Project concept

Energy prediction model using ML with specification : **4** methods

1) Direct: $\boldsymbol{E - E_0 = \Delta_{ML}}$
2) Harmonic: $\boldsymbol{E - E_0 = E_{har} + \Delta_{ML}}$
3) Normal: $\boldsymbol{E - E_0 = \sum_i E_i^{nor}(\chi_i) + \Delta_{ML}}$
4) Internal: $\boldsymbol{E - E_0 = \sum_i E_i^{int}(\chi_i) + \Delta_{ML}}$

$\Delta_{ML} = \text{Training energy data}$

$E - E_0 = \text{Molecular internal energy (from first} - \text{principle calculation, CCSD(T))}$

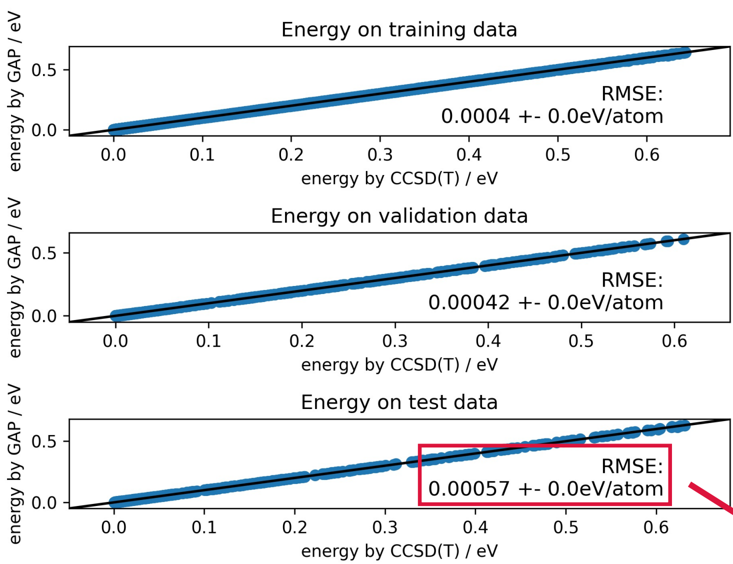
```
Lattice="15.0 0.0 0.0 0.0 0.0 15.0 0.0 0.0 0.0 15.0" Properties=species:S:1:pos:R:3:forces:R:3 energy=0.0122504762 pbc="T T T"
O        0.00000000       0.00000000       0.00000000      -0.26348127       0.98232935       0.00000000
H        0.95454173       0.00000000       0.00000000       0.32589082      -0.21701854       0.00000000
H       -0.19271003       0.95610648       0.00000000      -0.06240955      -0.76531081       0.00000000
```

```
Lattice="15.0 0.0 0.0 0.0 15.0 0.0 0.0 0.0 15.0" Properties=species:S:1:pos:R:3 energy=-0.0008384372 pbc="T T T"
O         0.00000000         0.00000000         0.00000000
H         0.95454173         0.00000000         0.00000000
H        -0.19271003         0.95610648         0.00000000
```

Method 1: Direct

$$\boldsymbol{E - E_0 = \Delta_{ML}}$$

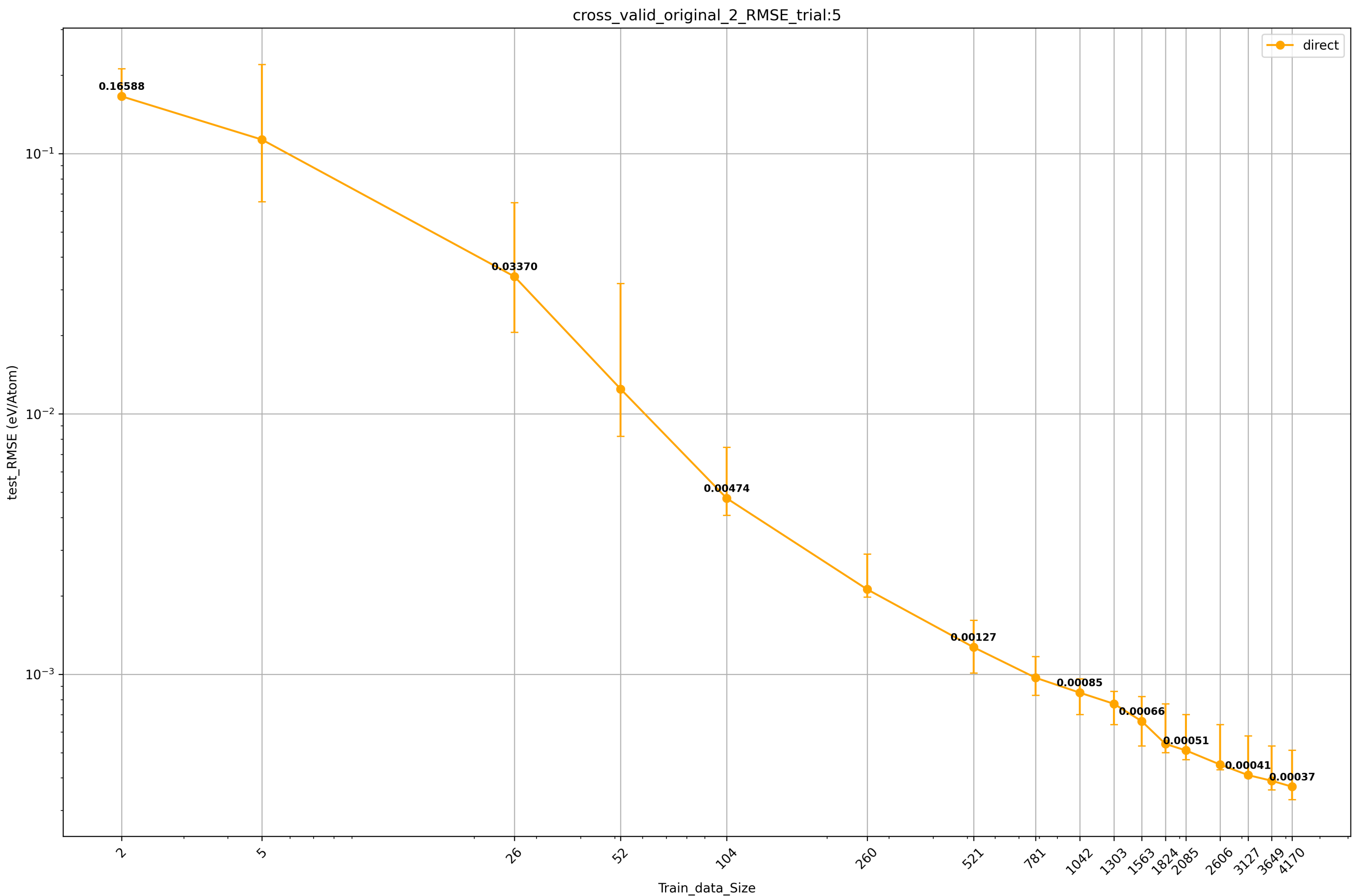
Δ_{ML} is just CCSD(T) data. Using GAP without specification.
Since we take energy data as it is, ML result form this method becomes the **reference**.



ML Train Δ_{ML}

Using trained ML potential
to predict energy for Test set

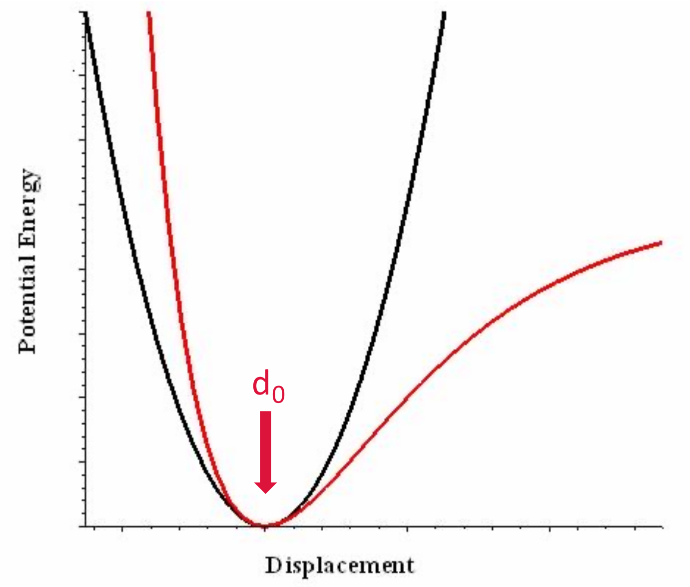
Test set's RMSE indicates the accuracy of the ML model.
The smaller it is, the more accurate the model is.



(Benchmark)Direct: $E - E_0 = \Delta_{ML}$

Method 2: Harmonic

$$\widehat{H}_{el}\psi_{el} = U * \psi_{el} \quad ; U(\{\mathrm{r}_\alpha\})$$



Taylor series expansion

$$\emptyset(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

$$u = u(d_0) + \underbrace{\cancel{\frac{\partial u(d_0)}{\partial d}}}_{}^{0}(d - d_0) + \frac{1}{2!}\frac{\partial^2 u}{\partial d^2}(d - d_0)^2 + \cdots$$

$$u = u(d_0) + \frac{1}{2!}\boxed{\frac{\partial^2 u}{\partial d^2}}(d - d_0)^2 + \Delta^{ML}$$

Hessian Matrix = force constant

Hessian: an n x n square matrix composed of the second-order partial derivatives of a function of n variables

Method 2: Harmonic

$$\boldsymbol{E - E_0 = E_{har} + \Delta_{ML}}$$

Δ_{ML} is the difference between CCSD(T) data and Harmonic Energy.
Δ_{ML} = Single-mode anharmonicity + Anharmonic coupling

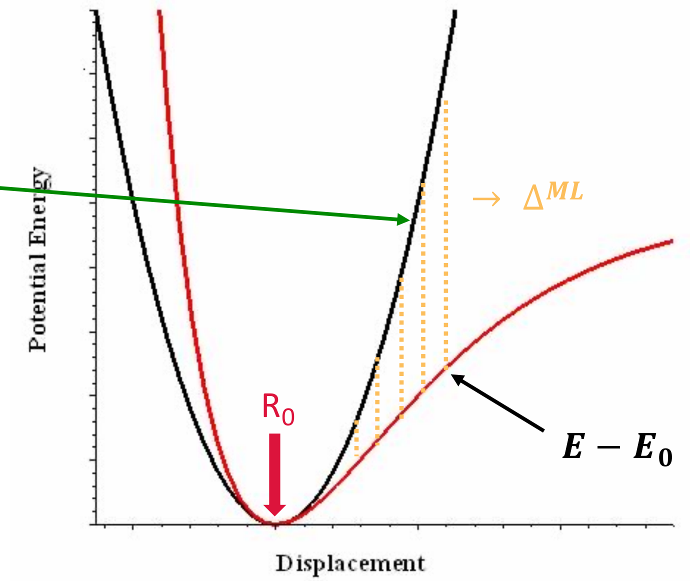
$$E_{har} = \frac{1}{2} \overset{1 \times 3N}{(x_1 \cdots x_{3N})} \overset{3N \times 3N}{\begin{pmatrix} \frac{\partial^2 U}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 U}{\partial x_1 \partial x_{3N}} \\ \cdots & \ddots & \cdots \\ \frac{\partial^2 U}{\partial x_{3N} \partial x_1} & \cdots & \frac{\partial^2 U}{\partial x_{3N} \partial x_{3N}} \end{pmatrix}} \overset{3N \times 1}{\begin{pmatrix} x_1 \\ \cdots \\ x_{3N} \end{pmatrix}} = \boxed{\frac{1}{2} x^T H_U x}$$

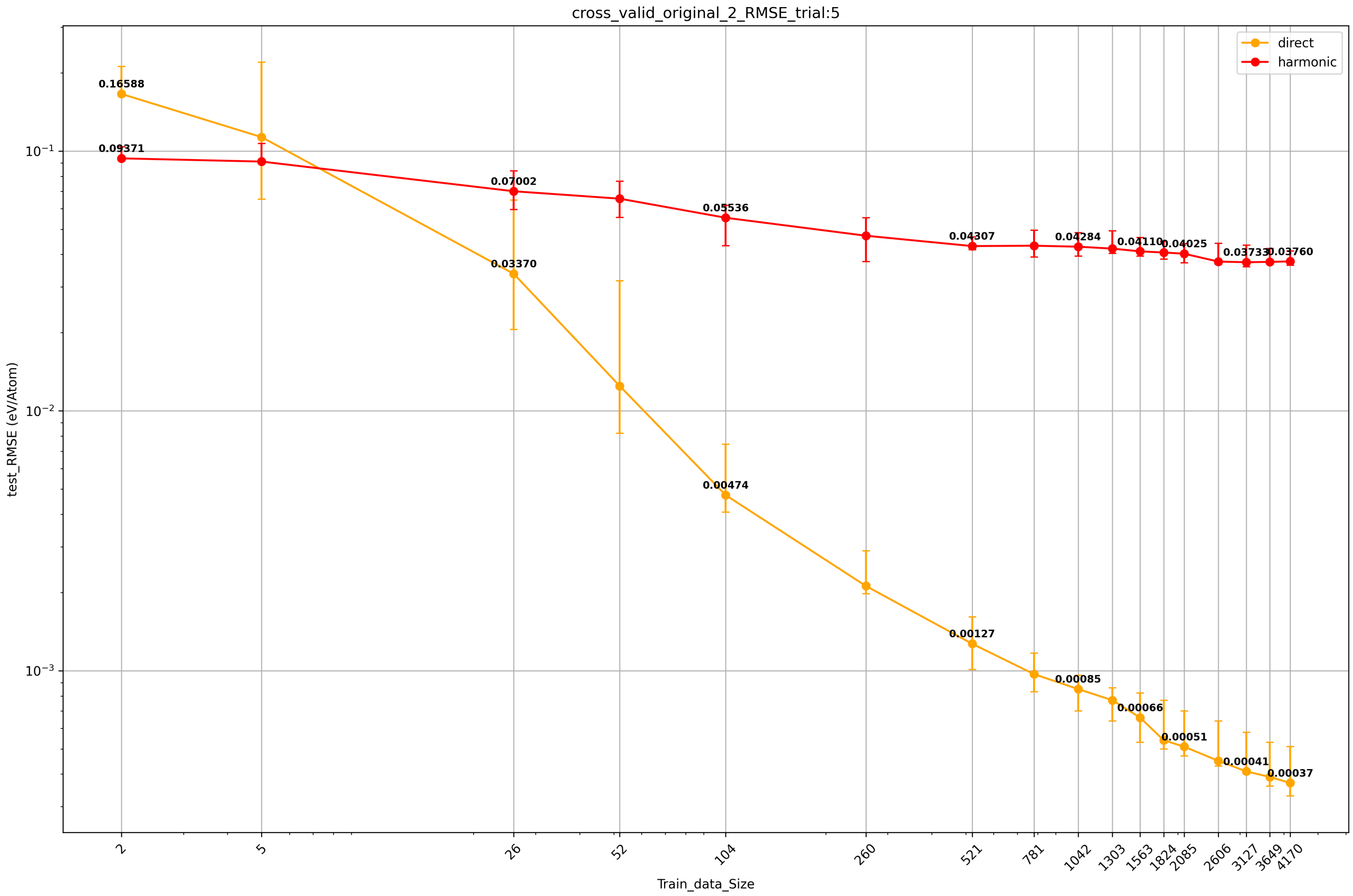
Harmonic Energy

transposed displacement vector

displacement vector

Non-mass weighted Hessian matrix(cartesian coordinate) -> force constant





Harmonic: $E - E_0 = E_{har} + \Delta_{ML}$

(Benchmark)Direct: $E - E_0 = \Delta_{ML}$

고려대학교 KOREA UNIVERSITY

Method 3: Normal

$< Qi|Qj > = \delta_{ij}$

Displacement vectors are orthonormalized

$$E - E_0 = \sum_i E_i^{nor}(\chi_i) + \Delta_{ML}$$

Δ_{ML} is the difference between CCSD(T) data and sum of Interpolated Energy from each **normal** mode .
Δ_{ML} = Anharmonic coupling

For water, $E - E_0 = E_1^{nor} + E_2^{nor} + E_3^{nor} + \Delta_{ML}$

initial coordinates

Displacement vector (assigned to normal mode)

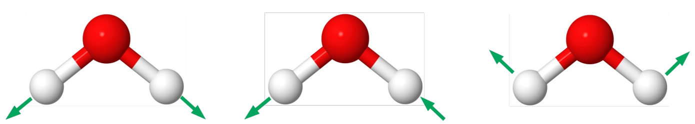
$$R = R_0 + \sum_{i=1}^{3N-6} c_i Q_i$$

current coordinates

coefficient

For water,
$R - R_0 = c_1Q_1 + c_2Q_2 + c_3Q_3$

3 Vibrational modes
Normal modes



symmetric stretching | asymmetric stretching | bending

~3650 cm-1 | ~3750 cm-1 | ~1600 cm-1

e.g.,

$$
\begin{aligned}
& < R - R_0|Q_1 > \\
&=< c_1Q_1 + c_2Q_2 + c_3Q_3|Q_1 > \\
&= c_1 < Q_1|Q_1 > \\
&+c_2 < Q_2|Q_1 > \\
&+c_3 < Q_3|Q_1 > \\
&= c_1 * 1 + c_2 * 0 + c_3 * 0 \\
&= c_1
\end{aligned}
$$

$\longrightarrow$

$< R - R_0|Q_1 > = c_1$
$< R - R_0|Q_2 > = c_2$
$< R - R_0|Q_3 > = c_3$

Method 3: Normal

coefficient

Displacement vector

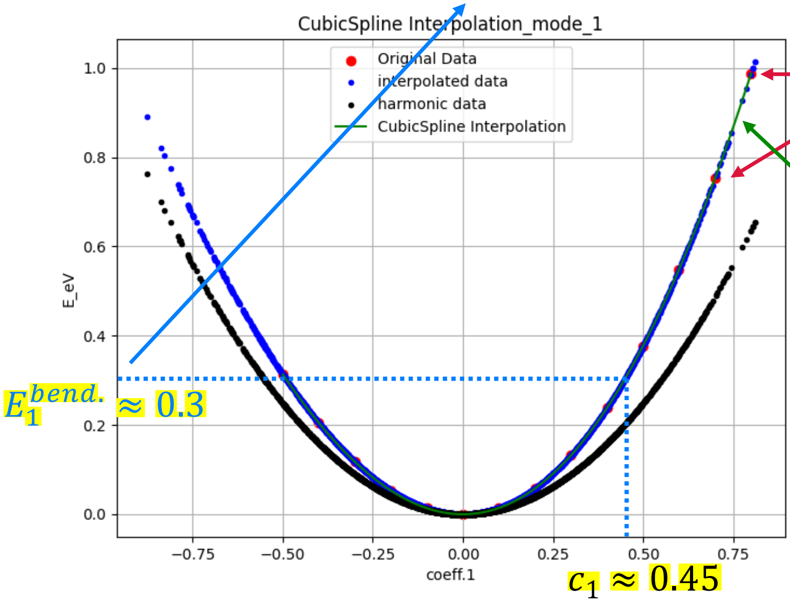
$R - R_0 = c_1 Q_1 + c_2 Q_2 + c_3 Q_3$

initial coordinates

$$\boldsymbol{E - E_0 = \sum_i E_i^{nor}(\chi_i) + \Delta_{ML}}$$

Δ_{ML} is the difference between CCSD(T) data and sum of Interpolated Energy from each **normal** mode .
Δ_{ML} = Anharmonic coupling

For H_2O, $\boldsymbol{E - E_0 = E_1^{bend.} + E_2^{sym.stretch.} + E_3^{asym.stretch.} + \Delta_{ML}}$

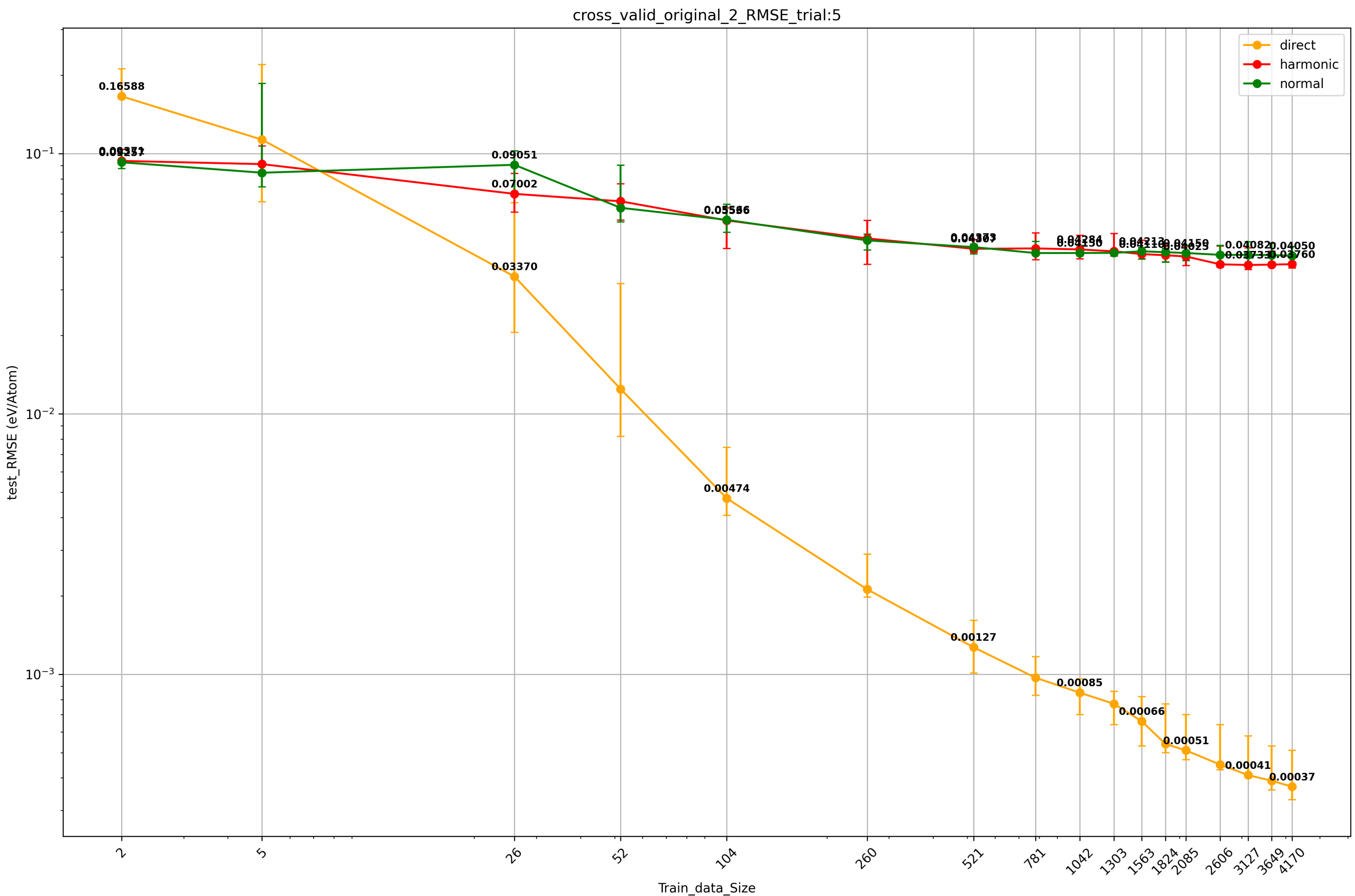


Energy from
single-mode distorted coordinates (e.g., $R = R_0 + c_1 Q_1$)
for interpolation

c: -0.5 *to* 0.8 (*increment*: 0.1)

CubicSpline interpolation

Step
1. Getting CCSD(T) energy for interpolation
2. Interpolation for each-normal mode
3. Getting $E_1^{nor} + E_2^{nor} + E_3^{nor}$ using c_1 ,c_2, c_3
4. Calculate $\boldsymbol{\Delta_{ML}}$

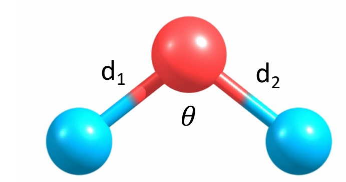


Normal: $E - E_0 = \sum_i E_i^{nor}(\chi_i) + \Delta_{ML}$

Harmonic: $E - E_0 = E_{har} + \Delta_{ML}$

(Benchmark)Direct: $E - E_0 = \Delta_{ML}$

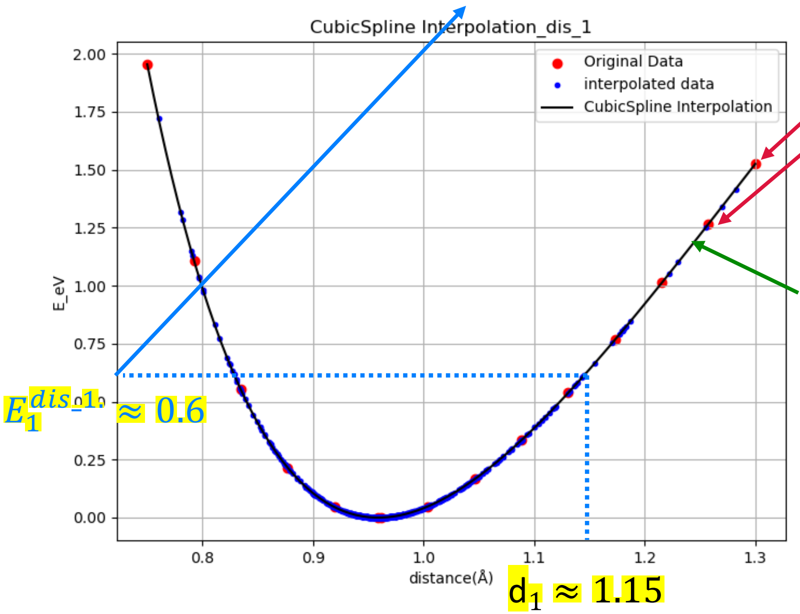
Method 4: Internal



$$\boldsymbol{E - E_0 = \sum_i E_i^{int}(\chi_i) + \Delta_{ML}}$$

Δ_{ML} is the difference between CCSD(T) data and sum of Interpolated Energy from each **internal** mode .
Δ_{ML} = Harmonic coupling + Anharmonic coupling

For H_2O, $\boldsymbol{E - E_0 = E_1^{dis\_1} + E_2^{dis\_2} + E_3^{ang} + \Delta_{ML}}$



Energy from
single-internal mode distorted coordinates ($distance, angle$)
for interpolation

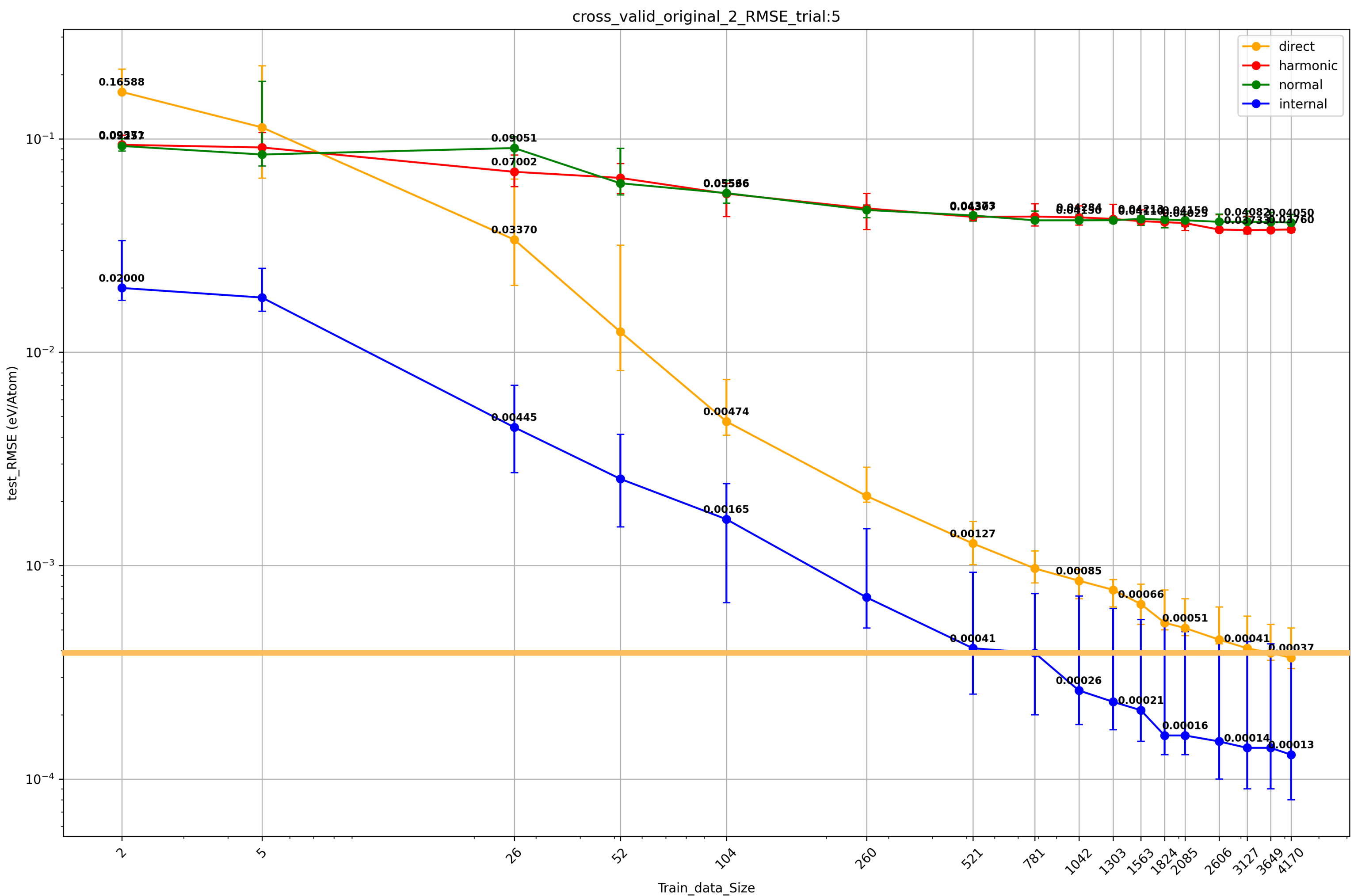
CubicSpline interpolation

Step
1. Getting CCSD(T) energy for interpolation
2. Interpolation for each-internal mode
3. Getting $E_1^{dis\_1} + E_2^{dis\_2} + E_3^{ang}$ using d_1 ,d_2, θ
4. Calculate $\boldsymbol{\Delta_{ML}}$

ML result(original data)

Original data sets: Train(4170 structures) / Valid(521 structures) / Test(521 structures)

Hyperparameter Bayesian optimization was proceeded for each method to be fair



Normal: $E - E_0 = \sum_i E_i^{nor}(\chi_i) + \Delta_{ML}$

Harmonic: $E - E_0 = E_{har} + \Delta_{ML}$

Internal method

showed much better performance than Benchmark!

(Benchmark)Direct: $E - E_0 = \Delta_{ML}$

Internal: $E - E_0 = \sum_i E_i^{int}(\chi_i) + \Delta_{ML}$