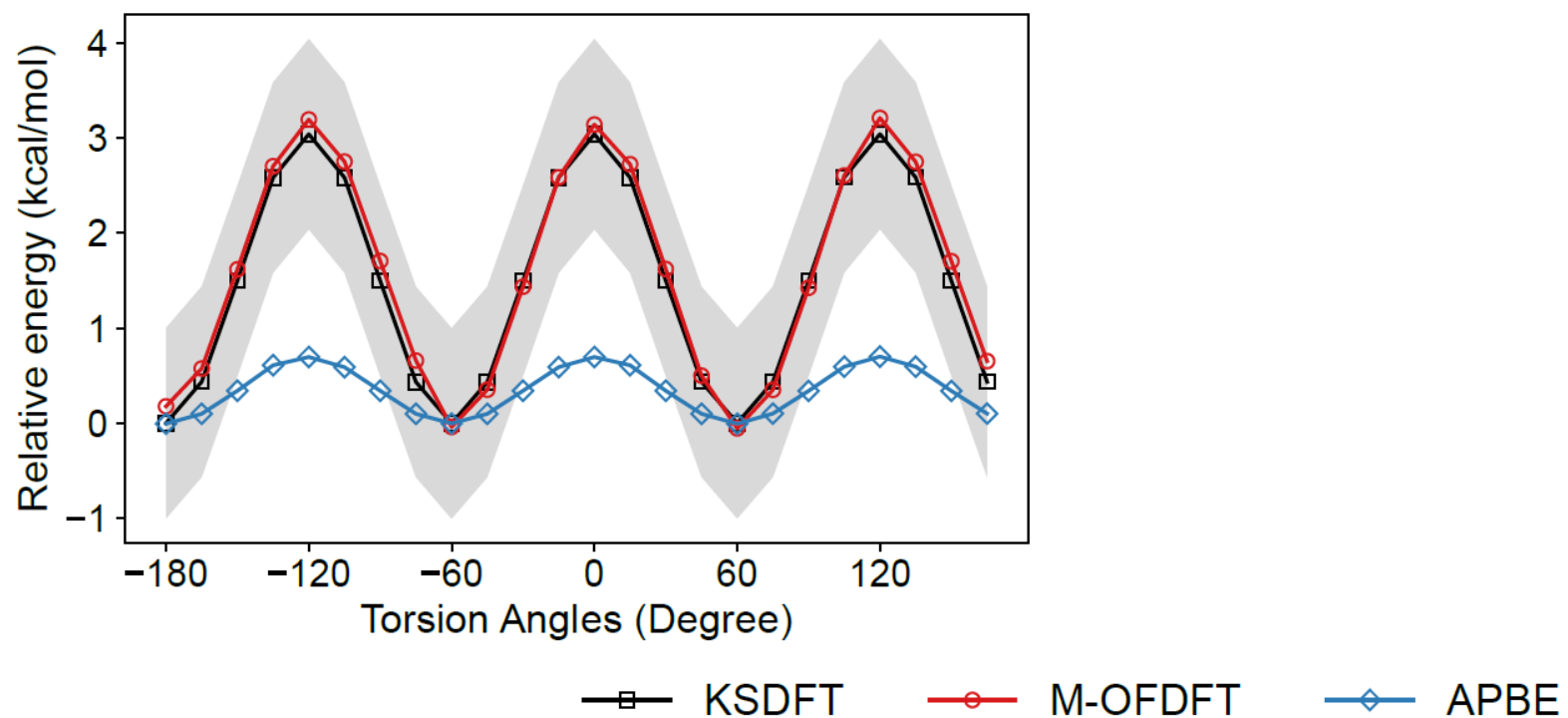
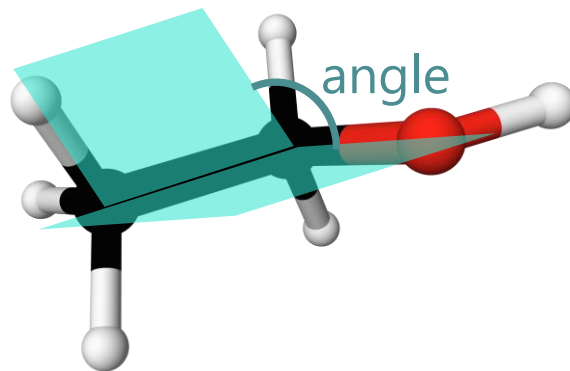


Within chemical accuracy  
(1 kcal/mol) to KSDFT

Classical OFDFT

# Better Accuracy than Classical OFDFT

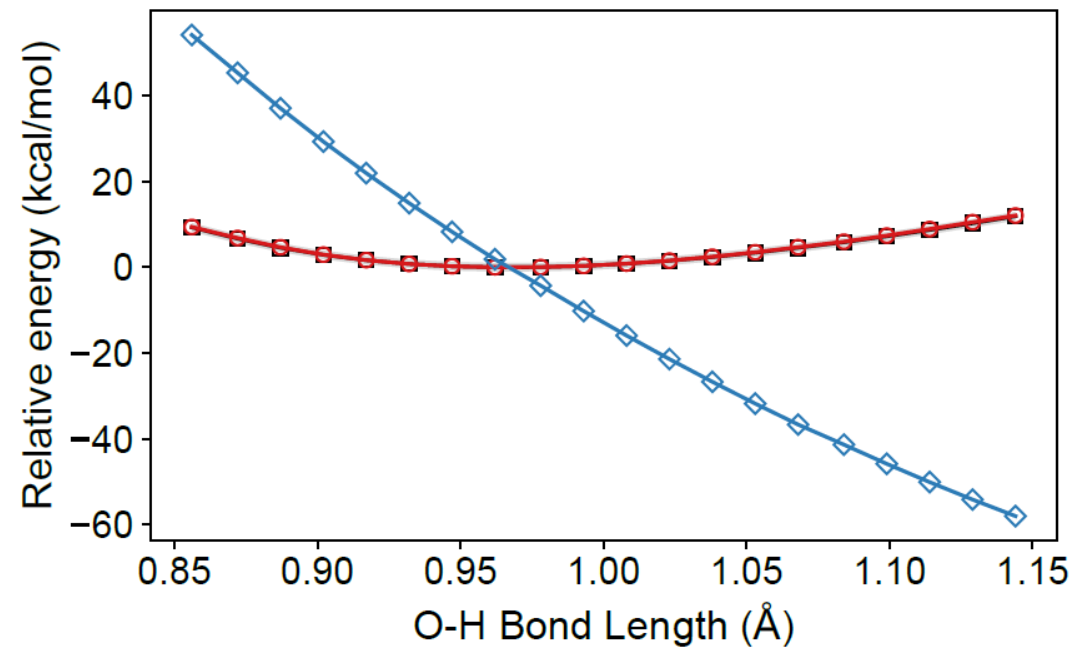
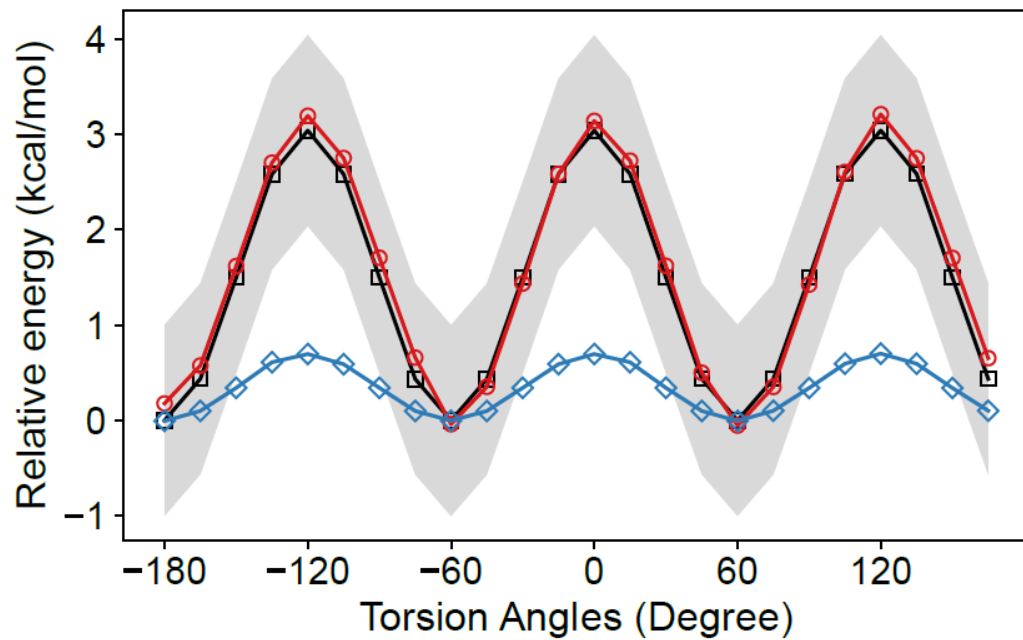
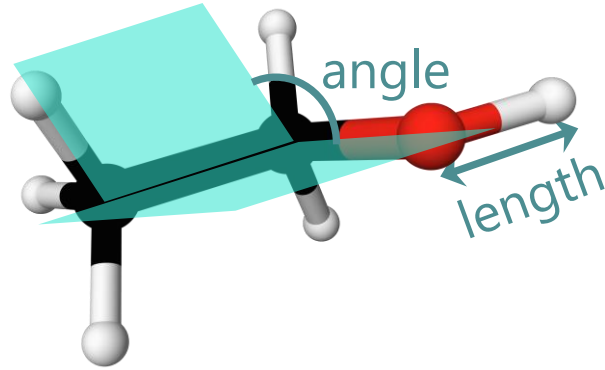
- Energy surface:





# Better Accuracy than Classical OFDFT

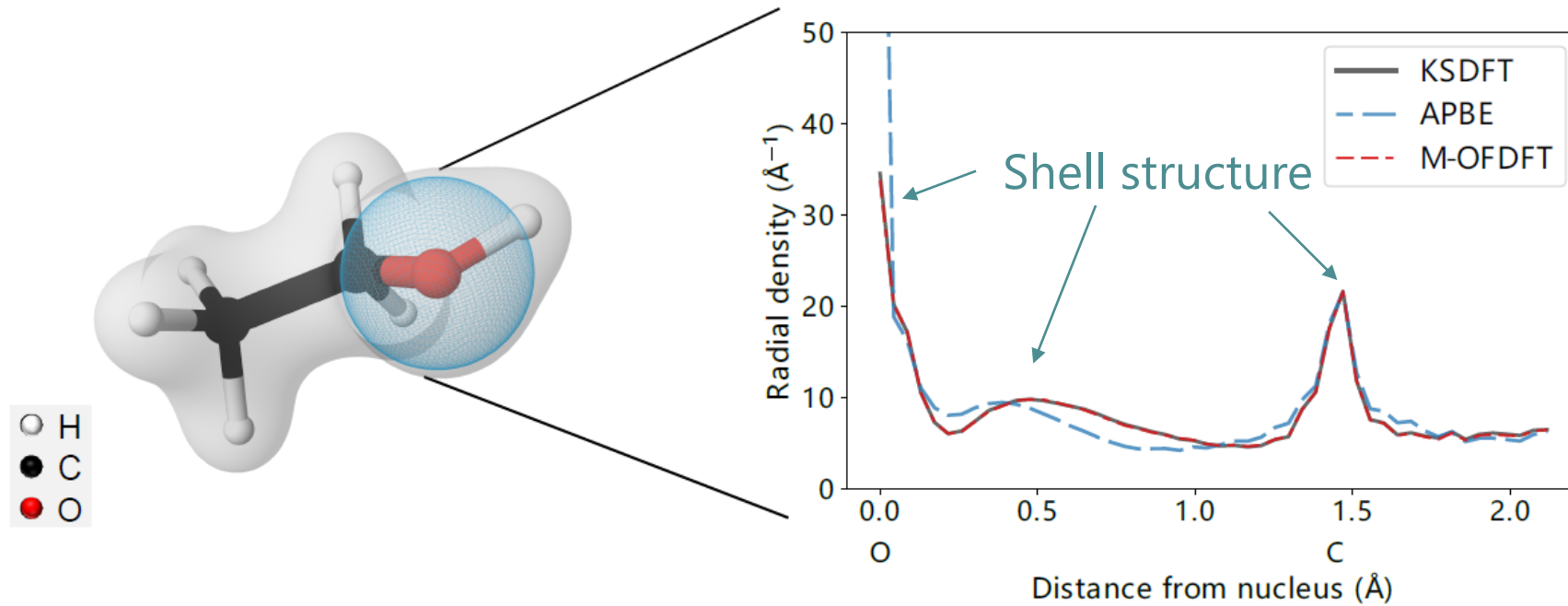
- Energy surface:



—□— KSDFT    —○— M-OFDFT    —◇— APBE

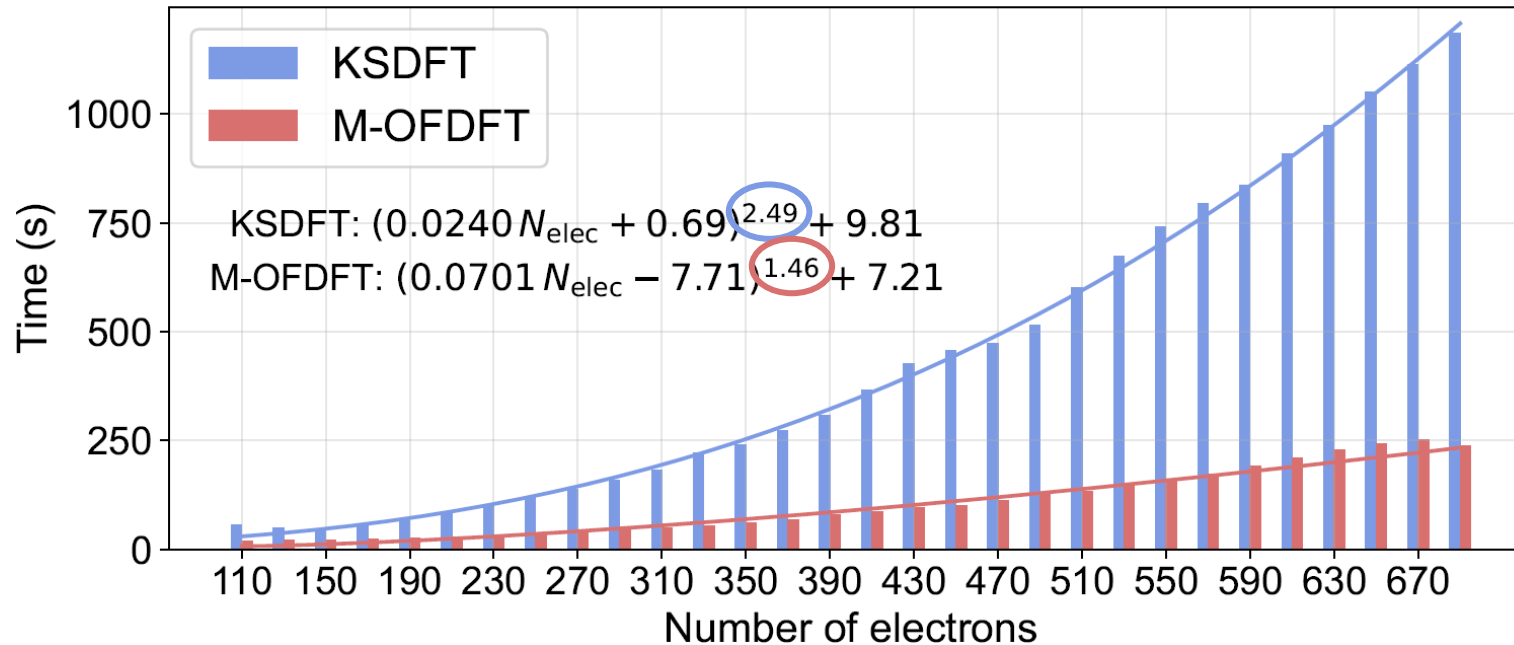
# Better Accuracy than Classical OFDFT

- Electron density:



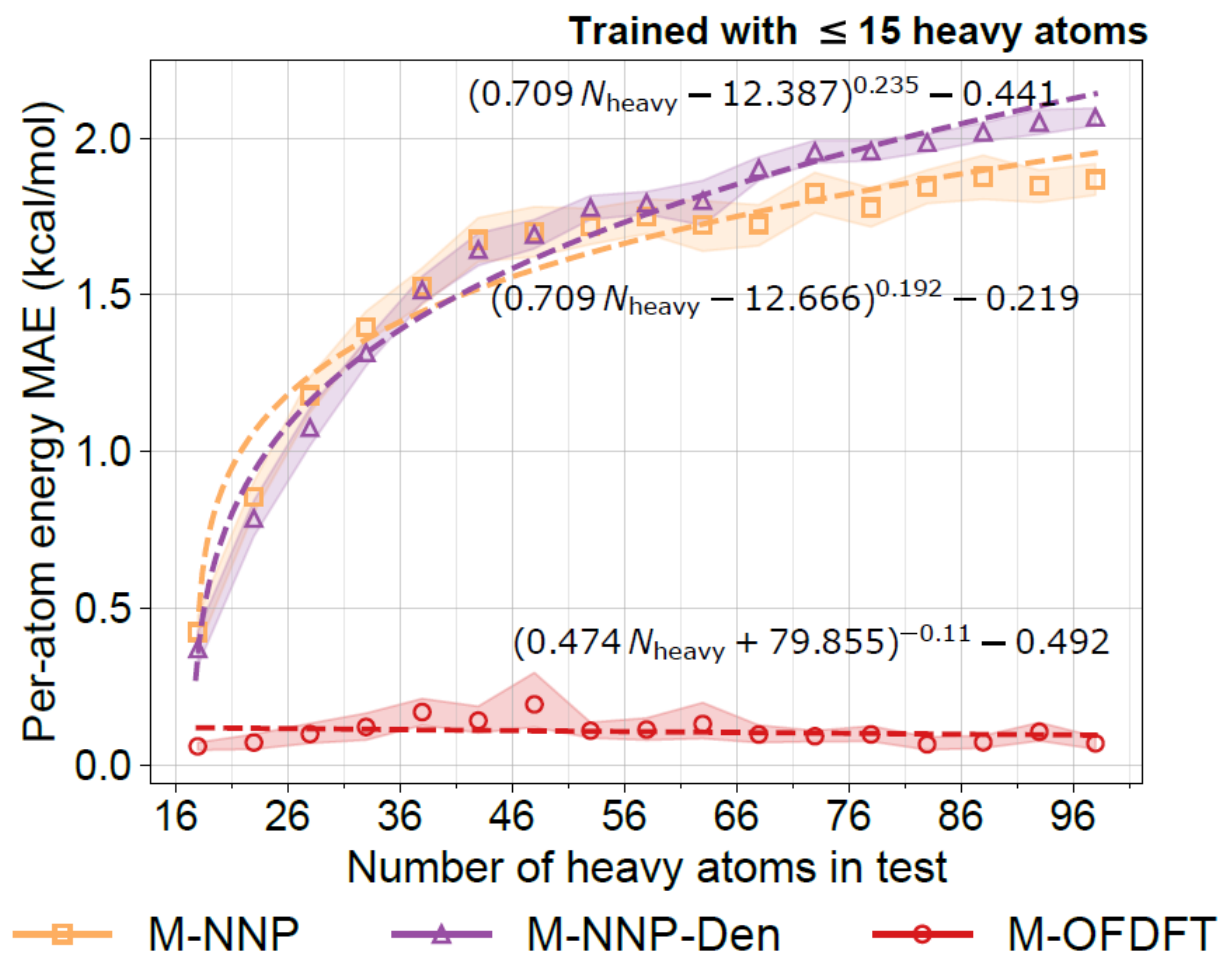
# Better Scalability than KSDFT

- Lower empirical cost scaling:



- Protein (PDB ID: 1PRB): 738 atoms, 2750 electrons  
→ 27.4-fold speed up (0.45 h vs. 12.3 h).

# Better Extrapolability than NN Potential



# In the Future

- Large-scale molecular dynamics simulation.
- Quantum embedding for multi-scale calculation.
- Towards universality by:
  - Large data and model.
  - Mathematical properties.

**Thank You!**