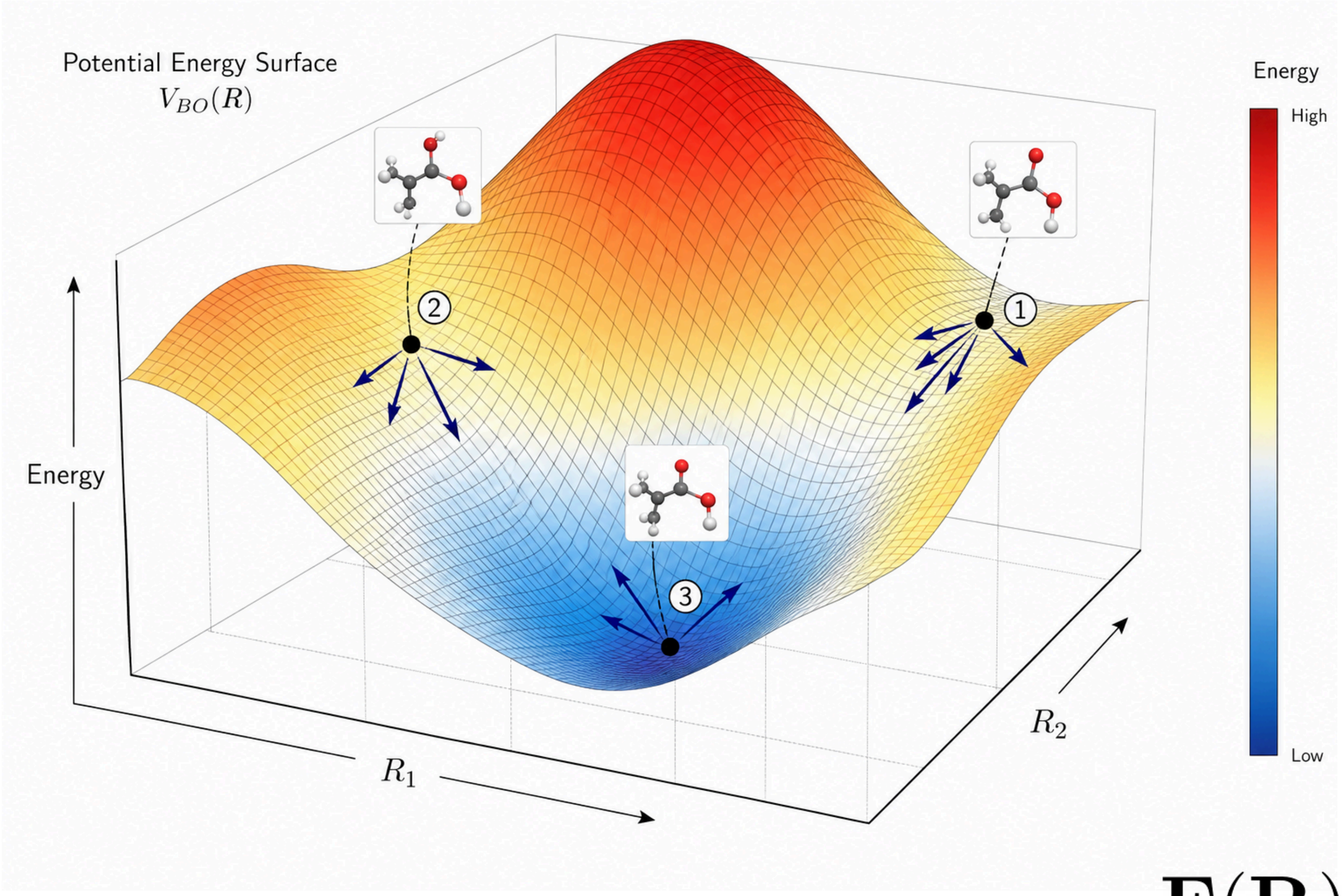
CUHK ZHANG LAB : JOURNAL CLUB

GDML:
Learning Molecular Force Fields in the Gradient Domain

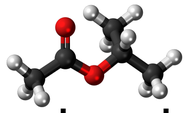
Jiseung Hong · August 6th 2026

GDML: Learning Molecular Force Fields in the Gradient Domain



$$\mathbf{F}(\mathbf{R}) = -\nabla_{\mathbf{R}} V_{\mathrm{BO}}(\mathbf{R})$$

The Computational-Chemistry Objects : BO-PES



Total molecular vector : $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \in \mathbb{R}^{3N}$

$r_{\text{atom } 1,x}$

$$\hat{H}\psi(\mathbf{R}) = E\psi(\mathbf{R})$$

BO approximated Potential Energy : $V_{\text{BO}}(\mathbf{R})$

The Computational-Chemistry Objects : location vector to Force vector

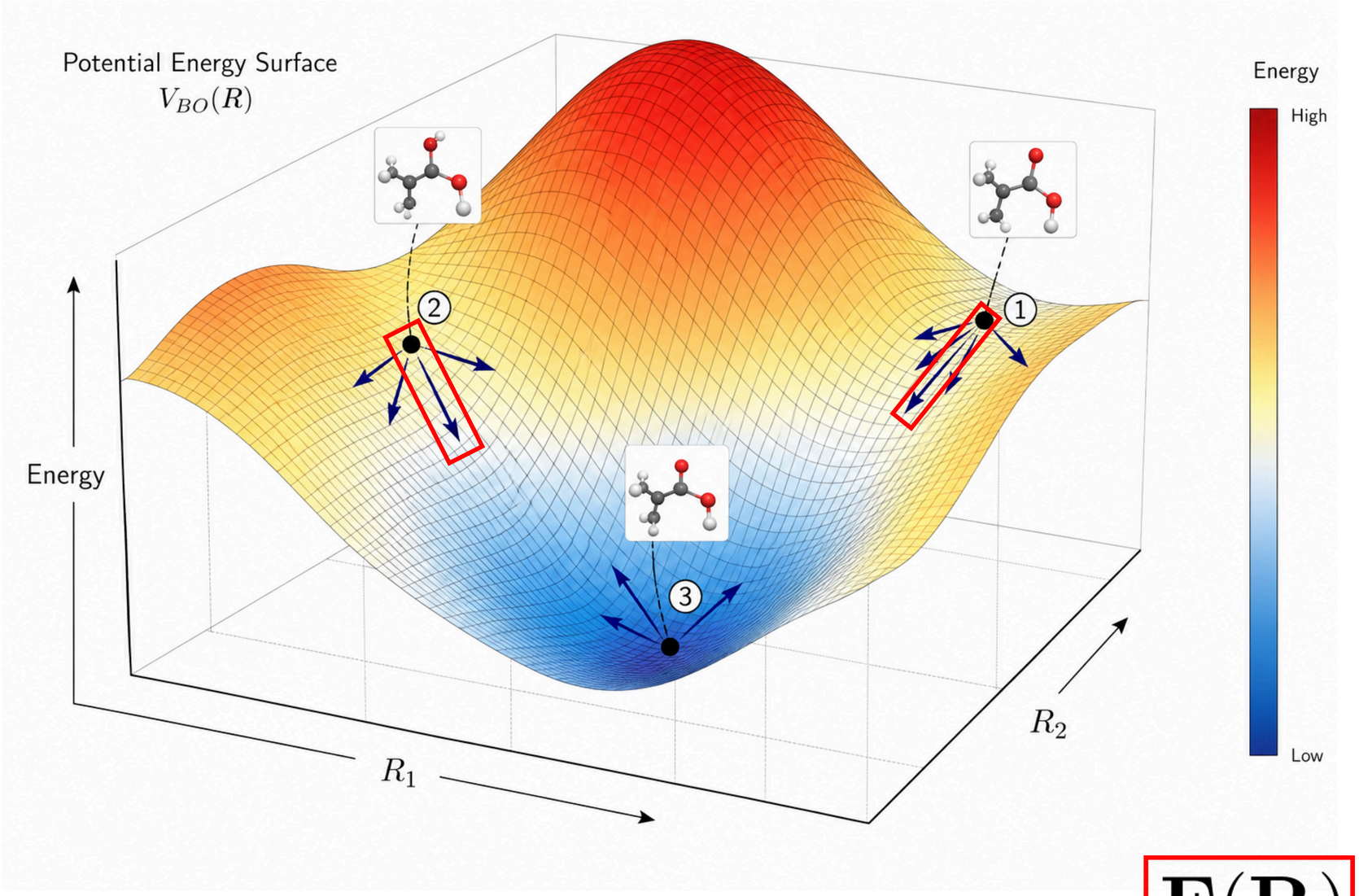
Effective Force Vector :

BO approximated Potential Energy

$$\mathbf{F}_i(\mathbf{R}) = -\nabla_{\mathbf{r}_i} V_{\mathrm{BO}}(\mathbf{R})$$

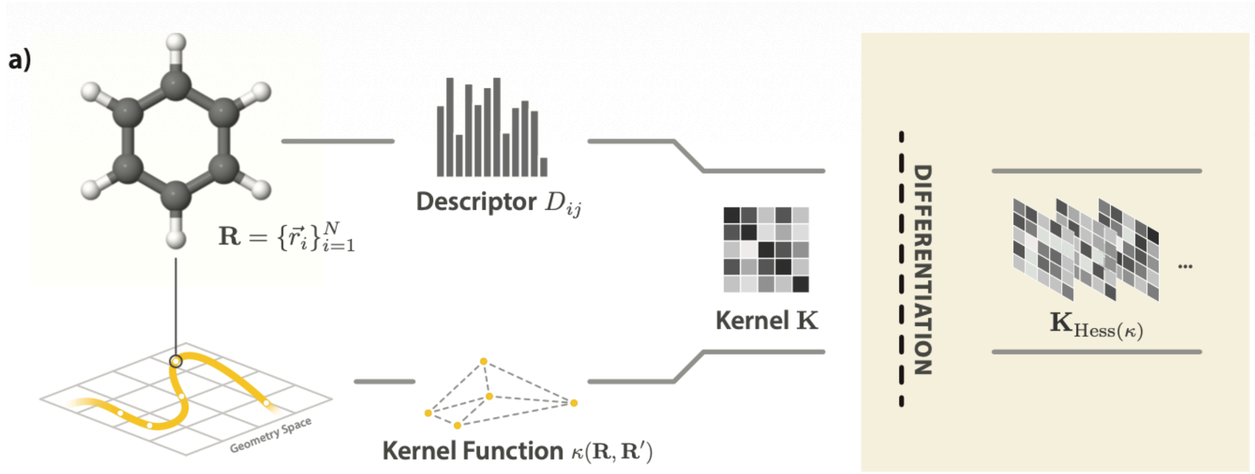
negative gradient operator

The Computational-Chemistry Objects : location vector to Force vector



$$\mathbf{F}(\mathbf{R}) = -\nabla_{\mathbf{R}} V_{\mathrm{BO}}(\mathbf{R})$$

How GDML Constructs a Conservative Vector Field



Descriptor encodes molecular structure.

$$D_{ij} = \begin{cases} \|\vec{r}_i - \vec{r}_j\|^{-1} & \text{for } i > j \\ 0 & \text{for } i \leq j \end{cases}$$

Kernel function measures the similarity between pairs of inputs.

$\kappa : \langle \phi(\mathbf{D}), \phi(\mathbf{D}') \rangle_{\mathcal{H}}$

prevents the informational mutation by translate, rotate

this kernel measures the similarity of two molecular pair

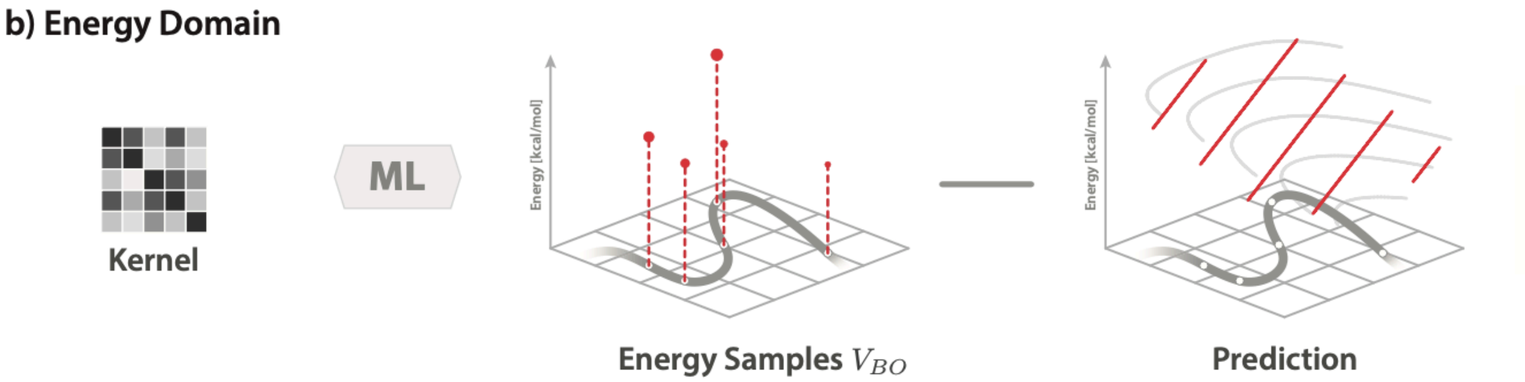
$$\mathbf{K}_F(\mathbf{R}, \mathbf{R}') = \nabla_{\mathbf{R}} \nabla_{\mathbf{R}'}^{\top} \kappa(\mathbf{R}, \mathbf{R}')$$

Differentiate the kernel to construct the force covariance.

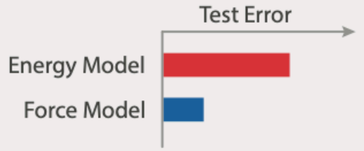
$$\left(\mathbf{K}_{\mathrm{Hess}(\kappa)} + \lambda \mathbf{I}\right) \boldsymbol{\alpha} = \nabla V_{\mathrm{BO}} = -\mathbf{F}$$

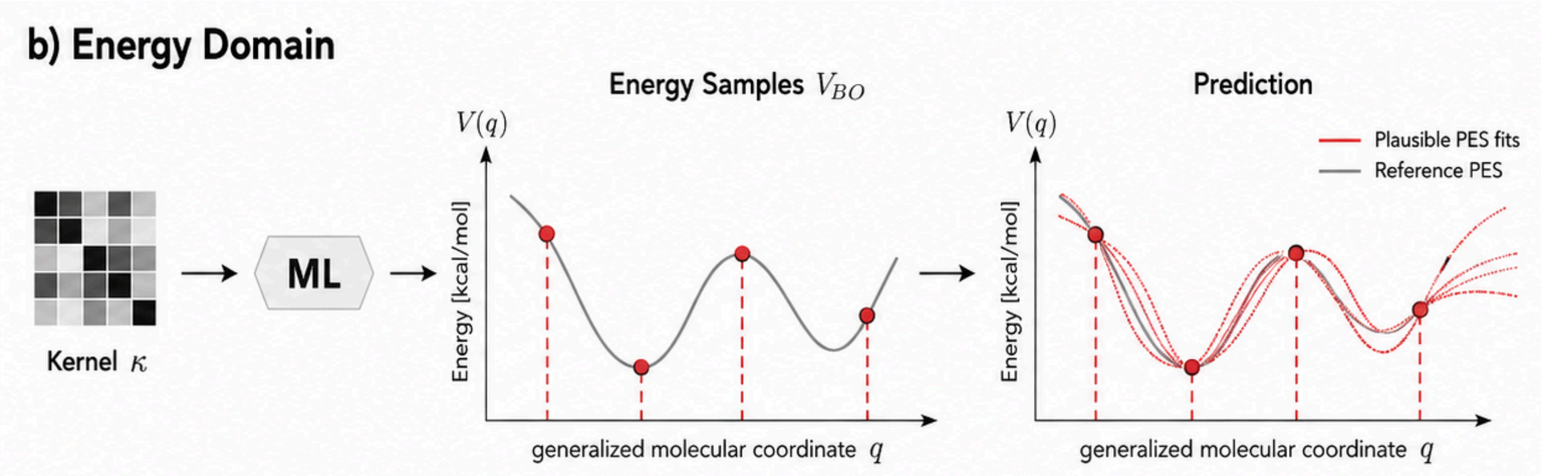
I = indentity matrix (Bias)
α = weight

Why Force-Domain Supervision Contains More PES Information

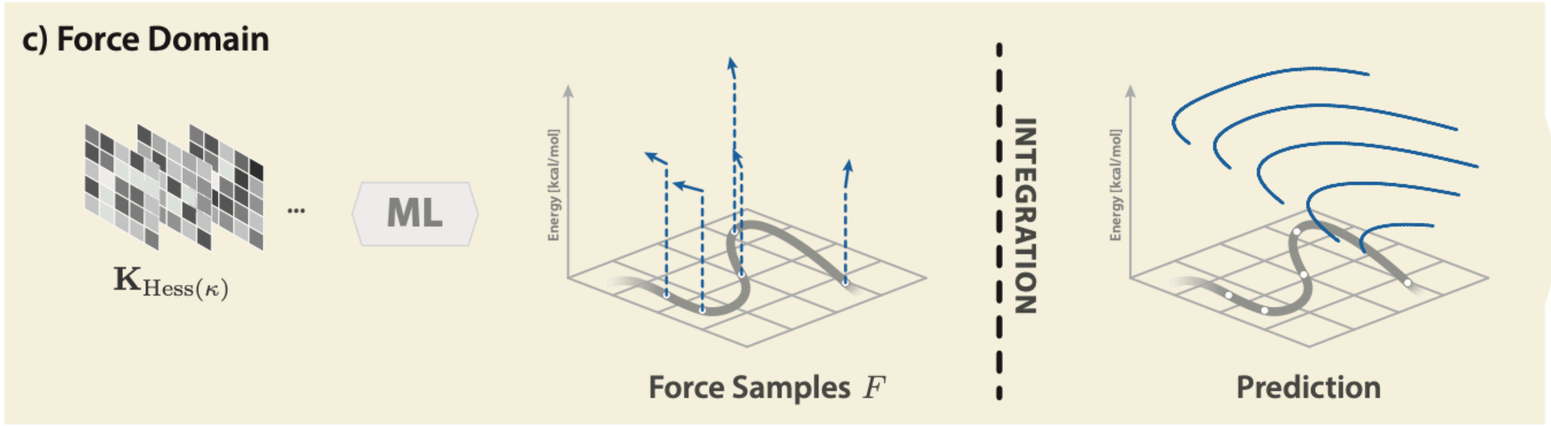


Problem:
Energy-based model lacks detail in under-sampled regions.





Why Force-Domain Supervision Contains More PES Information



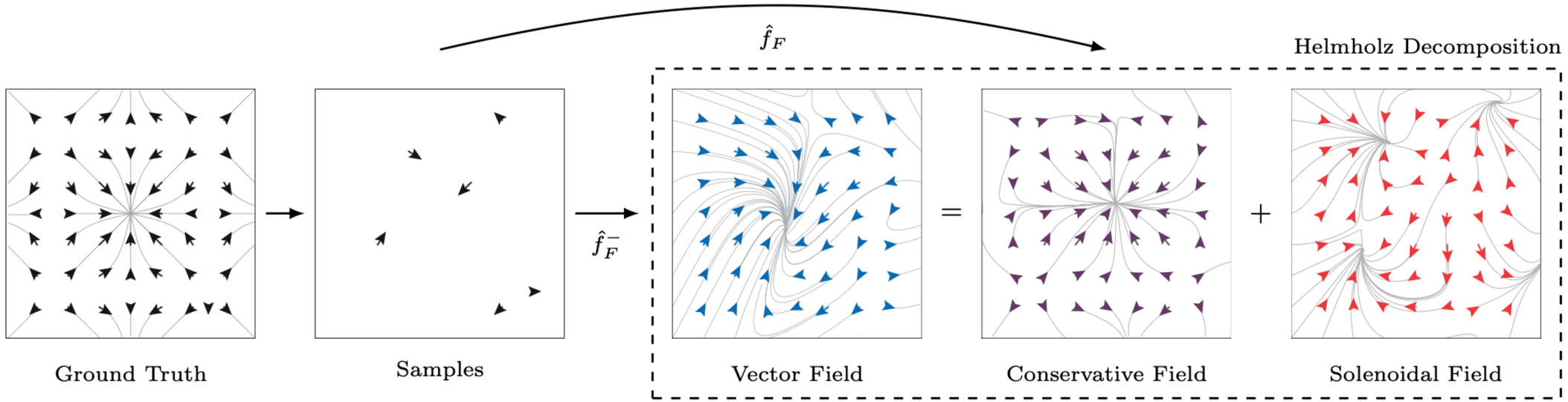
Solution:
Training in the force domain accurately reproduces PES topology.

$$\hat{f}_F(\vec{x}) = \sum_{i=1}^{M} \sum_{j=1}^{3N} (\vec{\alpha}_i)_j \frac{\partial}{\partial x_j} \nabla\kappa(\vec{x}, \vec{x}_i)$$

$$\hat{f}_E(\vec{x}) = \sum_{i=1}^{M} \sum_{j=1}^{3N} (\vec{\alpha}_i)_j \frac{\partial}{\partial x_j} \kappa(\vec{x}, \vec{x}_i)$$

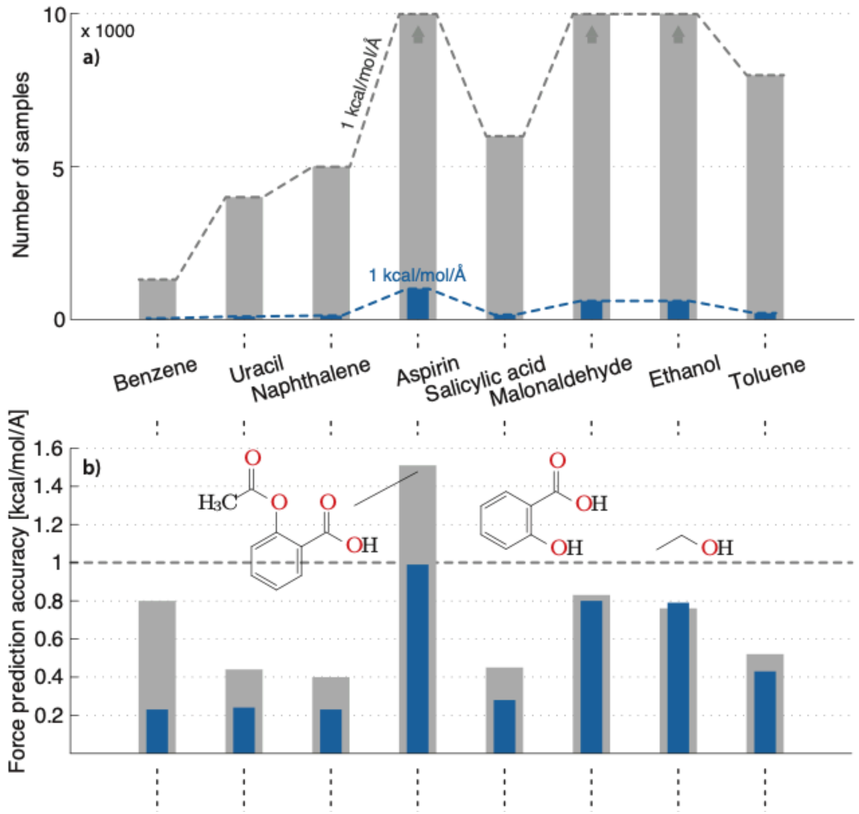
These general priors are sufficient
to reproduce good estimates from a restricted number of force samples.

What Energy Conservation Removes: The Solenoidal Error



true force field is conservative

Does the Physical Prior Improve Data Efficiency?



While GDML achieved the target force accuracy with approximately 1,000 geometries, energy-based models may require up to two orders of magnitude more samples.

Converged GDML exhibits a lower force error than an equivalent energy-based model.

From Force Accuracy to Molecular Dynamics

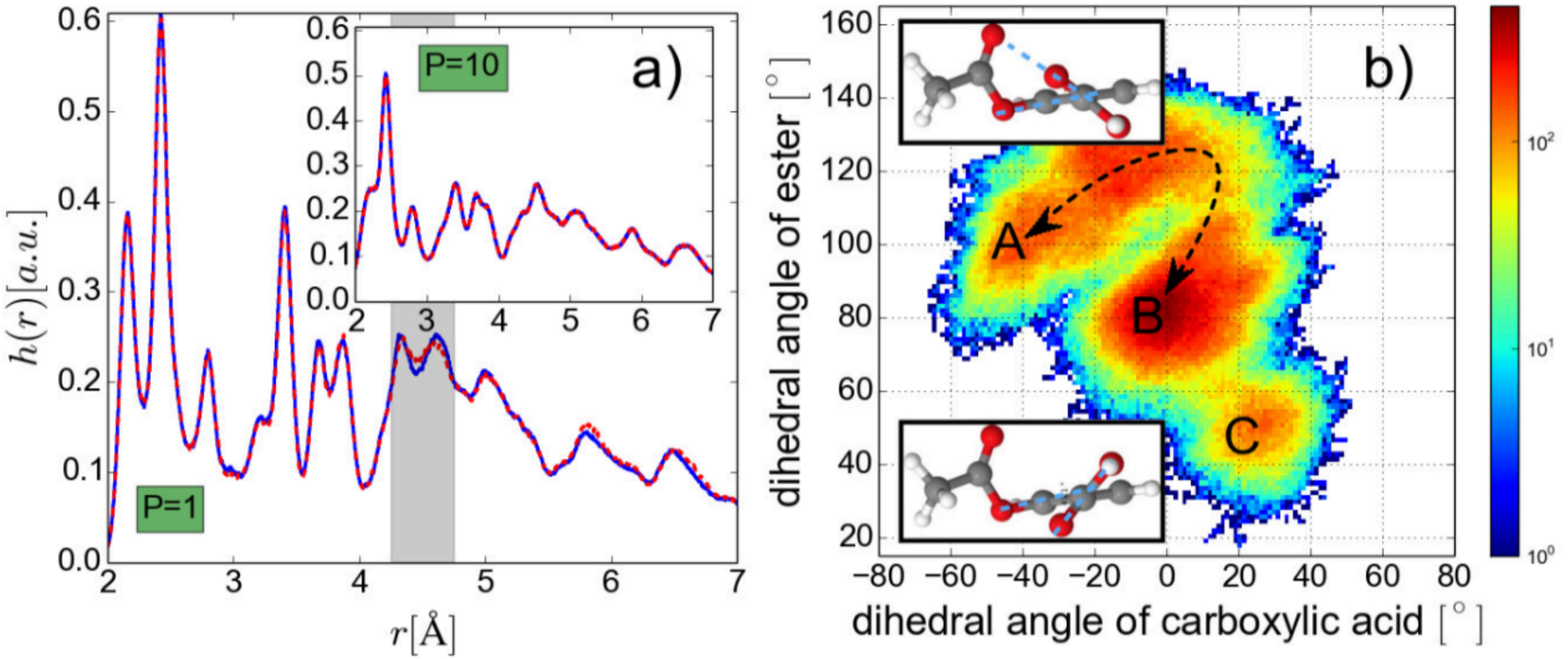
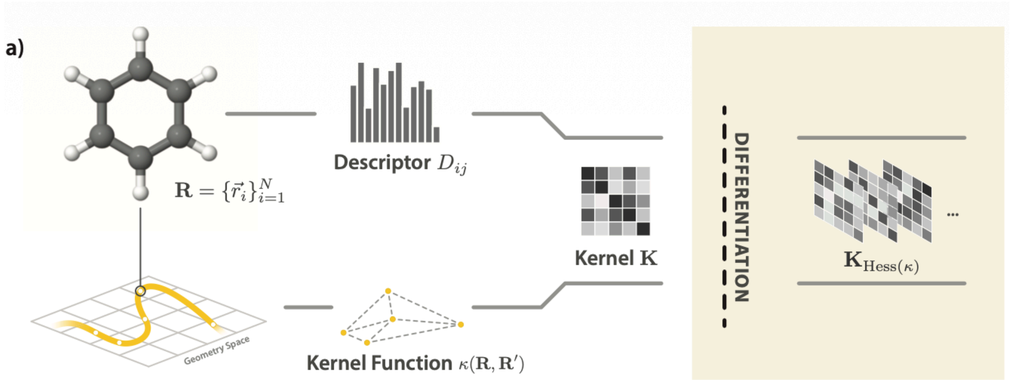
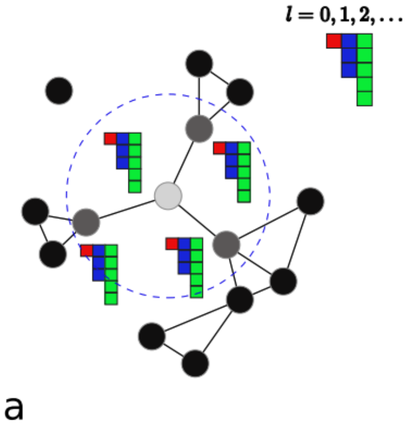


Fig. 4a: aspirin interatomic-distance distribution
Fig. 4b: dihedral-angle probability distribution

From Force Accuracy to Molecular Dynamics

GDML, 2017 ⟶ Nequlp, 2022





What survived?

1. Energy–force consistency
2. Physical structure is imposed, not learned from scratch
3. Forces are information-rich PES observations
4. The final target is molecular dynamics

Q&A?

Thanks!

JISEUNG HONG